Epidemiology of venous thromboembolism with focus on risk markers

Magdalena Johansson

Department of Public Health and Clinical Medicine
Umeå 2019
“An expert is one who knows more and more about less and less until he knows absolutely everything about nothing.”

Nicholas M. Butler (1862–1947)
Table of Contents

Abstract........................................................................................................ iv
   Background ................................................................................................. iv
   Aims ............................................................................................................ iv
   Methods ....................................................................................................... iv
   Results ......................................................................................................... v
   Conclusions ............................................................................................... v

Sammanfattning på svenska ........................................................................ vii
   Bakgrund .................................................................................................. vii
   Syfte .......................................................................................................... vii
   Metod ......................................................................................................... vii
   Resultat ....................................................................................................... viii
   Slutsatserser ............................................................................................. ix

Abbreviations .............................................................................................. x

Original papers ............................................................................................ xi

Introduction .................................................................................................. 1
   The formation of venous thrombi .............................................................. 1
   Deep vein thrombosis (DVT) .................................................................... 3
   Pulmonary embolism (PE) ....................................................................... 3
   Rare thrombosis locations ........................................................................ 4
   Incidence of VTE ...................................................................................... 4
   Accuracy of VTE diagnoses in administrative registries ......................... 5
   Provoked and unprovoked VTE ............................................................... 6
   Provoking factors ..................................................................................... 6
      Institutionization ................................................................................... 6
      Surgery .................................................................................................. 6
      Trauma .................................................................................................. 6
   Immobilization .......................................................................................... 7
   Central venous catheters ......................................................................... 7
   Pregnancy, puerperium and hormone therapy ......................................... 7
   Cancer ....................................................................................................... 8
      Non-malignant persistent provoking factors ........................................... 8
   Risk factors for VTE ................................................................................. 8
   Hereditary risk factors for VTE ............................................................... 9
      Family history ....................................................................................... 9
      Ethnicity ................................................................................................. 9
      ABO blood type and coagulation factor VIII ....................................... 9
      Antithrombin deficiency ..................................................................... 10
      Protein C and protein S deficiency ...................................................... 10
      Activated protein C resistance ............................................................ 10
      Prothrombin mutation ....................................................................... 10
   Acquired risk factors for VTE ............................................................... 11
      Age ....................................................................................................... 11
Antiphospholipid antibodies ................................................................. 11
Smoking ......................................................................................... 12
Obesity............................................................................................... 12
Diabetes and hyperglycemia .............................................................. 12
Alcohol use ..................................................................................... 14
Physical activity ............................................................................ 15
Treatment of VTE ............................................................................ 17
Prognosis of VTE ............................................................................ 18
Recurrence of VTE .......................................................................... 20
Risk factors and risk markers ........................................................... 20
Causality ........................................................................................ 22
Study designs .................................................................................. 23
Cohort studies ................................................................................ 24
Cohort effects ................................................................................ 26
Reverse causality ........................................................................... 26
Confounding ................................................................................... 26
Selection bias ................................................................................ 28
Information bias ............................................................................. 29

Objectives .......................................................................................... 30

Materials and Methods ................................................................. 31
Study population – paper I ............................................................... 31
Study population – papers II to V .................................................... 31
  The Västerbotten Intervention Programme (VIP) ................................ 31
  The Venous thromboEmbolism In Northern Sweden (VEINS) study .... 31
Cases ............................................................................................... 33
  Case identification – paper I............................................................. 33
  Case identification – papers II to