Department of Clinical Epidemiology

Department of Clinical Epidemiology
Aarhus University Hospital
Olof Palmes Allé 43-45
DK-8200 Aarhus N
Denmark

Tel.: +45 871 68063
Fax: +45 871 67215
www.kea.au.dk

Report no. 86

Venous thromboembolism: risk factors and risk of subsequent arterial thromboembolic events

Research Year Report | Henrik Solli

Aarhus University Hospital

AARHUS UNIVERSITY

FACULTY OF HEALTH SCIENCES, AARHUS UNIVERSITY, DENMARK

Venous thromboembolism:
risk factors and risk of subsequent arterial thromboembolic events

Research year report

Henrik Solli

Department of Clinical Epidemiology, Aarhus University Hospital

Supervisors

Henrik Toft Sørensen, MD, PhD, DMSc, Professor (main supervisor)
Department of Clinical Epidemiology
Aarhus University Hospital, Denmark

Morten Olsen, MD, PhD (co-supervisor)
Department of Clinical Epidemiology
Aarhus University Hospital, Denmark

Morten Schmidt, MD (co-supervisor)
Department of Clinical Epidemiology
Aarhus University Hospital, Denmark

Preface

This report is based on research projects conducted during my research year at the Department of Clinical Epidemiology, Aarhus University Hospital, from February 2013 to January 2014.

First, I would like to express my sincerest gratitude to my main supervisor, Professor Henrik Toft Sørensen, for formulating the project idea and for giving me the opportunity to conduct a research year project at a leading research institution. Although Henrik is an internationally esteemed researcher with a very full work schedule, he has always found time to discuss and improve the research project, both practically and scientifically.

I would also like to thank my co-supervisors, Morten Olsen and Morten Schmidt, for their patience and supervision, and for teaching me skills in epidemiologic research and scientific writing. They have both gone beyond the call of duty to make sure that I have gotten the greatest possible yield from my research year and I feel very privileged to have been able to benefit from their expertise on a day-to-day basis. Professor Lars Pedersen has also been indispensable to the project for extracting the dataset and for giving advice regarding statistical methodology.

During my research project, I had the pleasure of being a visiting researcher at the Department of Biostatistics, Kansas University Medical Center, Kansas City, USA. The visit gave me the opportunity to broaden my knowledge of biostatistics and to experience a different culture, and I consider my research stay to be a memory for life. For this, I would like to give my sincerest thanks to my day-to-day collaborator Dr. Michael Brimacombe and the rest of the Department. I would also like to thank Henrik Toft Sørensen for initiating contact with the Department of Biostatistics.

Finn Breinholt Larsen at the Center for Public Health and Quality Improvement deserves recognition for leading the development of the “Hvordan Har Du Det?/How Are You?” lifestyle survey and for allowing my research group access to the survey data. Larsen and colleagues have developed an extensive and well-designed questionnaire that is of the utmost applicability for use in epidemiological research.

Finally, I would like to thank all employees at Department of Clinical Epidemiology; the administration, statistical staff, and fellow PhD students and research year students, for providing a warm and encouraging working environment and for making my research year a good learning experience.

Henrik Solli

Funding

My research year project was made possible by financial support from:

Department of Clinical Epidemiology

The Danish Heart Foundation

Department of Clinical Medicine's travel grant

Fonden af 17.12.1981

Contact Information

Henrik Solli

E-mail: hsol@clin.au.dk

Address: The Department of Clinical Epidemiology,
Aarhus University Hospital
Olof Palmes Allé 43-45, 8200 Aarhus N

Telephone: +45 871 68 212

Fax: +45 871 67 215

List of abbreviations

ATC	Anatomical Therapeutic Chemical Classification System (drug class)
BMI	Body mass index
CCI	Charlson comorbidity index
CI	Confidence interval
DNDRP	Danish National Database of Reimbursed Prescriptions
DNRP	Danish National Registry of Patients
g	Gram
HAY	"How Are You?" (questionnaire)
HR	Hazard ratio
ICD	International Classification of Diseases
ICD-8	International Classification of Diseases, 8th revision
ICD-10	International Classification of Diseases, 10th revision
IR	Incidence rate
MICE	Multiple imputation by chained equations
PAI-1	Plasminogen activator inhibitor-1
PY	Person-year
Q	Question (in appendix)
VTE	Venous thromboembolism

Table of contents

Abstract in English 3
Abstract in Danish 5
Extract 7
Introduction 7
Methods 8
Results 12
Discussion 14
Conclusion 17
Supplement 19
Subanalysis 19
Additional methodological considerations 21
References 33
Figures 43
Tables 45
Appendix 53

Abstract in English

Background: The impact of smoking and obesity on the risk of venous thromboembolism (VTE) is not entirely clear. VTE is also a risk factor for arterial thromboembolic events, such as myocardial infarction and ischemic stroke, but whether the association is causal or due to confounding by shared risk factors is unknown.

Objectives: 1) To examine the risk of VTE for smokers compared with non-smokers and for obese individuals compared with persons of normal weight; and 2) to examine the risk of arterial thromboembolic events in VTE patients compared with individuals without VTE, while taking into account possible confounding by shared risk factors.

Methods: The study population was comprised of responders to the Danish survey "How Are You?", which contains self-reported data on several lifestyle factors for individuals aged 25 to 79 years. We used the Danish National Registry of Patients covering all Danish hospitals to identify diagnoses of VTE, myocardial infarction, ischemic stroke, and comorbidity. We used the Danish National Database of Reimbursed Prescriptions to identify the cohort's filled prescriptions of statins, vitamin K antagonists and low-dose aspirin. Survey participants were followed from return of questionnaire in spring 2006 until outcome, emigration, death, or 31 December 2012, whichever came first. We used Cox regression analysis to compute hazard ratios (HRs) with 95% confidence intervals (CIs), and adjusted for lifestyle factors, comorbidity, and drug use.

Results: We found an elevated risk of VTE among smokers compared with never smokers (HR: 1.33; 95% CI: 0.91-1.94) and among obese individuals compared with individuals of normal weight (HR: 1.83; 95% CI: 1.21-2.77). Compared with individuals without VTE, individuals with VTE had crude HRs of 1.05 (95% CI: 0.34-3.27) for myocardial infarction, 1.93 (95% CI: 0.86-4.33) for ischemic stroke, and 1.59 (95% CI: 0.82-3.07) for the combined outcome. The adjusted HRs were 0.60 (95% CI: 0.19-1.90) for myocardial infarction, 1.09 (95% CI: 0.48-2.47) for ischemic stroke, and 0.91 (95% CI: 0.47-1.77) for the combined outcome.

Conclusion: Both smoking and obesity were associated with an increased risk of VTE. In the analysis of VTE and risk of arterial thromboembolic events, the risk estimates diminished substantially upon adjustment for age and sex, indicating that most of the observed association may be explained by confounding by these two variables. Further adjustment for lifestyle factors, comorbidity, and medications had little impact on the risk estimates. The risk estimates are imprecise, and should be interpreted with caution.

Abstract in Danish

Baggrund: Indflydelsen af rygning og overvægt på risikoen for at udvikle venøs tromboemboli (VTE) er ukendt. VTE er også en risikofaktor for arterielle tromboemboliske events såsom myokardieinfarkt og iskæmisk apopleksi, men det er uklart om associationen er kausal eller skyldes fælles risikofaktorer.

Formål: 1) At undersøge risikoen for at udvikle VTE blandt rygere sammenlignet med ikke rygere og blandt overvægte sammenlignet med normalvægtige, og 2) at undersøge risikoen for arterielle tromboemboliske events blandt VTE patienter sammenlignet med individer foruden VTE ved at kontrollere for eventuel confounding fra fælles risikofaktorer.

Metode: Studiepopulationen bestod af respondenter til den danske sundhedsundersøgelse "Hvordan Har Du Det?", som indeholder selvrapporteret livsstilsdata for personer mellem 25 og 79 år. Vi benyttede Landspatientregisteret til at udtrække diagnoser for VTE, myokardieinfarkt, iskæmisk apopleksi samt comorbiditet, og Dansk Receptdatabase til at identificere kohortens brug af statiner, vitamin K antagonister og lavdosis aspirin. Undersøgelsesdeltagerne blev fulgt fra deres returnering af spørgeskemaet foråret 2006 og frem til den første af følgende begivenheder: udfald, emigration, død eller 31. december 2012. Vi benyttede Cox regression til at beregne hazard ratioer (HR) med 95% konfidensinterval (CI), justerede for livsstilsfaktorer, comorbiditet og medicinbrug.

Resultater: Vi fandt en øget risiko for VTE for rygere sammenlignet med ikke-rygere (HR: 1.33; 95% CI: 0.91-1.94) og for overvægtige sammenlignet med normalvægtige (HR: 1.83; 95% CI: 1.21-2.77). Personer med VTE havde en ikke-justeret HR på 1.05 (95% CI: 0.34-3.27) for myokardieinfarkt, 1.93 (95% CI: 0.86-4.33) for iskæmisk apopleksi og 1.59 (95% CI: 0.82-3.07) for det kombinerede udfald sammenlignet med personer foruden VTE. De tilsvarende fuldt justerede HR var 0.60 (95% CI: 0.19-1.90) for myokardieinfarkt, 1.09 (95% CI: 0.48-2.47) for iskæmisk apopleksi og 0.91 (95% CI: 0.47-1.77) for det kombinerede udfald.

Konklusion: Både rygning og overvægt var associeret med en øget risiko for VTE. For associationen mellem VTE og efterfølgende arterielle tromboemboliske events faldt risikoestimaterne betragteligt ved justering for alder og køn. Dette indikerer at det meste af den observerede association kan forklares med confounding fra disse to variabler. Yderligere justering for livsstilsfaktorer, comorbiditet og medicinbrug havde meget lidt indvirkning på risikoestimaterne. Risikoestimaterne er upræcise, og bør tolkes med forsigtighed.

Extract

Introduction

Venous thromboembolism (VTE), which includes both deep venous thrombosis and pulmonary embolism, has a reported incidence rate per 1000 person-years of 1.5 in individuals aged 15 and older,[1] rising to 3.1 in individuals aged 85-89 years.[2,3] The disease is a serious and potentially lethal condition,[4] and is a leading cause of preventable in-hospital deaths in the United States.[5]

Conditions leading to a challenged cardiovascular system, such as pregnancy, surgery, trauma, cancer, and immobilization, are all established risk factors for VTE.[2] However, lifestyle factors and their contribution to risk of VTE have not been well examined and understood. Although smoking is a well-established risk factor for arterial thromboembolic events, it remains controversial whether smoking is also a risk factor for venous thrombosis.[6-14] Obesity has continuously been cited as a risk factor for VTE.[9-14] However, fewer studies have investigated whether the association differs between provoked and unprovoked (idiopathic) VTE.[9,11,14] Provoked VTE is defined as a VTE event predating occult cancer or being secondary to fracture, surgery, trauma, pregnancy or prevalent cancer, while an unprovoked VTE event occurs in absence of these conditions.[15]

Several studies have pointed to VTE itself as risk factor for developing subsequent arterial thromboembolic events, such as myocardial infarction and ischemic stroke.[16-22] One study concluded that the elevated risk of acute arterial events among individuals with VTE was attributable solely to confounding by shared risk factors such as age, sex, obesity, smoking, comorbidity, and thrombophilia,[22] while most other studies were unable to adjust their risk estimates for the same covariates.[16-21]

We therefore identified responders to an extensive lifestyle questionnaire and conducted a cohort study with the following objectives: 1) To examine the risk of VTE for smokers compared with non-smokers and for obese individuals compared with persons of normal weight; and 2) to examine the risk of arterial thromboembolic events in VTE patients compared with individuals without VTE, while taking into account possible confounding by shared risk factors.

Methods

Setting

We conducted this cohort study among residents in one of Denmark's five regions, the Central Denmark Region, which has a population of approximately 1.2 million individuals.[23] The Danish National Health Service provides universal tax-supported healthcare, which guarantees free and unfettered access to general practitioners and hospitals, and partial reimbursement for prescribed medications.[24] Accurate and unambiguous individual-level linkage of all Danish registries is possible using the unique civil registration number assigned to all residents at birth or upon immigration.[25]

Study cohort

The study population included all individuals, who returned the questionnaire-based public health survey "Hvordan Har Du Det?"/"How Are You?", which was conducted by the Center for Public Health and Quality Improvement, Central Denmark Region.[26] The questionnaire was distributed by mail in February of 2006 to 31,500 randomly selected inhabitants of the Central Denmark Region. Eligible participants were defined as Danish citizens aged 25-79 years with at least one parent born in Denmark and who were residents of the Central Denmark Region at the time the questionnaire was issued.[27]

Study participation was voluntary, and 21,602 individuals agreed to answer the questionnaire, corresponding to a response proportion of 69%.[28] Participants provided self-reported answers to a detailed questionnaire regarding lifestyle factors containing approximately 400 questions.[28,29]

Data regarding age, sex, death, and emigration

We used the Danish Civil Registration System to collect data regarding age, sex, all-cause mortality, and emigration.[25] This registry is updated daily and contains vital statistics of members of the Danish population – including date of birth, change of address, date of emigration, and exact date of death – dating back to 1968.[25] We divided age into five groups: 25-39, 40-49, 50-59, 60-69, and 70-79 years.

Data regarding cardiovascular disease

We used the Danish National Registry of Patients (DNRP) to identify all individuals with a first-time diagnosis of VTE (defined as deep venous thrombosis or pulmonary embolism), myocardial infarction, or ischemic stroke. The DNRP records information on patients discharged from all Danish non-psychiatric hospitals since 1 January 1977 and from all emergency room and outpatient specialty clinic visits since 1995.[30] Each hospital discharge or outpatient visit is recorded in the registry with one primary diagnosis and one or more secondary diagnoses classified according to the International Classification of Diseases, 8th revision (ICD-8) until the end of 1993 and the 10th revision (ICD-10) thereafter.[30]

We further subdivided VTE diagnoses into provoked and unprovoked events (Figure 1). A VTE event was defined as provoked if the patient had a history of malignancy any time prior to or within 90 days after the diagnosis of VTE or a recorded discharge diagnosis of fracture, trauma, surgery, or pregnancy within 90 days before the VTE diagnosis.[15] Conditions defining provoked VTE were recorded in the DNRP. Remaining cases were classified as unprovoked.[15]

Data regarding lifestyle factors

Information regarding the following lifestyle factors was extracted from the "How Are You?" survey: smoking status, body mass index (BMI), physical activity, and level of education.

Smoking status was categorized into three categories: never smokers, former smokers, and current smokers. Among current smokers that also reported daily consumption of tobacco products, we calculated the total tobacco consumption in grams (g) per day using the following formula: 1 cigarette equals 1 g, 1 cigar equals 4.5 g, 1 cheroot equals 3 g, and 1 pipe stop equals 3 g of tobacco.[8]

BMI was calculated as weight in kilograms divided by height in meters squared and grouped into four categories, as defined by the World Health Organization: underweight (BMI < 18.5), normal weight (18.5 ≤ BMI < 25), overweight (25 ≤ BMI < 30), and obese (BMI ≥ 30).[31]

Physical activity was dichotomized with respect to whether or not the individual conducted any form of sport or other physical activity in their spare time on a regular basis.

Level of education was grouped into four categories: ≤ 7 years of primary education, 8-10 years of primary education, high school education, vocational training or equivalent secondary education, and finally higher education (post-secondary education of varying length).

Data regarding drug use

We used the Danish National Database of Reimbursed Prescriptions (DNDRP) to identify the use of different classes of cardiovascular medications among study participants.[32] This database encompasses the reimbursement records of all reimbursable drugs sold in community pharmacies and hospital-based outpatient pharmacies in Denmark since 2004. Stored in the database is information regarding the dispensed drug (name, form, strength, and pack size), patient, prescriber, and pharmacy.[32] Drug use was defined as a registered dispensement of a reimbursable drug within 90 days before start of follow-up. We collected data on the following classes of drugs: statins, vitamin K antagonists, and low-dose aspirin.

Data regarding comorbidity

We used the Charlson comorbidity index (CCI) to assess each individual's burden of comorbidity.[33] The CCI assigns between one to six points to a range of diseases depending in their severity and anticipated 1-year mortality.[33] Diagnoses included in the CCI were extracted from the DNRP.[34] For all 19 comorbidities in the index, we computed the cumulative score for each individual study participant and defined three categories of comorbidity based on scores of 0 (low), 1-2 (moderate), and ≥ 3 (high).[35]

Prevalent hypertension and diabetes were determined by extracting hypertension- and diabetes-related diagnoses from the DNRP, as well as prevalence of either disease stated in the "How Are You?" questionnaire. We also used the redemption of anti-diabetic drugs through the DNDRP to identify diabetic patients.

Statistical analysis

All individuals were followed from the date of return of the questionnaire in spring 2006 until the date of outcome, death, emigration, or 31 December 2012, whichever came first (Figure 1). All data

was linked at the level of the individual using the civil registration number. For individuals with missing information regarding one or more lifestyle factors in the "How Are You?" survey, we imputed the missing values using the multiple imputation by chained equations (MICE) method.[36] For our risk estimates, we compute incidence rates (IRs) per 1000 person-years for all exposure categories, and used Cox regression to compute crude and adjusted hazard ratios (HRs) with 95% confidence intervals (CIs) for all outcome events. The proportional hazard assumption was assessed graphically for all analyses by plotting log(-log(survival function)) against time, and it was found to be valid.

For the analysis of smoking and obesity as risk factors for VTE, we excluded all individuals with a recorded diagnosis of VTE before start of follow-up from the study population (n=200). We calculated IRs and HRs for VTE overall, provoked VTE, and unprovoked VTE for both smoking status and BMI group. Because up to 30% of individuals that survive a first VTE event develop a recurrent event,[2] we censored all analysis time at the date of the first VTE event, regardless whether the event was provoked or unprovoked. We did not adjust for diabetes when BMI group was the exposure category, as diabetes may be an intermediary step on the causal pathway between obesity and VTE.

For the analysis of VTE and subsequent risk of arterial thromboembolic events, we excluded all individuals with a recorded diagnosis of either myocardial infarction or ischemic stroke before start of follow-up from the study population (n=682). Exposed individuals were defined as individuals with both a completed questionnaire and a diagnosis of VTE recorded in the DNRP between 1977 and the end of follow-up (Figure 2). If an individual had a recorded diagnosis of VTE before return of the questionnaire, their exposed risk time was recorded from the day that they returned the questionnaire. If a first diagnosis of VTE was recorded during the follow-up period, the individual started follow-up as unexposed on the day of return of the questionnaire, and the exposure status was changed to exposed on the day of the VTE diagnosis and until the end of follow-up. Individuals without a diagnosis of VTE before or during follow-up contributed all their risk-time as unexposed. For this analysis, myocardial infarction, cerebrovascular disease, and hemiplegia were removed from the CCI, because these conditions are outcome events or are strongly associated with the outcome. We assessed IRs and HRs for the outcomes for overall, provoked, and unprovoked VTE separately.

All statistical analyses were conducted using Stata software version 12.1 (STATA, College Station, Texas, USA).

Results

Smoking and risk of VTE

Study population characteristics according to smoking status are presented in Table 1. Current and former smokers were older, and had a higher burden of comorbidity and a higher proportion of males compared with never smokers. Current smokers had a lower proportion of individuals with a high education level and a lower proportion of physically active individuals, compared with both former and never smokers. Former smokers were more likely to be diagnosed with hypertension or to be users of statins, compared with both current and never smokers.

Risk estimates are presented in Table 2. We observed IRs per 1000 person-years for overall VTE of 0.94 (95% CI: 0.73-1.22) for never smokers, 1.15 (95% CI: 0.86-1.54) for former smokers, and 1.62 (95% CI: 1.25-2.09) for current smokers. Using never smokers as the reference, the adjusted HR was 1.33 (95% CI: 0.91-1.94) for current smokers and 0.76 (95% CI: 0.51-1.13) for former smokers regarding overall VTE. Among current smokers, we observed a higher adjusted point estimate for smokers of ≥ 20 g of tobacco/day (HR: 1.87; 95% CI: 1.19-2.96) than for smokers with a lower tobacco consumption (HR: 1.00; 95% CI: 0.61-1.63). Current smoking was more strongly associated with provoked VTE (HR: 1.60; 95% CI: 0.93-2.75) than unprovoked VTE (HR: 1.09; 95% CI: 0.63-1.89).

Obesity and risk of VTE

Study population characteristics according to BMI group are presented in Table 3. Compared with the total study population, underweight and normal weight individuals had a higher proportion of females, while overweight individuals had a higher proportion of males. Both underweight and obese individuals had a higher proportion of physically inactive individuals and greater burden of comorbidity than the total study population. Underweight individuals were also more likely to be current smokers. Obese individuals were more likely to be diagnosed with both diabetes and hypertension, and were more likely to be users of all three recorded drug classes, especially vitamin K antagonists and low-dose aspirin.

Risk estimates are presented in Table 4. We found the following IRs per 1000 person-years for overall VTE: 0.96 (95% CI: 0.75-1.22) for normal weight, 1.59 (95% CI: 0.51-4.93) for underweight, 1.28 (95% CI: 1.00-1.63) for overweight, and 2.14 (95% CI: 1.58-2.91) for obese individuals. The adjusted HR for VTE was 1.27 (95% CI: 0.39-4.08) for underweight, 1.10 (95% CI: 0.77-1.57) for overweight, and 1.83 (95% CI: 1.21-2.77) for obese individuals. For overweight and obesity, risk estimates did not differ substantially between provoked and unprovoked VTE. We did observe a stronger association for provoked VTE than for unprovoked VTE among underweight individuals. However, these estimates were imprecise owing to the low number of outcomes in this BMI group.

VTE and risk of subsequent arterial thromboembolic events

Study population characteristics according to VTE status are presented in Table 5. Compared with individuals without VTE at the start of follow-up, the VTE cohort were older; had a higher BMI class and a lower proportion of physically active individuals; and also exhibited a higher burden of comorbidity, hypertension, and diabetes. Individuals in the VTE cohort were also more likely to be users of all three recorded drug classes.

Risk estimates are presented in Table 6. For the VTE cohort, we calculated an IR per 1000 person-years of 1.98 (95% CI: 0.64-6.15) for myocardial infarction and 4.00 (95% CI: 1.80-8.91) for ischemic stroke. The corresponding rates for the non-VTE cohort were 1.89 (95% CI: 1.67-2.14) and 2.06 (95% CI: 1.83-2.31), respectively. For the crude estimates, we observed a null association for myocardial infarction (HR: 1.05; 95% CI: 0.34-3.27) and a two-fold increased risk for ischemic stroke (HR: 1.93; 95% CI: 0.86-4.33). Upon adjusting for age and sex only, the estimates decreased severely [HR of 0.64 (95% CI: 0.21-2.01) and 1.14 (95% CI: 0.51-2.56, respectively]. Adding adjustment for important lifestyle factors and chronic diseases (smoking, BMI, diabetes, hypertension, and CCI score) had very little effect on the estimates compared with the age- and sex-adjusted estimates. Finally, additional adjustment for medications, physical exercise, and level of education had a negligible impact on the estimates. Unprovoked VTE was more strongly associated with ischemic stroke than provoked VTE; however, the opposite was true for myocardial infarction.

Discussion

In our analyses, we found that both current smoking and obesity were associated with an increased risk of VTE. Smoking might be more strongly associated with provoked VTE than with unprovoked VTE, while the risk estimates for obesity were similar for the two sub-categories of VTE. Most of the observed association between VTE and arterial thromboembolic events could be explained by confounding by age and sex, while lifestyle factors, comorbidity, and medications had little impact on the risk estimates.

Comparison with other studies

Smoking and risk of VTE

Although some studies have observed a positive association between smoking and VTE,[7,8,11-13] others have reported no elevated risk.[6,9,10,14] A recent meta-analysis on the topic pooled the results of 32 observational studies involving 3,966,184 participants and 35,151 VTE events.[37] The authors found a relative risk of VTE of 1.17 (95% CI: 1.09-1.25) for ever smokers, 1.23 (95% CI: 1.14-1.33) for current smokers, and 1.10 (95% CI: 1.03-1.17) for former smokers. Including only studies that adjusted for BMI yielded a higher relative risk for current smokers (1.30; 95% CI: 1.24-1.37). Our findings of a fully adjusted HR of 1.33 (95% CI: 0.91-1.94) for current *vs.* never smokers are in accordance with these findings. We also observed a stronger association for provoked versus unprovoked VTE. Most previous studies that divided VTE into provoked and unprovoked cases have also reached this conclusion,[6,8,9] apart from one study that found a reverse relation.[11]

Obesity and risk of VTE

Obesity is one lifestyle factor that has been consistently shown to be associated with VTE.[9-14] In 2008, a meta-analysis pooling nine studies with a total of 8,125 VTE events found an odds ratio of 2.33 (95% CI: 1.68-3.24) for VTE for obese individuals. Upon restricting the analysis to high-quality studies in which BMI was adequately measured, the odds ratio decreased to 1.84 (95% CI: 1.55-2.18).[38] Our results are in accordance with these findings.

VTE and risk of subsequent arterial thromboembolic events

The largest study on the association was conducted in Denmark by Sørensen et al.[16] Here, the authors found a HR of 1.60 (95% CI: 1.35-1.91) for myocardial infarction and 1.85 (95% CI: 1.44-2.37) for ischemic stroke among patients with deep venous thrombosis, and a HR of 2.60 (95% CI: 2.14-3.14) for myocardial infarction and 2.34 (95% CI: 1.66-3.31) for ischemic stroke among patients with pulmonary embolism, all describing risks within 1 year of follow-up.[16] Extending the follow-up period to 2-20 years provided less markedly elevated risks, at approximately 20-30% for both arterial cardiovascular events.[16] Although we found no association between VTE and arterial thromboembolic events after adjusting for the full set of covariates in our study, the low precision of the risk estimates renders interpretation of the results and comparisons with other studies difficult. The observed null association in our study does not necessary mean that a true positive association does not exist.[39] We especially believe that the observed negative association for myocardial infarction is not causal, but rather due to lack of precision.

To our knowledge, only one other study has been able to adjust analyses of the association between VTE and arterial thromboembolic events for potential confounding by shared risk factors.[22] Roach et al. concluded that the increased risk of arterial thromboembolic events after VTE was due to confounding by shared risk factors, as their HRs diminished from 2.2 (95% CI: 1.2-3.8) in patients compared with random digit dialing controls and 1.5 (95% CI: 1.0-2.3) in patients compared with partners to 1.8 (95% CI: 0.8-4.2) and 1.3 (95% CI: 0.7-2.5), respectively, after adjustment for all confounders (age, sex, BMI, smoking, chronic disease, malignancy, genetic thrombophilia, and procoagulant markers).

Although we also observed a null association for the combined outcome of myocardial infarction and ischemic stroke, we have reached a different conclusion regarding the impact of common risk factors on the association. After the estimates were adjusted for age and sex, further adjustment for possible confounding factors had very little impact on the estimates. Roach et al. did not present risk estimates that were solely adjusted for age and sex, thereby rendering interpretations of the impact of lifestyle factors alone on the association difficult to quantify.[22]

Strengths and limitations

Some issues should be considered when interpreting our results. The free access to health care provided by the Danish Health System and the computerized manner in which discharge diagnoses are recorded minimized referral and diagnostic biases of the diagnoses of VTE, myocardial infarction, and ischemic stroke. Regarding the validity of these diagnoses recorded in the DNRP, the positive predictive value has been reported as approximately 94% for myocardial infarction,[40] approximately 88-94% for ischemic stroke,[34,41] and approximately 75% for VTE.[42] As we believe the diagnostic specificity to be independent of exposure status, any misclassification of the outcome diagnoses due to lack of specificity would bias the risk estimates towards the null.

A major strength of our study is the fact that we supplemented DNRP data regarding diabetes mellitus and hypertension with self-reported prevalence of these diseases from the "How Are You?" survey, as well as recorded usage of anti-diabetic medications from the DNDRP. Because some diabetic and hypertensive individuals are treated solely by their general practitioner, relying on discharge data to identify diabetic and hypertensive patients may lead to only identifying the most severe cases of these two diseases, which again may lead to an overestimation of their contribution as risk factors.

The systematic underreporting of BMI among overweight individuals and tobacco consumption among smokers is well documented in the literature.[43,44] Because the information in the questionnaire was collected at the start of follow-up, we find it unlikely that the manner in which the questionnaire was answered should be associated with whether or not the individual experienced a study outcome.

The study design of VTE and risk of arterial thromboembolic events allowed individuals to obtain a status of exposed either at onset of follow-up or during follow-up; we are unable to determine whether an individual that was diagnosed with VTE after return of the questionnaire subsequently changed one or more aspects of their lifestyle profile. It can be argued that a person suffering from a disease with potential substantial morbidity will most likely change his or her lifestyle profile in a healthier rather than an unhealthier direction.[4]

Finally, the low number of outcomes (*e.g.*, only nine myocardial infarctions or ischemic strokes after a VTE event) was a major limitation in all analyses. Even though our study population was large, it was comprised of a sample of the general population rather than a hospital population.

Therefore, the vast majority of study participants did not experience an outcome, leading to risk estimates with a low precision. This emphasizes the need for additional studies with a larger study population and longer follow-up time in order to clarify the investigated associations.

Conclusion

Both smoking and obesity were associated with an increased risk of VTE. In the analysis of VTE and risk of arterial thromboembolic events, the risk estimates diminished substantially upon adjustment for age and sex, indicating that most of the observed association may be explained by confounding by these two variables. Further adjustment for lifestyle factors, comorbidity, and medications had little impact on the risk estimates. The risk estimates are imprecise, and should be interpreted with caution.

.

Supplement

The following section of the research year report contains supplementary information, namely a subanalysis followed by general methodological considerations concerning the whole research year project.

Subanalysis

Introduction

It is unknown whether the observed association between obesity and VTE is mediated through physical inactivity. We therefore performed a subanalysis of obesity and risk of VTE among physically active and inactive individuals separately.

Methods

The same approach as described in the methods section of the extract apply to this subanalysis. We formed two subcohorts, one consisting of individuals that reported practicing sports or other kinds of physical activities on a regular basis in their spare time in the "How Are You?"-questionnaire, and another consisting of individuals that did not. Only individuals with non-missing information with regards to physical activity were included in this subanalysis. The risk estimates were adjusted for the same sets of covariates as the main analysis of obesity and risk of VTE, apart from physical activity, which showed no variability within the two subcohorts.

Results

Risk estimates for physically active and physically inactive individuals are presented in Supplementary table 1 and 2, respectively. The physically active subcohort consisted of 9,626 individuals. The physically inactive subcohort consisted of 11,351 individuals. A total of 425 individuals had missing information regarding physical activity and were excluded from analysis. 54 overall VTE events occurred in the physically active subcohort and 118 overall VTE events occurred in the physically inactive subcohort. We found the following IRs per 1000 person-years for physically active individuals: 0.56 (95% CI: 0.36-0.89) for normal weight, 0.92 (95% CI: 0.60-1.42) for overweight, and 1.98 (95% CI: 1.15-3.42) for obese individuals. Calculating an IR for

physically active underweight individuals was not possible, as no VTE events occurred within this BMI group. For physically inactive individuals, we found the following IRs per 1000 person-years: 1.33 (95% CI: 0.99-1.80) for normal weight, 2.46 (95% CI: 0.79-7.63) for underweight, 1.51 (95% CI: 1.11-2.05) for overweight, and 2.29 (95% CI: 1.58-3.31) for obese individuals.

The fully adjusted HRs for overall VTE for the physically active subcohort were as follows: 1.41 (95% CI: 0.74-2.66) for overweight and 3.32 (95% CI: 1.59-6.92) for obese individuals, both using normal weight individuals as reference. The HRs did not differ substantially for provoked VTE and unprovoked VTE. The fully adjusted HRs for overall VTE for the physically inactive subcohorts were: 1.62 (95% CI: 0.49-5.30) for underweight, 0.97 (95% CI: 0.63-1.51) for overweight, and 1.50 (95% CI: 0.91-2.47) for obese individuals, all using normal weight individuals as reference. Again, the risk estimates were similar for provoked and unprovoked VTE.

Discussion

In this subanalysis, our aim was to assess whether the effect of obesity on risk of VTE differed among physically active and physically inactive individuals. We hypothesized that the risk estimates would be lower in the physically active subcohort compared with the physically inactive, as physical inactivity has been proposed as a risk factor for VTE.[45] We observed that the fully adjusted risk estimates for obesity and risk of VTE differed between physically active (HR: 3.32; 95% CI: 1.59-6.92) and inactive individuals (HR: 1.50; 95% CI: 0.91-2.47), implying that physical activity may act as an effect measure modifier on the association. However, the results contradicted our initial hypothesis, as the HRs were higher for obese, physically active individuals than for obese, physically inactive individuals.

Comparing BMI groups across subcohorts, the IRs were consistently higher for physically inactive individuals compared with physically active individuals, with the exception of obesity and risk of provoked VTE, which had an equal IR for the two subcohorts. This might indicate that obesity and physical inactivity are additive risk factors for VTE. Furthermore, more than double the amount of VTE events occurred in the physically inactive subcohort compared with the physically active subcohort (118 and 54, respectively), while the number of person-years were fairly similar in the two subcohorts (74494 and 64529, respectively). Our surprising findings might in part be explained by a differentiated misclassification, as the perceived definition of what physical activity

encompasses may differ between BMI groups. To our knowledge, no other study has assessed physical activity as a possible effect measure modifier on the association between obesity and VTE.

Additional methodological considerations

Elaboration on the statistical analysis

In our study population, two individuals died in the timespan between filling out the questionnaire and the questionnaire being registered at Center for Health and Quality Improvement. Since follow-up started on the day the returned questionnaire was registered, these individuals did not contribute any risk-time, and were therefore excluded from the analyses. A total of 11 individuals turned 80 years old in the time between being randomly selected to receive a questionnaire and returning the questionnaire. These individuals were included in the age group covering individuals between 70 and 79 years of age. One individual reported an unrealistic height of 270 cm. The BMI of this individual was replaced with missing information and later imputed based on a set of predictive covariates.

Upon calculating the CCI, we excluded the less severe of two overlapping diseases from the index score for each individual. If an individual was diagnosed with moderate to severe liver disease, the index score of a possible mild liver disease was omitted. Similarly, a diagnosis of diabetes with end organ failure omitted a diagnosis of diabetes without end organ failure, and a diagnosis of metastatic cancer omitted a diagnosis of a non-metastatic solid tumor.

With regards to the definition of outcome diagnoses, we used primary ICD-8 and ICD-10 diagnoses to define first-time myocardial infarction and ischemic stroke, and both primary and secondary ICD-8 and ICD-10 diagnoses to define first-time VTE. Because the statistical analysis relied on time-to-event data, we used primary diagnoses to determine the timing of the outcome event as accurately as possible. The positive predictive values are also often higher for primary diagnoses than for secondary diagnoses. However, because VTE is often a concomitant disease acquired during hospitalization,[2] secondary diagnoses were included in the definition VTE.

For the analyses of smoking and obesity as risk factors for VTE, individuals were censored at any first-time VTE event, regardless of whether the outcome in question was overall, provoked or unprovoked VTE. As mentioned in the extract, this was done because the risk of a second VTE

event is not independent of a first-time VTE event. For this reason, the attributed person-years for each exposure category are the same for the analyses of overall, provoked and unprovoked VTE.

Choice of covariates in the regression model

With the analysis of VTE and risk of subsequent myocardial infraction and ischemic stroke, our aim was to assess whether the association was causal or resulted from confounding by shared risk factors. For this reason, fitting a relevant regression model was paramount to the interpretation of the results. Using several sources of data, in particular the "How Are You?"-questionnaire, we defined several potential covariates for the purpose of adjusting our risk estimates. For the final regression model, we settled on a set of covariates that showed an uneven distribution among exposure categories and were clinically relevant (Table 5). Covariates that were considered for inclusion in the regression model, but later excluded, are presented in the following table:

	Total study population	No VTE before start of follow-up	VTE before start of follow-up
Diet			
Unhealthy diet	2,793 (13.4)	2,769 (13.4)	24 (13.5)
Intermediate diet	13,318 (63.7)	13,212 (63.7)	106 (59.6)
Healthy diet	4,381 (20.9)	4,339 (20.9)	42 (23.6)
Missing information	428 (02.1)	422 (02.0)	6 (03.4)
Alcohol consumption			
Low intake	16,951 (81.0)	16,800 (81.0)	151 (84.8)
Moderate intake	2,260 (10.8)	2,245 (10.8)	15 (08.4)
High intake	1,232 (05.9)	1,225 (05.9)	7 (03.9)
Missing information	477 (02.3)	472 (02.3)	5 (02.8)
Household income			
Low	1,125 (05.4)	1,109 (05.4)	16 (09.0)
Medium	9,850 (47.1)	9,767 (47.1)	83 (46.6)
High	6,164 (29.5)	6,136 (29.6)	28 (15.7)
Missing information	3,781 (18.1)	3,730 (18.0)	51 (28.7)
Hormone replacement therapy	526 (02.5)	522 (02.5)	4 (02.3)

The distributions of diet and alcohol consumption were evenly distributed among individuals with and without VTE at start of follow-up, and both were relatively weak independent risk factors for the outcomes (data not shown). Diet may also be considered to be a proxy variable for both level of education and BMI group, both of which were included in the final regression model. Use of

hormone replacement therapy was also evenly distributed among exposure groups. Although the distribution of household income was somewhat skewed between exposure groups, with a higher proportion of low income among the VTE cohort and a higher proportion of high income among the cohort without VTE, we chose to exclude this covariate from the regression model. First, household income had a higher proportion of missing information than any other variable, which gave rise to concern for the validity of this covariate. The reason for the high amount of missing values in this category might be due to some individuals being less likely to disclose what is considered a personal issue. Although we utilized multiple imputation analysis to impute missing values in our dataset, a proportion of missing information exceeding 20 to 30 percent is considered less suited for imputation.[46] Second, household income is closely correlated to level of education, which was included in the final regression model.

In general, we chose to include covariates in the regression model which could be considered to fulfill the three requirements needed for a variable to act as a confounder on the association between VTE and arterial thromboembolic events (see later section for definition of a confounder). Based on the number of events, 259 myocardial infarctions in the unexposed cohort and 3 in the exposed cohort, and 282 ischemic stokes in the unexposed cohort and 6 in the exposed cohort, we predicted that the covariates were to be cautiously implemented in order for the Cox regression model not to collapse. We proceeded with a stepwise approach where we adjusted gradually for an increasing set of covariates (Table 6). The set of covariates used in the fully adjusted analysis contained a total of 12 covariates, of which seven were dichotomous and five were categorical. Traditionally, there has been a rule of thumb dictating that for any Cox or logistic regression model, a minimum of ten outcome events should be present per variable in the regression model. In recent years, however, a revision of this rule has been proposed.[47]

Multiple imputation

Because the "How Are You?"-questionnaire was distributed to study participants by mail and relied on self-reported information by the recipient, some of the questions for some individuals were left unanswered, leading to missing information. Missing information in health surveys is common, and needs to be handled appropriately. The missing information may arise due to failure to interpret the questionnaires instructions, an unwillingness to answer a particular question, often regarding private matters, or other reasons.

One way for researchers to address the problem with missing data is to include only individuals with complete information on all study covariates. This method of analysis is referred to as a complete case analysis. However, this method may lead to a bias and to exclusion of a substantial proportion of the study population, leading to fewer study outcomes and a loss of precision and power.[46]

One approach to handling missing information is multiple imputation.[46] With this method, multiple copies of the initial dataset are created. The missing values in each dataset are then replaced with imputed values sampled from a predictive model based on the observed data. The reason for creating multiple datasets is that predicting the missing values is associated with uncertainty. The statistical model therefore injects appropriate variability to the multiple imputed values. The true value of the missing data can never be known; it can only be approximated by statistical methods based on the predictive variables. Using standard statistical methods (*e.g.*, Cox regression), a model is fitted for each imputed dataset. The associated risk estimates will differ between datasets, owing to the variation introduced during the imputation of the missing values. The risk estimates are then averaged together, and standard errors, taking account of the variability of the imputed datasets, are calculated using Rubin's rules.[48]

Modeling a multiple imputation analysis requires careful consideration, and failure to do so may result in severe bias.[46] The assumption that data is "missing at random", meaning that any systematic difference between missing values and observed values can be explained by differences in the observed data, needs to be fulfilled in order for the multiple imputation analysis to be valid. This assumption cannot be tested by observing the data directly, but rather approximated by including a relevant and sufficient set of predictive variables in the imputation model. In general, as many predictive variables as possible should be included in the imputation model, and the model should include variables that are correlated with the imputed variable, variables that are associated with the missingness of the imputed variable, and outcome variables.[49] For our analyses we used the multiple imputation by chained equations (MICE) method. This method of multiple imputation is described elsewhere.[49]

Before applying the exclusion criteria, the following covariates in our dataset had missing values:

Covariate	Number of individuals with missing information and percentage of total study population
Smoking status	793 (3.7%)
BMI group	475 (2.2%)
Physical activity	428 (2.0%)
Level of education	567 (2.6%)

As can be seen from the table, all covariates had a very low proportion of missing values, making analysis by multiple imputation suitable. Smoking status had a larger proportion of missing values compared with the other covariates. This might be because smoking status was determined by aggregating questionnaire responses from several questions. We chose not to impute values for consumption of tobacco products in grams per day, as this covariate was conditional on smoking status, giving rise to statistical problems in the imputation model due to perfect prediction.[49]

Although it is difficult to completely rule out whether missing information in one covariate is independent of missingness in another covariate, the majority of individuals with missing information had only a single covariate missing, as illustrated in the following table:

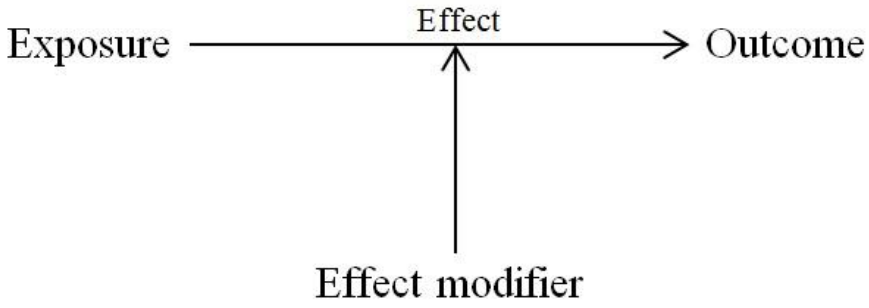
Number of missing values per individual	Number of individuals and percentage of total study population
0	19,680 (91.1%)
1	1,678 (7.8%)
2	177 (0.8%)
3	37 (0.2%)
4	30 (0.1%)

As can be seen from Table 1 and Table 3, individuals with missing information on either smoking status or BMI group were more likely to also have missing information on other lifestyle factors compared with the total study population. However, for individuals with complete information about smoking status or BMI group, the proportion of missing information for any of the other covariates were similar for all three smoking categories and all four BMI groups.

An adequate imputation model should include all variables (both covariates and outcomes) as predicative variables.[49] We therefore imputed the following variables: smoking status, BMI group, physical activity, and level of education, using the following set of predictive variables: age group, sex, smoking status, BMI group, physical activity, level of education, diabetes, hypertension, Charlson comorbidity index score, use of statins, vitamin K antagonists, low-dose aspirin, occurrence of VTE, myocardial infarction, and ischemic stroke.

Effect measure modification

Effect measure modification occurs when a measure of effect changes over values of some other variable (the effect modifier),[50] as illustrated below:



Effect measure modification represents a departure from additivity, that is, the effect of the exposure on the outcome and the effect modificator coinciding is different from the sum of effects expected from the exposure and effect modificator alone. This relationship can be observed when the exposure-outcome relationship differs within different levels of a variable which acts as an effect modifier. Consider the following example: the risk difference for asbestos exposure and risk of lung cancer is 5-1 = 4 cases per 100,000 for non-smokers, while it is 50-10 = 40 cases per 100,000 among smokers. Interestingly, the relative risk of lung cancer after asbestos exposure is identical among nonsmokers and smokers (relative risk = 5). This illustrates an inherent property of effect measure modification; it is dependent on the choice of effect measure.[50]

In our study, we investigated physical activity as a possible effect measure modifier on the association between obesity and VTE. Effect measure modification was present because the risk estimates for a given BMI group differed between subcohorts.

Confounding

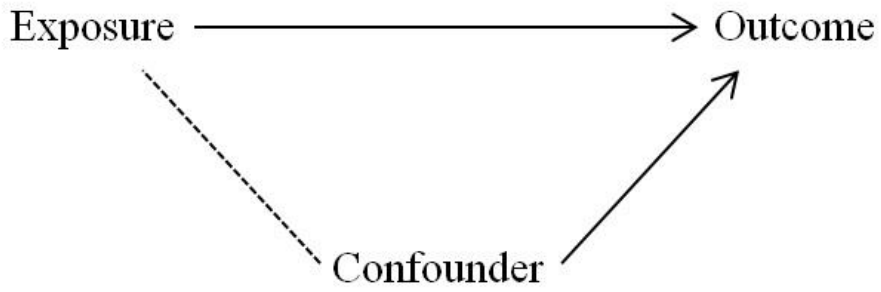
Confounding is a central issue in all study designs in epidemiological research, and may give rise to bias due to a confusion of effects meaning that the effect of the exposure is mixed with the effect of another variable, the confounding variable.[50] Three criteria must be met for a variable to act as a confounder on a given association:

1) The confounding variable is associated with the exposure variable (unevenly distributed across exposure categories).

2) The confounding variable is associated with the outcome, either as a cause or as a proxy for a cause, but not as an effect of the outcome.

3) The confounding variable is not an intermediary step on the causal path between the exposure variable and the outcome.

This relationship can also be expresses graphically:



There are several different forms of confounding: residual confounding, unknown confounding, unmeasured confounding and confounding identified and adjusted for.[50]

Residual confounding arises when a variable is divided into too broad categories. Examples of variables that are broadly defined in our study and may give rise to residual confounding are smoking status and physical activity. For smoking status, never smokers are defined as individuals who have never smoked on a daily basis (neither formerly, nor currently), but who may smoke occasionally (on a weekly or less than weekly basis). Hence, the group of never smokers is comprised of a mixture of individuals who have never smoked in their lives and of individuals who tend to smoke on a less than daily basis. Residual confounding was evident, as the risk estimates for overall VTE were increased for current smokers once we restricted the reference group to only

include individuals who had never smoked at any point in their lives (data not shown). The definition of physical activity used in our study may potentially give rise to some residual confounding, as the variable was defined using a yes/no answer to the following question: "Do you practice sports or conduct other kinds of physical activities on a regular basis in your spare time, that gives you exercise" (translated from Danish). The interpretation of what sort of activities qualify to be classified as sports or physical exercise may vary from individual to individual. Also, the group that responded positively to the question consists of a mixture of individuals ranging from once a week joggers to professional athletes.

One method to avoid residual confounding is to divide variables into more categories, so that the within-category variance is diminished. However, dividing a variable into to many categories may result in too few events within each category and hence imprecise estimates. Because the number of outcomes in our study population was fairly low, we employed a restrictive approach with regards to subdividing variables. In the case of occasional smokers who had never smoked on a daily basis, this group only made up a very little proportion of the study population (1.1%). Therefore, assigning this group to a separate exposure category would lead to few if any outcomes in this category.

Unknown confounding is present when a confounding variable is not identified and hence not adjusted for in the risk estimates. Unmeasured confounding is present when a variable is identified as a confounder, but adjustment is not possible because information on the variable is not available. For many registry based epidemiological studies, unmeasured confounding may arise, as information on lifestyle factors such as smoking status, BMI, physical activity, and level of education is not available within the medical registry, and thus represents a limitation of the study. By restricting our study population to responders to the lifestyle questionnaire "How Are You?", adjustment for a vast array of lifestyle factors was possible, as the dataset consisted of more than 300 variables. On the other hand, restricting our study population to questionnaire responders limited the precision of the estimates compared with previous studies, in which the study population was defined by nationwide medical registries such as the DNRP.[16]

In general, confounding can be addressed by means of study design through restriction, matching or randomizing, or by statistical analysis through stratification, standardization or regression analysis.[50]

External validity

In general, responders to health questionnaires tend to be healthier than individuals who decline to participate. In spite of this, our study population had comparable proportions of both diabetic and hypertensive patients to that of the general Danish population.[51,52] The distribution of BMI groups among questionnaire responders was also similar to what have previously been reported for the Danish population.[53] With regards to the validity of the set of questions which constitute the questionnaire, several well-validated scoring systems were included, such as the CAGE-C score for alcohol addiction,[54] the dietary quality score for diet,[55] and the perceived stress scale for level of day-to-day stress.[56]

Studies in which the study population consists of a specific subsample of the general population are often criticized for lacking external validity, and the reported risk estimates' applicability to the general population may be doubted. However, statistical inference made on the basis of a less representative study population may still be valuable, and a design implementing a non-representative study sample is sometimes a necessity.[57] One example of sound statistical inference being made on a non-representative study population is the famous study by Doll and Hill on the mortality of male British physicians in relation to their smoking habits.[58] The study population did not represent the general population with regards to sex, race, ethnicity, socioeconomic status, and many other variables. The study's conclusion, that smoking increases mortality, was applied to the population in general, and the findings of Doll and Hill have since been reproduced in a variety of different study populations. This is because the biological mechanism of smoking and the resulting elevated mortality does not differ substantially between study populations. In the same way, the biochemical effects of the associations reported in our study would most likely not differ between questionnaire responders and questionnaire decliners. Therefore, we will argue that sound generalizable statistical inference can be made on our findings. As Rothman states: "It is not representativeness of the study subjects that enhances the generalization, it is knowledge of specific conditions and an understanding of the mechanism that makes for a proper generalization."[57]

Pathophysiological explanations

Smoking and risk of VTE

The pathophysiological relationship between smoking and risk of VTE remains controversial. The total accumulated tobacco exposure, measured in pack-years, is associated with the risk of arterial thromboembolic events in a dose-response relationship, in part mediated through atherosclerosis. On the other hand, it has been proposed that smoking has a more acute and transient effect on risk of VTE.[8] This hypothesis may be biologically plausible, as smoking has been shown to increase levels of coagulation factors and inflammatory mediators in the blood, both of which are associated with an elevated risk of VTE.[59,60] This prothrombotic state may be reversible, as other studies have shown that the elevated fibrinogen concentration in smokers decreased quickly to levels similar to that of never smokers upon smoking cessation.[61] Our finding of current smoking being stronger associated with provoked VTE compared with unprovoked VTE supports this theory, as the hypercoaguable and inflammatory state caused by smoking may further challenge a cardiovascular system already compromised by conditions such as pregnancy, surgery, trauma, cancer, and immobilization. The discrepancy between the risk estimates for current and former smokers found in our study may also indicate that the effect of smoking on risk of VTE diminishes after smoking cessation.

Obesity and risk of VTE

As mentioned in the extract, obesity has consistently been cited as a risk factor for VTE. As the world is experiencing an obesity pandemic, with an estimated proportion of obese individuals reaching 41% among US adult and a world-wide prevalence of 700 million by 2015, a refined understanding of the pathophysiological mechanism for the observed association is needed.[62] Body fat, especially abdominal fat, may raise the intra-abdominal pressure and in turn limit venous return from the lower extremities.[62] This can be observed clinically as a decreased blood velocity in the femoral vein, a frequent clinical finding among obese individuals. An understimulated skeletal muscle pump due to physical inactivity in addition to poor gait may further limit venous return.[63] Several biochemical mechanisms have also been suggested to cause the elevated risk of VTE. Leptin, a hormone which acts on the hypothalamus to decrease appetite and food intake, is present in elevated plasma levels in obese individuals, as the central nervous system of these individuals seems to be progressively resistant the its effect.[62] Previous studies have shown that leptin may

induce the transcription of plasminogen activator inhibitor-1 (PAI-1), which inhibits fibrinolysis and thereby maintains a pro-thrombotic state.[64] BMI and waist-to-hip ratio have also been shown to be positively correlated with levels of factor VII, factor VIIIc, fibrinogen, and von Willebrand factor.[62] Furthermore, obesity might lead to increased inflammation, oxidative stress and endothelial dysfunction, all mechanisms proposed elevate the risk of VTE.[62]

VTE and risk of subsequent arterial thromboembolic events

The pathophysiological relationship between venous and arterial thrombosis is difficult to assess, as these disorders have traditionally been viewed as two distinct pathophysiological entities.[65] A venous thrombus is referred to as a red thrombus consisting mainly of red blood cells and fibrin while an arterial thrombus is referred to as a white thrombus consisting mainly of platelets. The first study to propose a possible link between venous and arterial thrombosis was published in 2003 and showed that atherosclerosis was twice as prevalent in patients with unprovoked VTE as in age- and sex-matched controls.[17] Atherosclerosis may therefore represent a common cause of both venous and arterial thrombosis, where the elevated risk of VTE is mediated through increased levels of hemostatic and inflammatory markers present in individuals with atherosclerosis. However, little evidence is available to support this hypothesis.[66] As the risk of arterial thromboembolic events following VTE is highest in the first year after the VTE event,[16] the inflammatory process instigated by the venous thrombus may lead to arterial thrombosis.[66] Later studies on VTE and risk of subsequent arterial thromboembolic events have suggested that the observed association is mediated through shared risk factors for venous and arterial thrombosis.[22] Another study which assessed risk factor profiles for cardiovascular disease within the same study population concluded that coronary artery disease and stroke had broadly comparable risk factor profiles that differed widely from that of VTE.[15] The pathophysiological relationship between venous and arterial thrombosis therefore remains controversial, and further studies on the association is needed.

References

1. 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):585-593.

2. Heit JA, Silverstein MD, Mohr DN, et al. The epidemiology of venous thromboembolism in the community. *Thromb Haemost*. 2001;86(1):452-463.

3. Kniffin WD,Jr, Baron JA, Barrett J, Birkmeyer JD, Anderson FA,Jr. The epidemiology of diagnosed pulmonary embolism and deep venous thrombosis in the elderly. *Arch Intern Med*. 1994;154(8):861-866.

4. Cohen AT, Agnelli G, Anderson FA, et al. Venous thromboembolism (VTE) in europe. the number of VTE events and associated morbidity and mortality. *Thromb Haemost*. 2007;98(4):756-764.

5. Beckman MG, Hooper WC, Critchley SE, Ortel TL. Venous thromboembolism: A public health concern. *Am J Prev Med*. 2010;38(4 Suppl):S495-501. doi: 10.1016/j.amepre.2009.12.017; 10.1016/j.amepre.2009.12.017.

6. Enga KF, Braekkan SK, Hansen-Krone IJ, le Cessie S, Rosendaal FR, Hansen JB. Cigarette smoking and the risk of venous thromboembolism: The tromso study. *J Thromb Haemost*. 2012;10(10):2068-2074. doi: 10.1111/j.1538-7836.2012.04880.x; 10.1111/j.1538-7836.2012.04880.x.

7. Sweetland S, Parkin L, Balkwill A, et al. Smoking, surgery, and venous thromboembolism risk in women: United kingdom cohort study. *Circulation*. 2013;127(12):1276-1282. doi: 10.1161/CIRCULATIONAHA.113.001428; 10.1161/CIRCULATIONAHA.113.001428.

8. 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):1297-1303. doi: 10.1111/j.1538-7836.2009.03490.x; 10.1111/j.1538-7836.2009.03490.x.

9. Glynn RJ, Rosner B. Comparison of risk factors for the competing risks of coronary heart disease, stroke, and venous thromboembolism. *Am J Epidemiol*. 2005;162(10):975-982. doi: 10.1093/aje/kwi309.

10. Braekkan SK, Hald EM, Mathiesen EB, et al. Competing risk of atherosclerotic risk factors for arterial and venous thrombosis in a general population: The tromso study. *Arterioscler Thromb Vasc Biol*. 2012;32(2):487-491. doi: 10.1161/ATVBAHA.111.237545; 10.1161/ATVBAHA.111.237545.

11. Wattanakit K, Lutsey PL, Bell EJ, et al. Association between cardiovascular disease risk factors and occurrence of venous thromboembolism. A time-dependent analysis. *Thromb Haemost*. 2012;108(3):508-515. doi: 10.1160/TH11-10-0726; 10.1160/TH11-10-0726.

12. Holst AG, Jensen G, Prescott E. Risk factors for venous thromboembolism: Results from the copenhagen city heart study. *Circulation*. 2010;121(17):1896-1903. doi: 10.1161/CIRCULATIONAHA.109.921460; 10.1161/CIRCULATIONAHA.109.921460.

13. Severinsen MT, Overvad K, Johnsen SP, et al. Genetic susceptibility, smoking, obesity and risk of venous thromboembolism. *Br J Haematol*. 2010;149(2):273-279. doi: 10.1111/j.1365-2141.2010.08086.x; 10.1111/j.1365-2141.2010.08086.x.

14. 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):1182-1189.

15. Glynn RJ, Rosner B. Comparison of risk factors for the competing risks of coronary heart disease, stroke, and venous thromboembolism. *Am J Epidemiol*. 2005;162(10):975-982. doi: 10.1093/aje/kwi309.

16. Sorensen HT, Horvath-Puho E, Pedersen L, Baron JA, Prandoni P. Venous thromboembolism and subsequent hospitalisation due to acute arterial cardiovascular events: A 20-year cohort study. *Lancet*. 2007;370(9601):1773-1779. doi: 10.1016/S0140-6736(07)61745-0.

17. Prandoni P, Bilora F, Marchiori A, et al. An association between atherosclerosis and venous thrombosis. *N Engl J Med*. 2003;348(15):1435-1441. doi: 10.1056/NEJMoa022157.

18. Prandoni P, Ghirarduzzi A, Prins MH, et al. Venous thromboembolism and the risk of subsequent symptomatic atherosclerosis. *J Thromb Haemost*. 2006;4(9):1891-1896. doi: 10.1111/j.1538-7836.2006.02058.x.

19. Bova C, Marchiori A, Noto A, et al. Incidence of arterial cardiovascular events in patients with idiopathic venous thromboembolism. A retrospective cohort study. *Thromb Haemost*. 2006;96(2):132-136.

20. Klok FA, Mos IC, Broek L, et al. Risk of arterial cardiovascular events in patients after pulmonary embolism. *Blood*. 2009;114(8):1484-1488. doi: 10.1182/blood-2009-05-220491; 10.1182/blood-2009-05-220491.

21. Becattini C, Agnelli G, Prandoni P, et al. A prospective study on cardiovascular events after acute pulmonary embolism. *Eur Heart J*. 2005;26(1):77-83. doi: 10.1093/eurheartj/ehi018.

22. Roach RE, Lijfering WM, Flinterman LE, Rosendaal FR, Cannegieter SC. Increased risk of CVD after VT is determined by common etiologic factors. *Blood*. 2013;121(24):4948-4954. doi: 10.1182/blood-2013-01-479238; 10.1182/blood-2013-01-479238.

23. 4100-08-054folder-21x21.indd - w4100-08-054folder.pdf. http://www.rm.dk/files/Regionshuset/Kommunikationsafdelingen/English/w4100-08-054folder.pdf. Accessed 1/8/2014, 2014.

24. Frank L. Epidemiology. when an entire country is a cohort. *Science*. 2000;287(5462):2398-2399.

25. Pedersen CB. The danish civil registration system. *Scand J Public Health*. 2011;39(7 Suppl):22-25. doi: 10.1177/1403494810387965; 10.1177/1403494810387965.

26. In english - public health and quality improvement, central denmark region- CFK folkesundhed og kvalitetsudvikling. http://www.cfk.rm.dk/om+os/in+english. Accessed 1/8/2014, 2014.

27. Edwards P, Roberts I, Clarke M, et al. Increasing response rates to postal questionnaires: Systematic review. *BMJ*. 2002;324(7347):1183.

28. HHDD profil - revideret 2008.pdf. http://www.cfk.rm.dk/files/Sundhed/CFK/Projekter/HHDD - sundhedsprofil/HVORDAN HAR DU DET_2010/Hjemmesiden/HHDD profil - Revideret 2008.pdf. Accessed 1/8/2014, 2014.

29. Hvordan har du det? 2001-2006- CFK folkesundhed og kvalitetsudvikling. http://www.cfk.rm.dk/projekter/befolkningsundersøgelser/hvordan+har+du+det-c7-/historie/hvordan+har+du+det-c7-+2001-2006. Accessed 1/8/2014, 2014.

30. Lynge E, Sandegaard JL, Rebolj M. The danish national patient register. *Scand J Public Health*. 2011;39(7 Suppl):30-33. doi: 10.1177/1403494811401482; 10.1177/1403494811401482.

31. WHO :: Global database on body mass index. http://apps.who.int/bmi/index.jsp?introPage=intro_3.html. Accessed 1/8/2014, 2014.

32. Johannesdottir SA, Horvath-Puho E, Ehrenstein V, Schmidt M, Pedersen L, Sorensen HT. Existing data sources for clinical epidemiology: The danish national database of reimbursed prescriptions. *Clin Epidemiol*. 2012;4:303-313. doi: 10.2147/CLEP.S37587; 10.2147/CLEP.S37587.

33. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis*. 1987;40(5):373-383.

34. Thygesen SK, Christiansen CF, Christensen S, Lash TL, Sorensen HT. The predictive value of ICD-10 diagnostic coding used to assess charlson comorbidity index conditions in the population-based danish national registry of patients. *BMC Med Res Methodol*. 2011;11:83-2288-11-83. doi: 10.1186/1471-2288-11-83; 10.1186/1471-2288-11-83.

35. Jacobs DR,Jr, Kroenke C, Crow R, et al. PREDICT: A simple risk score for clinical severity and long-term prognosis after hospitalization for acute myocardial infarction or unstable angina: The minnesota heart survey. *Circulation*. 1999;100(6):599-607.

36. Multiple imputation by chained equations (MICE): Implementation in stata - paper. http://www.jstatsoft.org/v45/i04/paper. Accessed 3/7/2014, 2014.

37. Cheng YJ, Liu ZH, Yao FJ, et al. Current and former smoking and risk for venous thromboembolism: A systematic review and meta-analysis. *PLoS Med*. 2013;10(9):e1001515. doi: 10.1371/journal.pmed.1001515; 10.1371/journal.pmed.1001515.

38. Ageno W, Dentali F, Romualdi E, Squizzato A. Cardiovascular disease and venous thromboembolism. different disorders or possible intriguing association? *Recenti Prog Med*. 2008;99(11):572-577.

39. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ*. 1995;311(7003):485.

40. Madsen M, Davidsen M, Rasmussen S, Abildstrom SZ, Osler M. The validity of the diagnosis of acute myocardial infarction in routine statistics: A comparison of mortality and hospital discharge data with the danish MONICA registry. *J Clin Epidemiol*. 2003;56(2):124-130.

41. Johnsen SP, Overvad K, Sorensen HT, Tjonneland A, Husted SE. Predictive value of stroke and transient ischemic attack discharge diagnoses in the danish national registry of patients. *J Clin Epidemiol*. 2002;55(6):602-607.

42. Severinsen MT, Kristensen SR, Overvad K, Dethlefsen C, Tjonneland A, Johnsen SP. Venous thromboembolism discharge diagnoses in the danish national patient registry should be used with

caution. *J Clin Epidemiol*. 2010;63(2):223-228. doi: 10.1016/j.jclinepi.2009.03.018; 10.1016/j.jclinepi.2009.03.018.

43. Visscher TL, Viet AL, Kroesbergen IH, Seidell JC. Underreporting of BMI in adults and its effect on obesity prevalence estimations in the period 1998 to 2001. *Obesity (Silver Spring)*. 2006;14(11):2054-2063. doi: 10.1038/oby.2006.240.

44. Fendrich M, Mackesy-Amiti ME, Johnson TP, Hubbell A, Wislar JS. Tobacco-reporting validity in an epidemiological drug-use survey. *Addict Behav*. 2005;30(1):175-181. doi: 10.1016/j.addbeh.2004.04.009.

45. Borch KH, Hansen-Krone I, Braekkan SK, et al. Physical activity and risk of venous thromboembolism. the tromso study. *Haematologica*. 2010;95(12):2088-2094. doi: 10.3324/haematol.2009.020305 [doi].

46. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ*. 2009;338:b2393. doi: 10.1136/bmj.b2393.

47. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and cox regression. *Am J Epidemiol*. 2007;165(6):710-718. doi: 10.1093/aje/kwk052.

48. Rubin DB. *Multiple imputation for nonresponse in surveys.* New York: Wiley; 1987.

49. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011;30(4):377-399. doi: 10.1002/sim.4067; 10.1002/sim.4067.

50. Rothman KJ. *Epidemiology: An introduction.* 2nd ed. USA: Oxford University Press; 2012.

51. https://www.sundhed.dk/sundhedsfaglig/laegehaandbogen/endokrinologi/tilstande-og-sygdomme/diabetes-mellitus/diabetes-type-2/. Accessed 04/14/ 2014, 2014.

52. Https://Www.sundhed.dk/sundhedsfaglig/laegehaandbogen/hjerte-kar/symptomer-og-tegn/hoejt-blodtryk/. Accessed 04/14/ 2014, 2014.

53. Http://Sundhedsstyrelsen.dk/da/sundhed/overvaegt/forekomst. Accessed 04/14/ 2014,2014.

54. Zierau F, Hardt F, Henriksen JH, et al. Validation of a self-administered modified CAGE test (CAGE-C) in a somatic hospital ward: Comparison with biochemical markers. *Scand J Clin Lab Invest*. 2005;65(7):615-622. doi: PN4861871V771482 [pii].

55. Toft U, Kristoffersen LH, Lau C, Borch-Johnsen K, Jorgensen T. The dietary quality score: Validation and association with cardiovascular risk factors: The Inter99 study. *Eur J Clin Nutr*. 2007;61(2):270-278. doi: 10.1038/sj.ejcn.1602503.

56. Morgan ES, Umberson K, Hertzog C. Construct validation of self-reported stress scales. *Psychol Assess*. 2014;26(1):90-99. doi: 10.1037/a0034714; 10.1037/a0034714.

57. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol*. 2013;42(4):1012-1014. doi: 10.1093/ije/dys223; 10.1093/ije/dys223.

58. DOLL R, HILL AB. The mortality of doctors in relation to their smoking habits; a preliminary report. *Br Med J*. 1954;1(4877):1451-1455.

59. Miller GJ, Bauer KA, Cooper JA, Rosenberg RD. Activation of the coagulant pathway in cigarette smokers. *Thromb Haemost*. 1998;79(3):549-553.

60. Fox EA, Kahn SR. The relationship between inflammation and venous thrombosis. A systematic review of clinical studies. *Thromb Haemost*. 2005;94(2):362-365. doi: 10.1267/THRO05020362.

61. Feher MD, Rampling MW, Brown J, et al. Acute changes in atherogenic and thrombogenic factors with cessation of smoking. *J R Soc Med*. 1990;83(3):146-148.

62. Allman-Farinelli MA. Obesity and venous thrombosis: A review. *Semin Thromb Hemost*. 2011;37(8):903-907. doi: 10.1055/s-0031-1297369; 10.1055/s-0031-1297369.

63. Darvall KA, Sam RC, Silverman SH, Bradbury AW, Adam DJ. Obesity and thrombosis. *Eur J Vasc Endovasc Surg*. 2007;33(2):223-233. doi: S1078-5884(06)00579-X [pii].

64. Singh P, Peterson TE, Barber KR, et al. Leptin upregulates the expression of plasminogen activator inhibitor-1 in human vascular endothelial cells. *Biochem Biophys Res Commun*. 2010;392(1):47-52. doi: 10.1016/j.bbrc.2009.12.158 [doi].

65. Prandoni P. Links between arterial and venous disease. *J Intern Med*. 2007;262(3):341-350. doi: JIM1815 [pii].

66. Lijfering WM, Flinterman LE, Vandenbroucke JP, Rosendaal FR, Cannegieter SC. Relationship between venous and arterial thrombosis: A review of the literature from a causal perspective. *Semin Thromb Hemost*. 2011;37(8):885-896. doi: 10.1055/s-0031-1297367; 10.1055/s-0031-1297367.

Figures

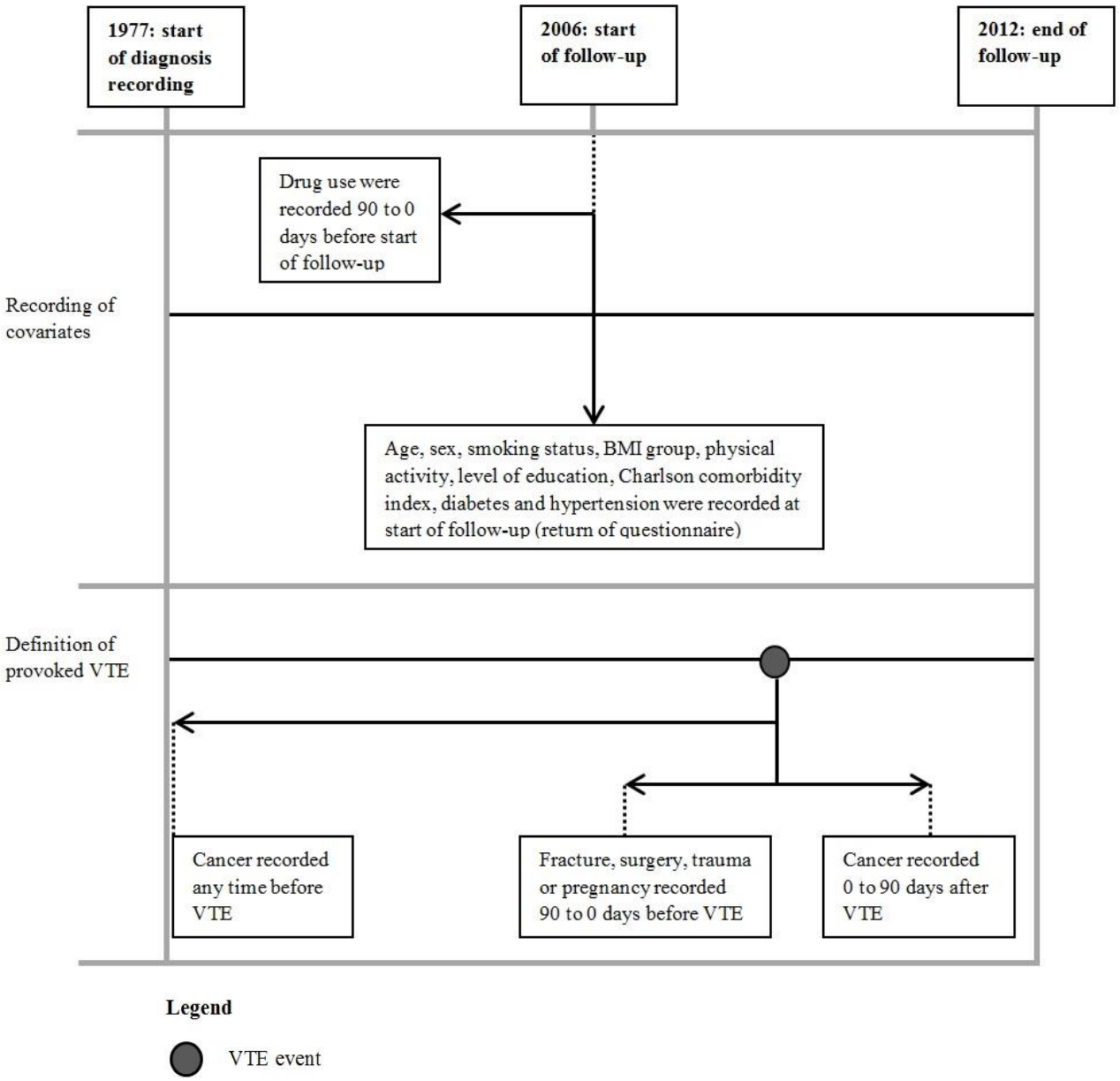


Figure 1. Visualization of the definition and timing of covariates and provoked VTE.

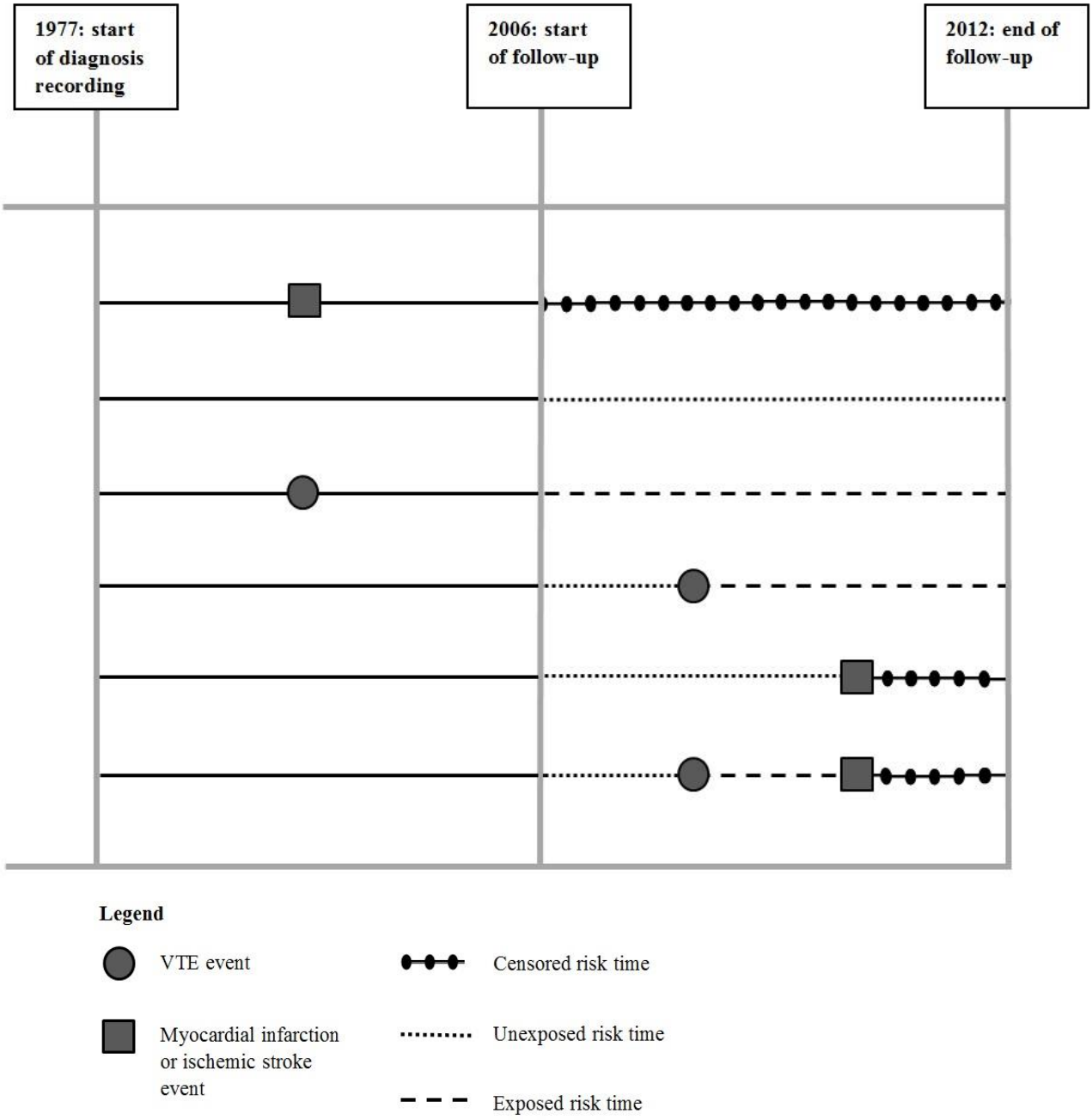


Figure 2. Visualization of the statistical analysis of VTE and risk of subsequent arterial thromboembolic events.

Tables

Table 1. Study population characteristics according to smoking status

	Total study population	Never daily smokers	Former daily smokers	Current daily smokers	Missing information
Sex					
Male	10,024 (46.8)	3,909 (42.5)	3,041 (51.2)	2,815 (51.3)	259 (33.1)
Female	11,378 (53.2)	5,280 (57.5)	2,904 (48.9)	2,670 (48.7)	524 (66.9)
Age					
25-39	5,106 (23.9)	2,996 (32.6)	925 (15.6)	1,121 (20.4)	64 (08.2)
40-49	4,829 (22.6)	2,204 (24.0)	1,224 (20.6)	1,316 (24.0)	85 (10.9)
50-59	4,928 (23.0)	1,852 (20.2)	1,570 (26.4)	1,400 (25.5)	106 (13.5)
60-69	4,189 (19.6)	1,527 (16.6)	1,339 (22.5)	1,113 (20.3)	210 (26.8)
70-79	2,350 (11.0)	610 (06.6)	887 (14.9)	535 (9.8)	318 (40.6)
BMI					
Underweight	306 (01.4)	123 (01.3)	47 (00.8)	127 (02.3)	9 (01.2)
Normal weight	10,078 (47.1)	4,401 (47.9)	2,531 (42.6)	2,862 (52.2)	284 (36.3)
Overweight	7,651 (35.8)	3,213 (35.0)	2,353 (39.6)	1,790 (32.6)	295 (37.7)
Obese	2,903 (13.6)	1,274 (13.9)	914 (15.4)	595 (10.9)	120 (15.3)
Missing information	464 (02.2)	178 (01.9)	100 (01.7)	111 (02.0)	75 (09.6)
Physical activity					
No	11,351 (53.0)	4,262 (46.4)	3,151 (53.0)	3,558 (64.9)	294 (37.6)
Yes	9,626 (45.0)	4,817 (52.4)	2,690 (45.3)	1,825 (33.3)	380 (48.5)
Missing information	425 (02.0)	110 (1.20)	104 (01.8)	102 (01.9)	109 (13.9)
Level of education					
7 years or less	2,752 (12.9)	857 (09.3)	834 (14.0)	751 (13.7)	310 (39.6)
8-10 years	2,889 (13.5)	1,047 (11.4)	655 (11.0)	1,047 (19.1)	140 (17.9)
High school education or vocational training	7,355 (34.4)	3,198 (34.8)	2,050 (34.5)	1,979 (36.1)	128 (16.4)
Higher education	7,848 (36.7)	3,852 (41.9)	2,268 (38.2)	1,599 (29.2)	129 (16.5)
Missing information	558 (02.6)	235 (02.6)	138 (02.3)	109 (02.0)	76 (09.7)
Comorbidity					
Low	17,944 (83.8)	8,229 (89.6)	4,702 (79.1)	4,429 (80.8)	584 (74.6)
Moderate	3,013 (14.1)	873 (09.5)	1,051 (17.7)	918 (16.7)	171 (21.8)
High	445 (02.1)	87 (01.0)	192 (03.2)	138 (02.5)	28 (03.6)
Diabetes	955 (04.5)	321 (03.5)	333 (05.6)	228 (04.2)	73 (09.3)
Hypertension	4,648 (21.7)	1,723 (18.8)	1,565 (26.3)	1,100 (20.1)	260 (33.2)
Statins	1,459 (06.8)	374 (04.1)	595 (10.0)	375 (06.8)	115 (14.7)
Vitamin K antagonists	237 (01.1)	63 (00.7)	85 (01.4)	62 (01.1)	27 (03.5)
Low-dose aspirin	1,295 (06.1)	324 (03.5)	512 (08.6)	342 (06.2)	117 (14.9)
Total	21,402	9,189	5,945	5,485	783

Table 2. Incidence rates and hazard ratios for VTE by smoking status

Outcome	Person-years	Number of outcomes	IR per 1000 PY (95% CI)	Crude HR (95% CI)	Adjusted HR (95% CI)*			
					Model 1	Model 2	Model 3	Model 4
VTE Overall (n=178)								
Never smoker	61,707	58	0.94 (0.73-1.22)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Former smoker	39,148	45	1.15 (0.86-1.54)	1.25 (0.85-1.84)	0.83 (0.56-1.23)	0.77 (0.51-1.14)	0.76 (0.51-1.13)	0.76 (0.51-1.13)
Current smoker	35,829	58	1.62 (1.25-2.09)	1.76 (1.22-2.52)	1.41 (0.98-2.04)	1.35 (0.93-1.96)	1.34 (0.92-1.95)	1.33 (0.91-1.94)
< 20 grams of tobacco/day	19,620	24	1.22 (0.82-1.83)	1.30 (0.81-2.09)	1.04 (0.64-1.68)	1.01 (0.62-1.65)	1.00 (0.61-1.64)	1.00 (0.61-1.63)
≥ 20 grams of tobacco/day	15,785	33	2.09 (1.49-2.94)	2.22 (1.45-3.41)	2.03 (1.30-3.16)	1.91 (1.21-3.01)	1.90 (1.20-2.99)	1.87 (1.19-2.96)
Missing information	4,994	17	3.40 (2.11-5.48)	-	-	-	-	-
VTE Provoked (n=90)								
Never smoker	61,707	25	0.41 (0.27-0.60)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Former smoker	39,148	23	0.59 (0.39-0.88)	1.50 (0.85-2.63)	0.97 (0.55-1.72)	0.83 (0.46-1.48)	0.84 (0.47-1.51)	0.83 (0.47-1.50)
Current smoker	35,829	33	0.92 (0.65-1.30)	2.26 (1.35-3.80)	1.79 (1.06-3.03)	1.60 (0.93-2.74)	1.60 (0.93-2.74)	1.60 (0.93-2.75)
< 20 grams of tobacco/day	19,620	16	0.82 (0.50-1.33)	2.02 (1.08-3.78)	1.59 (0.85-2.99)	1.52 (0.80-2.91)	1.54 (0.81-2.94)	1.54 (0.81-2.95)
≥ 20 grams of tobacco/day	15,785	16	1.01 (0.62-1.65)	2.51 (1.34-4.70)	2.29 (1.20-4.40)	2.11 (1.09-4.11)	2.10 (1.08-4.09)	2.12 (1.08-4.14)
Missing information	4,994	9	1.80 (0.94-3.46)	-	-	-	-	-
VTE Unprovoked (n=88)								
Never smoker	61,707	33	0.53 (0.38-0.75)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Former smoker	39,148	22	0.56 (0.37-0.85)	1.06 (0.62-1.83)	0.72 (0.42-1.25)	0.72 (0.41-1.26)	0.70 (0.40-1.22)	0.70 (0.40-1.23)
Current smoker	35,829	25	0.70 (0.47-1.03)	1.37 (0.80-2.32)	1.11 (0.65-1.90)	1.14 (0.65-1.97)	1.12 (0.65-1.94)	1.09 (0.63-1.89)
< 20 grams of tobacco/day	19,620	8	0.41 (0.20-0.82)	0.76 (0.35-1.64)	0.62 (0.28-1.34)	0.61 (0.28-1.33)	0.59 (0.27-1.30)	0.58 (0.27-1.29)
≥ 20 grams of tobacco/day	15,785	17	1.08 (0.67-1.73)	2.00 (1.12-3.60)	1.82 (0.99-3.35)	1.76 (0.94-3.38)	1.73 (0.93-3.24)	1.68 (0.89-3.14)
Missing information	4,994	8	1.60 (0.80-3.20)	-	-	-	-	-

* Adjusted for the following sets of covariates:
Model 1: age, sex
Model 2: age, sex, BMI group, physical activity, diabetes, hypertension, Charlson comorbidity index score
Model 3: age, sex, BMI group, physical activity, diabetes, hypertension, Charlson comorbidity index score, statins, vitamin K antagonists, low-dose aspirin
Model 4: age, sex, BMI group, physical activity, diabetes, hypertension, Charlson comorbidity index score, statins, vitamin K antagonists, low-dose aspirin, level of education

Table 3. Study population characteristics according to BMI group

	Total study population	Underweight	Normal weight	Overweight	Obese	Missing information
Sex						
Male	10,024 (46.8)	40 (13.1)	3,892 (38.6)	4,526 (59.2)	1,415 (48.7)	151 (32.5)
Female	11,378 (53.2)	266 (86.9)	6,186 (61.4)	3,125 (40.8)	1,488 (51.3)	313 (67.5)
Age						
25-39	5,106 (23.9)	88 (28.8)	2,703 (26.8)	1,589 (20.8)	639 (22.0)	87 (18.8)
40-49	4,829 (22.6)	46 (15.0)	2,450 (24.3)	1,643 (21.5)	613 (21.1)	77 (16.6)
50-59	4,928 (23.0)	66 (21.6)	2,224 (22.1)	1,862 (24.3)	701 (24.2)	75 (16.2)
60-69	4,189 (19.6)	54 (17.7)	1,717 (17.0)	1,685 (22.0)	636 (21.9)	97 (20.9)
70-79	2,350 (11.0)	52 (17.0)	984 (09.8)	872 (11.4)	314 (10.8)	128 (27.6)
Smoking status						
Never smoker	9,189 (42.9)	123 (40.2)	4,401 (43.7)	3,213 (42.0)	1,274 (43.9)	178 (38.4)
Former smoker	5,945 (27.8)	47 (15.4)	2,531 (25.1)	2,353 (30,8)	914 (31.5)	100 (21.6)
Current smoker	5,485 (25.6)	127 (41.5)	2,862 (28.4)	1,790 (23.4)	595 (20.5)	111 (23.9)
Missing information	783 (03.7)	9 (02.9)	284 (02.8)	295 (03.9)	120 (04.1)	75 (16.2)
Physical activity						
No	11,351 (53.0)	201 (65.7)	4,915 (48.8)	4,127 (53.9)	1,868 (64.4)	240 (51.7)
Yes	9,626 (45.0)	95 (31.1)	5,011 (49.7)	3,393 (44.4)	980 (33.8)	147 (31.7)
Missing information	425 (02.0)	10 (03.3)	152 (01.5)	131 (01.7)	55 (01.9)	77 (16.6)
Level of education						
7 years or less	2,752 (12.9)	44 (14.4)	1,027 (10.2)	1,054 (13.8)	511 (17.6)	116 (25.0)
8-10 years	2,889 (13.5)	62 (20.3)	1,240 (12.3)	1,053 (13.8)	468 (16.1)	66 (14.2)
High school education or vocational training	7,355 (34.4)	85 (27.8)	3,336 (33.1)	2,796 (36.5)	1,019 (35.1)	119 (25.7)
Higher education	7,848 (36.7)	106 (34.6)	4,262 (42.3)	2,557 (33.4)	821 (28.3)	102 (22.0)
Missing information	558 (02.6)	9 (02.9)	213 (02.1)	191 (02.5)	84 (02.9)	61 (13.2)
Comorbidity						
Low	17,502 (81.8)	255 (73.5)	8,483 (84.2)	6,279 (82.1)	2,180 (75.1)	335 (72.2)
Moderate	3,335 (15.6)	66 (21.6)	1,374 (13.6)	1,170 (15.3)	617 (21.3)	108 (23.3)
High	565 (02.6)	15 (04.9)	221 (02.2)	202 (02.6)	106 (03.7)	21 (04.5)
Diabetes	955 (04.5)	10 (03.3)	250 (02.5)	349 (04.6)	305 (10.5)	41 (08.8)
Hypertension	4,648 (21.7)	35 (11.4)	1,529 (15.2)	1,848 (24.2)	1,114 (38.4)	122 (26.3)
Statins	1,459 (06.8)	20 (06.5)	460 (04.6)	612 (08.0)	327 (11.3)	40 (08.6)
Vitamin K antagonists	237 (01.1)	2 (00.7)	84 (00.8)	80 (01.1)	62 (02.1)	9 (01.9)
Low-dose aspirin	1,295 (06.1)	18 (05.9)	421 (04.2)	502 (06.6)	293 (10.1)	61 (13.2)
Total	21,402	306	10,078	7,651	2,903	464

Table 4. Incidence rates and hazard ratios for VTE by BMI group

Outcome	Person-years	Number of outcomes	IR per 1000 PY (95% CI)	Crude HR (95% CI)	Adjusted HR (95% CI)*			
					Model 1	Model 2	Model 3	Model 4
VTE Overall (n=178)								
Normal weight	66,889	64	0.96 (0.75-1.22)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Underweight	1,888	3	1.59 (0.51-4.93)	1.65 (0.52-5.26)	1.60 (0.50-5.12)	1.32 (0.41-4.24)	1.30 (0.40-4.17)	1.27 (0.39-4.08)
Overweight	50,808	65	1.28 (1.00-1.63)	1.34 (0.95-1.89)	1.10 (0.77-1.56)	1.11 (0.78-1.58)	1.11 (0.78-1.58)	1.10 (0.77-1.57)
Obesity	19,148	41	2.14 (1.58-2.91)	2.26 (1.53-3.35)	1.95 (1.31-2.89)	1.84 (1.22-2.77)	1.85 (1.23-2.78)	1.83 (1.21-2.77)
Missing information	2,945	5	1.70 (0.71-4.08)	-	-	-	-	-
VTE Provoked (n=90)								
Normal weight	66,889	32	0.48 (0.34-0.68)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Underweight	1,888	2	1.06 (0.27-4.24)	2.19 (0.53-9.16)	2.15 (0.51-9.08)	1.54 (0.36-6.57)	1.48 (0.35-6.32)	1.49 (0.35-6.34)
Overweight	50,808	34	0.67 (0.48-0.94)	1.39 (0.85-2.25)	1.11 (0.68-1.81)	1.14 (0.70-1.88)	1.15 (0.70-1.89)	1.17 (0.71-1.92)
Obesity	19,148	20	1.04 (0.67-1.62)	2.19 (1.25-3.82)	1.84 (1.05-3.22)	1.77 (0.99-3.16)	1.79 (1.00-3.20)	1.83 (1.02-3.27)
Missing information	2,945	2	0.68 (0.17-2.72)	-	-	-	-	-
VTE Unprovoked (n=88)								
Normal weight	66,889	32	0.48 (0.34-0.68)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Underweight	1,888	1	0.53 (0.07-3.76)	1.10 (0.15-8.08)	1.05 (0.14-7.75)	0.99 (0.13-7.35)	0.99 (0.13-7.32)	0.95 (0.13-7.08)
Overweight	50,808	31	0.61 (0.43-0.87)	1.29 (0.79-2.12)	1.09 (0.66-1.79)	1.07 (0.65-1.77)	1.06 (0.64-1.76)	1.04 (0.63-1.72)
Obesity	19,148	21	1.10 (0.72-1.68)	2.34 (1.35-4.07)	2.06 (1.18-3.59)	1.89 (1.06-3.37)	1.89 (1.06-3.38)	1.81 (1.01-3.24)
Missing information	2,945	3	1.02 (0.33-3.16)	-	-	-	-	-

* Adjusted for the following sets of covariates:
Model 1: age, sex
Model 2: age, sex, smoking status, physical activity, hypertension, Charlson comorbidity index score
Model 3: age, sex, smoking status, physical activity, hypertension, Charlson comorbidity index score, statins, vitamin K antagonists, low-dose aspirin
Model 4: age, sex, smoking status, physical activity, hypertension, Charlson comorbidity index score, statins, vitamin K antagonists, low-dose aspirin, level of education

Table 5. Study population characteristics according to VTE status at start of follow-up

	Total study population	No VTE before start of follow-up	VTE before start of follow-up
Sex			
Male	9,654 (46.2)	9,580 (46.2%)	74 (41.6%)
Female	11,266 (53.9)	11,162 (53.8%)	104 (58.4%)
Age			
25-39	5,122 (24.5)	5,103 (24.6)	19 (10.7)
40-49	4,807 (23.0)	4,790 (23.1)	17 (09.6)
50-59	4,866 (23.3)	4,822 (23.3)	44 (24.7)
60-69	4,004 (19.1)	3,942 (19.0)	62 (34.8)
70-79	2,121 (10.1)	2,085 (10.1)	36 (20.2)
Smoking			
Never smoker	9,135 (43.7)	9,077 (43.8)	58 (32.6)
Former smoker	5,703 (27.3)	5,648 (27.2)	55 (30.9)
Current smoker	5,343 (25.5)	5,285 (25.5)	58 (32.6)
Missing information	739 (03.5)	732 (03.5)	7 (03.9)
BMI			
Underweight	301 (01.4)	295 (01.4)	6 (03.4)
Normal weight	9,918 (47.4)	9,860 (47.5)	58 (32.6)
Overweight	7,430 (35.5)	7,365 (35.5)	65 (36.5)
Obese	2,825 (13.5)	2,786 (13.4)	39 (21.9)
Missing information	446 (02.1)	436 (02.1)	10 (05.6)
Physical activity			
No	11,048 (52.8)	10,929 (52.7)	119 (66.9)
Yes	9,475 (45.3)	9,418 (45.4)	57 (32.0)
Missing information	397 (01.9)	395 (1.9)	2 (01.1)
Level of education			
7 years or less	2,576 (12.3)	2,532 (12.2)	44 (24.7)
8-10 years	2,839 (13.6)	2,814 (13.6)	25 (14.0)
High school education or vocational training	7,191 (34.4)	7,133 (34.4)	58 (32.6)
Higher education	7,769 (37.1)	7,724 (37.2)	45 (25.3)
Missing information	545 (02.6)	539 (02.6)	6 (03.4)
Comorbidity			
Low	18,275 (87.4)	18,161 (87.6)	114 (64.0)
Moderate	2,343 (11.2)	2,296 (11.1)	47 (26.4)
High	302 (01.4)	285 (01.4)	17 (09.6)
Diabetes	871 (04.2)	850 (04.1)	21 (11.8)
Hypertension	4,298 (20.5)	4,245 (20.5)	53 (29.8)
Statins	1,124 (05.4)	1,101 (05,3)	23 (12.9)
Vitamin K antagonists	227 (01.1)	907 (04.4)	30 (16.9)
Low-dose aspirin	937 (04.5)	197 (01.0)	30 (16.9)
Total	20,920	20,742	178

Table 6. Incidence rates and hazard ratios for myocardial infarction and ischemic stroke for individuals with and without VTE

Outcome	Person-years	Number of outcomes	IR per 1000 PY (95% CI)	Crude HR (95% CI)	Adjusted HR (95% CI)*		
					Model 1	Model 2	Model 3
Myocardial infarction (n=262)							
No VTE	136,931	259	1.89 (1.67-2.14)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
VTE	1,513	3	1.98 (0.64-6.15)	1.05 (0.34-3.27)	0.64 (0.21-2.01)	0.60 (0.19-1.87)	0.60 (0.19-1.90)
Provoked VTE	446	2	4.48 (1.12-17.93)	2.37 (0.59-9.51)	1.33 (0.33-5.36)	1.25 (0.31-5.06)	1.30 (0.32-5.27)
Unprovoked VTE	1,067	1	0.94 (0.13-6.66)	0.49 (0.07-3.53)	0.32 (0.04-2.26)	0.29 (0.04-2.09)	0.29 (0.04-2.09)
Ischemic stroke (n=288)							
No VTE	136,922	282	2.06 (1.83-2.31)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
VTE	1,498	6	4.00 (1.80-8.91)	1.93 (0.86-4.33)	1.14 (0.51-2.56)	1.08 (0.48-2.45)	1.09 (0.48-2.47)
Provoked VTE	444	1	2.25 (0.32-16.00)	1.08 (0.15-7.71)	0.61 (0.09-4.33)	0.53 (0.07-3.82)	0.54 (0.07-3.84)
Unprovoked VTE	1,055	5	4.74 (1.97-11.39)	2.28 (0.94-5.53)	1.38 (0.57-3.35)	1.36 (0.56-3.32)	1.37 (0.56-3.35)
Myocardial infarction or ischemic stroke (n=529)							
No VTE	136,182	520	3.82 (3.50-4.16)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
VTE	1,481	9	6.08 (3.16-11.68)	1.59 (0.82-3.07)	0.96 (0.50-1.86)	0.90 (0.47-1.75)	0.91 (0.47-1.77)
Provoked VTE	433	3	6.92 (2.23-21.46)	1.81 (0.58-5.62)	1.01 (0.33-3.16)	0.93 (0.30-2.89)	0.94 (0.30-2.93)
Unprovoked VTE	1,047	6	5.73 (2.57-12.75)	1.50 (0.67-3.35)	0.94 (0.42-2.09)	0.89 (0.40-2.00)	0.90 (0.40-2.02)

* Adjusted for the following sets of covariates:
Model 1: age, sex
Model 2: age, sex, smoking status, BMI group, diabetes, hypertension, Charlson comorbidity index score
Model 3: age, sex, smoking status, BMI group, diabetes, hypertension, Charlson comorbidity index score, physical activity, level of education, statins, vitamin K antagonists, low-dose aspirin

Supplementary table 1. Incidence rates and hazard ratios for VTE by BMI group among physically active individuals (n=9,626)

Outcome	Person-years	Number of outcomes	IR per 1000 PY (95% CI)	Crude HR (95% CI)	Adjusted HR (95% CI)*			
					Model 1	Model 2	Model 3	Model 4
VTE Overall (n=54)								
Normal weight	33,656	19	0.56 (0.36-0.89)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Underweight	609	0	-	-	-	-	-	-
Overweight	22,747	21	0.92 (0.60-1.42)	1.64 (0.88-3.05)	1.32 (0.70-2.49)	1.41 (0.75-2.66)	1.42 (0.75-2.68)	1.41 (0.74-2.66)
Obesity	6,556	13	1.98 (1.15-3.42)	3.59 (1.77-7.26)	3.24 (1.60-6.57)	3.30 (1.59-6.84)	3.33 (1.61-6.90)	3.32 (1.59-6.92)
Missing information	961	1	1.04 (0.15-7.38)	-	-	-	-	-
VTE Provoked (n=28)								
Normal weight	33,656	10	0.30 (0.16-0.55)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Underweight	609	0	-	-	-	-	-	-
Overweight	22,747	10	0.44 (0.24-0.82)	1.49 (0.62-3.58)	1.13 (0.46-2.76)	1.22 (0.50-2.98)	1.25 (0.51-3.06)	1.25 (0.51-3.06)
Obesity	6,556	7	1.07 (0.51-2.24)	3.73 (1.42-9.78)	3.26 (1.24-8.60)	3.03 (1.12-8.22)	2.99 (1.10-8.14)	3.15 (1.16-8.57)
Missing information	961	1	1.04 (0.15-7.38)	-	-	-	-	-
VTE Unprovoked (n=26)								
Normal weight	33,656	9	0.27 (0.14-0.52)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Underweight	609	0	-	-	-	-	-	-
Overweight	22,747	11	0.48 (0.27-0.87)	1.81 (0.75-4.37)	1.57 (0.64-3.86)	1.66 (0.67-4.10)	1.65 (0.67-4.07)	1.58 (0.64-3.91)
Obesity	6,556	6	0.92 (0.41-2.04)	3.42 (1.22-9.62)	3.20 (1.14-9.04)	3.66 (1.26-10.60)	3.64 (1.25-10.58)	3.34 (1.14-9.78)
Missing information	961	0	-	-	-	-	-	-

* Adjusted for the following sets of covariates:
Model 1: age, sex
Model 2: age, sex, smoking status, hypertension, Charlson comorbidity index score
Model 3: age, sex, smoking status, hypertension, Charlson comorbidity index score, statins, vitamin K antagonists, low-dose aspirin
Model 4: age, sex, smoking status, hypertension, Charlson comorbidity index score, statins, vitamin K antagonists, low-dose aspirin, level of education

Supplementary table 2. Incidence rates and hazard ratios for VTE by BMI group among physically inactive individuals (n=11,351)

Outcome	Person-years	Number of outcomes	IR per 1000 PY (95% CI)	Crude HR (95% CI)	Adjusted HR (95% CI)*			
					Model 1	Model 2	Model 3	Model 4
VTE Overall (n=118)								
Normal weight	32,283	43	1.33 (0.99-1.80)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Underweight	1,220	3	2.46 (0.79-7.63)	1.81 (0.56-5.84)	1.79 (0.55-5.82)	1.71 (0.52-5.60)	1.68 (0.51-5.48)	1.62 (0.49-5.30)
Overweight	27,218	41	1.51 (1.11-2.05)	1.13 (0.74-1.73)	0.97 (0.63-1.49)	0.98 (0.63-1.52)	0.97 (0.63-1.51)	0.97 (0.63-1.51)
Obesity	12,252	28	2.29 (1.58-3.31)	1.72 (1.07-2.77)	1.53 (0.95-2.48)	1.51 (0.92-2.47)	1.51 (0.92-2.47)	1.50 (0.91-2.47)
Missing information	1,521	3	1.97 (0.64-6.12)	-	-	-	-	-
VTE Provoked (n=62)								
Normal weight	32,283	22	0.68 (0.45-1.04)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Underweight	1,220	2	1.64 (0.41-6.56)	2.39 (0.56-10.18)	2.37 (0.55-10.21)	1.97 (0.45-8.57)	1.84 (0.42-8.04)	1.85 (0.42-8.05)
Overweight	27,218	24	0.88 (0.59-1.32)	1.28 (0.72-2.28)	1.07 (0.60-1.92)	1.13 (0.63-2.04)	1.13 (0.62-2.04)	1.13 (0.62-2.05)
Obesity	12,252	13	1.06 (0.62-1.83)	1.53 (0.77-3.04)	1.33 (0.67-2.65)	1.41 (0.69-2.88)	1.41 (0.69-2.88)	1.41 (0.69-2.90)
Missing information	1,521	1	0.66 (0.09-4.67)	-	-	-	-	-
VTE Unprovoked (n=56)								
Normal weight	32,283	21	0.65 (0.42-1.00)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Underweight	1,220	1	0.82 (0.12-5.82)	1.21 (0.16-8.98)	1.19 (0.16-8.93)	1.31 (0.17-9.88)	1.33 (0.18-10.03)	1.25 (0.17-9.45)
Overweight	27,218	17	0.62 (0.39-1.00)	0.98 (0.52-1.85)	0.86 (0.45-1.63)	0.83 (0.43-1.58)	0.82 (0.43-1.57)	0.82 (0.43-1.57)
Obesity	12,252	15	1.22 (0.74-2.03)	1.91 (0.98-3.72)	1.76 (0.90-3.42)	1.58 (0.79-3.16)	1.58 (0.79-3.16)	1.55 (0.77-3.13)
Missing information	1,521	2	1.31 (0.33-5.26)	-	-	-	-	-

* Adjusted for the following sets of covariates:
Model 1: age, sex
Model 2: age, sex, smoking status, hypertension, Charlson comorbidity index score
Model 3: age, sex, smoking status, hypertension, Charlson comorbidity index score, statins, vitamin K antagonists, low-dose aspirin
Model 4: age, sex, smoking status, hypertension, Charlson comorbidity index score, statins, vitamin K antagonists, low-dose aspirin, level of education

Appendix

Appendix 1. Definition of cardiovascular events

	ICD-8	ICD-10	**Danish classification of surgical procedures and therapies** (1977 - 1995)	**NOMESCO classification of surgical procedures** (1995-present)
Myocardial infarction	410	I21		
Ischemic stroke	433-434	I63-I64		
Venous thromboembolism	451.00, 450.99	I80.1-3, I26		
Deep venous thrombosis	451.00	I80.1-3		
Pulmonary embolism	450.99	I26		
Provoked VTE*				
Cancer	140–209	C00–C99, B21		
Pregnancy or delivery	630–680	O00–O99		
Fracture or trauma	800–929, 950-959	S00–T14		
Surgery			01000-90860	KA-KQ

* Provoked VTE is defined as VTE coinciding with one or more states of elevated VTE risk within a set timeframe: diagnosis of fracture, surgery, trauma, or pregnancy within 90 days before VTE diagnosis, or a diagnosis of malignancy any time prior to or within 90 days after the VTE diagnosis. Unprovoked VTE was defined as all VTE cases that did not fit the definition of provoked VTE.

Appendix 2. Definition of covariates (source of data in parentheses)		
Smoking status (HAY)		
	Never daily smoker	Q. 17: ”Nej, jeg ryger ikke”, ”Ja, mindst én gang om ugen” or ”Ja, men sjældnere end hver uge” AND Q. 24: ”Nej, jeg har aldrig røget hver dag.”
	Former daily smoker	Q. 24: ”Ja, jeg har tidligere røget hver dag”
	Current daily smoker	Q. 17: ”Ja, hver dag”
BMI group (HAY)		
	Underweight	BMI < 18.5
	Normal weight	25 ≤ BMI < 30
	Overweight	BMI ≤ 30
	Obese	BMI ≤ 30
	Weight: Q. 54, given in whole kilograms. Height: Q. 53, given in whole centimeters.	
Physical activity (HAY)		
	Regular exercise	Q. 64: ”Ja”
	No regular exercise	Q. 64: ”Nej”
Level of education (HAY)		
	≤7 years	Q. 83: ”7 års skolegang eller mindre”
	8-10 years	Q. 83: ”8-9 års skolegang” eller ”10 års skolegang”
	High school education, vocational training or equivalent	Q. 83: ”Studentereksamen, HF, HH, HTX eller tilsvarende” or Q. 84: ”Faglært indenfor håndværk, handel, kontor m.v. (lærlinge- eller EFG-uddannelse)
	Higher education	Q. 84: ”Kort videregående uddannelse, under 3 år (fx social og sundhedsassistent, politibetjent, tekniker, merkonom) or ”Mellemlang videregående uddannelse, 3-4 år (fx folkeskolelærer, journalist, socialrådgiver, fysioterapeut)” or ”Lang videregående uddannelse, mere end 4 år (fx civilingeniør, cand.mag., læge, psykolog)”
Diabetes (DNRP, DNDRP, HAY)		
	Prevalence of diabetes	ICD-8: 249, 250 ICD-10: E10, E11, O24 (except O24.4) ATC: A10A, A10B HAY: Q. 14: Sukkersyge (diabetes): “Ja, det har jeg nu” OR “Ja, har haft det tidligere.”
Hypertension (DNRP, HAY)		
	Prevalence of hypertension	ICD-8: 400-404 ICD-10: I10-I15 HAY: Q. 14: Forhøjet blodtryk: “Ja, det har jeg nu” OR “Ja, har haft det tidligere.”
Statins (DNDRP)		
	Use of statins	ATC: C10AA
Vitamin K antagonists (DNDRP)		
	Use of vitamin K antagonists	ATC: B01AA03, B01AA04
Low-dose aspirin (DNDRP)		
	Use of low-dose aspirin	ATC: B01AC06 or N02BA01 (N02BA01 restricted to drug identification codes: 17048, 19019, 24503, 24514, 24618, 39178, 39187, 44642, 50526, 65052, 95828, 152231, 432682, 459065, 459479, 459883, 506485, 519371)

Appendix 3. Carlson comorbidity index (DNRP, DNDRP)	
Index score 1	Congestive heart failure: ICD-8: 427.09; 427.10; 427.11; 427.19; 428.99; 782.49. ICD-10: I50; I11.0; I13.0; I13.2.
	Peripheral vascular disease: ICD-8: 440; 441; 442; 443; 444; 445. ICD-10: I70; I71; I72; I73; I74; I77.
	Dementia: ICD-8: 290.09-290.19; 293.09. ICD-10: F00-F03; F05.1; G30.
	Chronic pulmonary disease: ICD-8: 490-493; 515-518. ICD-10: J40-J47; J60-J67; J68.4; J70.1; J70.3; J84.1; J92.0; J96.1; J98.2; J98.3. ATC: R03
	Rheumatoid arthritis and Connective tissue disease: ICD-8: 712; 716; 734; 446; 135.99. ICD-10: M05; M06; M08; M09; M30; M31; M32; M33; M34; M35; M36; D86.
	Ulcer disease: ICD-8: 530.91; 530.98; 531-534. ICD-10: K22.1; K25-K28
	Mild liver disease: ICD-8: 571; 573.01; 573.04. ICD-10: B18; K70.0-K70.3; K70.9; K71; K73; K74; K76.0.
	Diabetes mellitus: ICD-8: 249.00; 249.06; 249.07; 249.09; 250.00; 250.06; 250.07; 250.09 ICD-10: E10.0; E10.1; E10.9; E11.0; E11.1; E11.9; O24 (except O24.4). ATC: A10A, A10B
Index score 2	Diabetes with end organ damage: ICD-8: 249.01-249.05; 249.8; 250.01-250.05; 250.08 ICD-10: E10.2-E10.8; E11.2-E11.8
	Moderate to severe renal disease: ICD-8: 403; 404; 580-583; 584; 590.09; 593.19; 753.10-753.19; 792. ICD-10: I12; I13; N00-N05; N07; N11; N14; N17-N19; Q61.
	Non-metastatic cancer: ICD-8: 140-194. ICD-10: C00-C75.
	Leukemia: ICD-8: 204-207. ICD-10: C91-C95.
	Lymphoma: ICD-8: 200-203; 275.59. ICD-10: C81-C85; C88; C90; C96.
Index score 3	Moderate to severe liver disease: ICD-8: 070.00; 070.02; 070.04; 070.06; 070.08; 573.00; 456.00-456.09. ICD-10: B15.0; B16.0; B16.2; B19.0; K70.4; K72; K76.6; I85.
Index score 6	Metastatic solid tumor: ICD-8: 195-198; 199. ICD-10: C76-C80.
	AIDS: ICD-8: 079.83. ICD-10: B21-B24.

Reports/PhD theses from Department of Clinical Epidemiology

1. Ane Marie Thulstrup: Mortality, infections and operative risk in patients with liver cirrhosis in Denmark. Clinical epidemiological studies. *2000*.

2. Nana Thrane: Prescription of systemic antibiotics for Danish children. *2000*.

3. Charlotte Søndergaard. Follow-up studies of prenatal, perinatal and postnatal risk factors in infantile colic. *2001*.

4. Charlotte Olesen: Use of the North Jutland Prescription Database in epidemiological studies of drug use and drug safety during pregnancy. *2001*.

5. Yuan Wei: The impact of fetal growth on the subsequent risk of infectious disease and asthma in childhood. *2001*.

6. Gitte Pedersen. Bacteremia: treatment and prognosis. *2001*.

7. Henrik Gregersen: The prognosis of Danish patients with monoclonal gammopathy of undertermined significance: register-based studies. *2002*.

8. Bente Nørgård: Colitis ulcerosa, coeliaki og graviditet; en oversigt med speciel reference til forløb og sikkerhed af medicinsk behandling. *2002*.

9. Søren Paaske Johnsen: Risk factors for stroke with special reference to diet, Chlamydia pneumoniae, infection, and use of non-steroidal anti-inflammatory drugs. *2002*.

10. Elise Snitker Jensen: Seasonal variation of meningococcal disease and factors associated with its outcome. *2003*.

11. Andrea Floyd: Drug-associated acute pancreatitis. Clinical epidemiological studies of selected drugs. *2004*.

12. Pia Wogelius: Aspects of dental health in children with asthma. Epidemiological studies of dental anxiety and caries among children in North Jutland County, Denmark. *2004*.

13. Kort-og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme i Nordjyllands, Viborg og Århus amter 1985-2003. *2004*.

14. Reimar W. Thomsen: Diabetes mellitus and community-acquired bacteremia: risk and prognosis. *2004*.

15. Kronisk obstruktiv lungesygdom i Nordjyllands, Viborg og Århus amter 1994-2004. Forekomst og prognose. Et pilotprojekt. *2005*.

16. Lungebetændelse i Nordjyllands, Viborg og Århus amter 1994-2004. Forekomst og prognose. Et pilotprojekt. *2005*.

17. Kort- og langtidsoverlevelse efter indlæggelse for nyre-, bugspytkirtel- og leverkræft i Nordjyllands, Viborg, Ringkøbing og Århus amter 1985-2004. *2005*.

18. Kort- og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme i Nordjyllands, Viborg, Ringkøbing og Århus amter 1995-2005. *2005*.

19. Mette Nørgaard: Haematological malignancies: Risk and prognosis. *2006*.

20. Alma Becic Pedersen: Studies based on the Danish Hip Arthroplastry Registry. *2006*.

    Særtryk: Klinisk Epidemiologisk Afdeling - De første 5 år. *2006*.

21. Blindtarmsbetændelse i Vejle, Ringkjøbing, Viborg, Nordjyllands og Århus Amter. *2006*.

22. Andre sygdommes betydning for overlevelse efter indlæggelse for seks kræftsygdomme i Nordjyllands, Viborg, Ringkjøbing og Århus amter 1995-2005. *2006*.

23. Ambulante besøg og indlæggelser for udvalgte kroniske sygdomme på somatiske hospitaler i Århus, Ringkjøbing, Viborg, og Nordjyllands amter. *2006*.

24. Ellen M Mikkelsen: Impact of genetic counseling for hereditary breast and ovarian cancer disposition on psychosocial outcomes and risk perception: A population-based follow-up study. *2006*.

25. Forbruget af lægemidler mod kroniske sygdomme i Århus, Viborg og Nordjyllands amter 2004-2005. *2006*.

26. Tilbagelægning af kolostomi og ileostomi i Vejle, Ringkjøbing, Viborg, Nordjyllands og Århus Amter. *2006*.

27. Rune Erichsen: Time trend in incidence and prognosis of primary liver cancer and liver cancer of unknown origin in a Danish region, 1985-2004. *2007*.

28. Vivian Langagergaard: Birth outcome in Danish women with breast cancer, cutaneous malignant melanoma, and Hodgkin's disease. *2007*.

29. Cynthia de Luise: The relationship between chronic obstructive pulmonary disease, comorbidity and mortality following hip fracture. *2007*.

30. Kirstine Kobberøe Søgaard: Risk of venous thromboembolism in patients with liver disease: A nationwide population-based case-control study. *2007*.

31. Kort- og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme i Region Midtjylland og Region Nordjylland 1995-2006. *2007*.

32. Mette Skytte Tetsche: Prognosis for ovarian cancer in Denmark 1980-2005: Studies of use of hospital discharge data to monitor and study prognosis and impact of comorbidity and venous thromboembolism on survival. *2007*.

33. Estrid Muff Munk: Clinical epidemiological studies in patients with unexplained chest and/or epigastric pain. *2007*.

34. Sygehuskontakter og lægemiddelforbrug for udvalgte kroniske sygdomme i Region Nordjylland. *2007*.

35. Vera Ehrenstein: Association of Apgar score and postterm delivery with neurologic morbidity: Cohort studies using data from Danish population registries. *2007*.

36. Annette Østergaard Jensen: Chronic diseases and non-melanoma skin cancer. The impact on risk and prognosis. *2008*.

37. Use of medical databases in clinical epidemiology. *2008*.

38. Majken Karoline Jensen: Genetic variation related to high-density lipoprotein metabolism and risk of coronary heart disease. *2008*.

39. Blodprop i hjertet - forekomst og prognose. En undersøgelse af førstegangsindlæggelser i Region Nordjylland og Region Midtjylland. *2008*.

40. Asbestose og kræft i lungehinderne. Danmark 1977-2005. *2008*.

41. Kort- og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme i Region Midtjylland og Region Nordjylland 1996-2007. *2008*.

42. Akutte indlæggelsesforløb og skadestuebesøg på hospiter i Region Midtjylland og Region Nordjylland 2003-2007. Et pilotprojekt. *Not published*.

43. Peter Jepsen: Prognosis for Danish patients with liver cirrhosis. *2009*.

44. Lars Pedersen: Use of Danish health registries to study drug-induced birth defects – A review with special reference to methodological issues and maternal use of non-steroidal anti-inflammatory drugs and Loratadine. *2009*.

45. Steffen Christensen: Prognosis of Danish patients in intensive care. Clinical epidemiological studies on the impact of preadmission cardiovascular drug use on mortality. *2009*.

46. Morten Schmidt: Use of selective cyclooxygenase-2 inhibitors and nonselective nonsteroidal antiinflammatory drugs and risk of cardiovascular events and death after intracoronary stenting. *2009*.

47. Jette Bromman Kornum: Obesity, diabetes and hospitalization with pneumonia. *2009*.

48. Theis Thilemann: Medication use and risk of revision after primary total hip arthroplasty. *2009*.

49. Operativ fjernelse af galdeblæren. Region Midtjylland & Region Nordjylland. 1998-2008. *2009*.

50. Mette Søgaard: Diagnosis and prognosis of patients with community-acquired bacteremia. *2009*.

51. Marianne Tang Severinsen. Risk factors for venous thromboembolism: Smoking, anthropometry and genetic susceptibility. *2010*.

52. Henriette Thisted: Antidiabetic Treatments and ischemic cardiovascular disease in Denmark: Risk and outcome. *2010*.

53. Kort- og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme. Region Midtjylland og Region Nordjylland 1997-2008. *2010*.

54. Prognosen efter akut indlæggelse på Medicinsk Visitationsafsnit på Nørrebrogade, Århus Sygehus. *2010*.

55. Kaare Haurvig Palnum: Implementation of clinical guidelines regarding acute treatment and secondary medical prophylaxis among patients with acute stroke in Denmark. *2010*.

56. Thomas Patrick Ahern: Estimating the impact of molecular profiles and prescription drugs on breast cancer outcomes. *2010*.

57. Annette Ingeman: Medical complications in patients with stroke: Data validity, processes of care, and clinical outcome. *2010*.

58. Knoglemetastaser og skeletrelaterede hændelser blandt patienter med prostatakræft i Danmark. Forekomst og prognose 1999-2007. *2010*.

59. Morten Olsen: Prognosis for Danish patients with congenital heart defects - Mortality, psychiatric morbidity, and educational achievement. *2010*.

60. Knoglemetastaser og skeletrelaterede hændelser blandt kvinder med brystkræft i Danmark. Forekomst og prognose 1999-2007. *2010*.

61. Kort- og langtidsoverlevelse efter hospitalsbehandlet kræft. Region Midtjylland og Region Nordjylland 1998-2009. *2010*.

62. Anna Lei Lamberg: The use of new and existing data sources in non-melanoma skin cancer research. *2011*.

63. Sigrún Alba Jóhannesdóttir: Mortality in cancer patients following a history of squamous cell skin cancer – A nationwide population-based cohort study. *2011*.

64. Martin Majlund Mikkelsen: Risk prediction and prognosis following cardiac surgery: the EuroSCORE and new potential prognostic factors. *2011*.

65. Gitte Vrelits Sørensen: Use of glucocorticoids and risk of breast cancer: a Danish population-based case-control study. *2011*.

66. Anne-Mette Bay Bjørn: Use of corticosteroids in pregnancy. With special focus on the relation to congenital malformations in offspring and miscarriage. *2012*.

67. Marie Louise Overgaard Svendsen: Early stroke care: studies on structure, process, and outcome. *2012*.

68. Christian Fynbo Christiansen: Diabetes, preadmission morbidity, and intensive care: population-based Danish studies of prognosis. *2012*.

69. Jennie Maria Christin Strid: Hospitalization rate and 30-day mortality of patients with status asthmaticus in Denmark – A 16-year nationwide population-based cohort study. *2012*.

70. Alkoholisk leversygdom i Region Midtjylland og Region Nordjylland. 2007-2011. *2012*.

71. Lars Jakobsen: Treatment and prognosis after the implementation of primary percutaneous coronary intervention as the standard treatment for ST-elevation myocardial infarction. *2012*.

72. Anna Maria Platon: The impact of chronic obstructive pulmonary disease on intensive care unit admission and 30-day mortality in patients undergoing colorectal cancer surgery: a Danish population-based cohort study. *2012*.

73. Rune Erichsen: Prognosis after Colorectal Cancer - A review of the specific impact of comorbidity, interval cancer, and colonic stent treatment. *2013*.

74. Anna Byrjalsen: Use of Corticosteroids during Pregnancy and in the Postnatal Period and Risk of Asthma in Offspring - A Nationwide Danish Cohort Study. *2013*.

75. Kristina Laugesen: In utero exposure to antidepressant drugs and risk of attention deficit hyperactivity disorder (ADHD). *2013*.

76. Malene Kærslund Hansen: Post-operative acute kidney injury and five-year risk of death, myocardial infarction, and stroke among elective cardiac surgical patients: A cohort study. *2013*.

77. Astrid Blicher Schelde: Impact of comorbidity on the prediction of first-time myocardial infarction, stroke, or death from single-photon emission computed tomography myocardial perfusion imaging: A Danish cohort study. *2013*.

78. Risiko for kræft blandt patienter med kronisk obstruktiv lungesygdom (KOL) i Danmark. (Online publication only). *2013*.

79. Kirurgisk fjernelse af milten og risikoen for efterfølgende infektioner, blodpropper og død. Danmark 1996-2005. (Online publication only). *2013*.

Jens Georg Hansen: Akut rhinosinuitis (ARS) – diagnostik og behandling af voksne i almen praksis. *2013*.

80. Henrik Gammelager: Prognosis after acute kidney injury among intensive care patients. *2014*.

81. Dennis Fristrup Simonsen: Patient-Related Risk Factors for Postoperative Pneumonia following Lung Cancer Surgery and Impact of Pneumonia on Survival. *2014*.

82. Anne Ording: Breast cancer and comorbidity: Risk and prognosis. *2014*.

83. Kristoffer Koch: Socioeconomic Status and Bacteremia: Risk, Prognosis, and Treatment. *2014*.

84. Anne Fia Grann: Melanoma: the impact of comorbidities and postdiagnostic treatments on prognosis. *2014*.

85. Michael Dalager-Pedersen: Prognosis of adults admitted to medical departments with community-acquired bacteremia. *2014*.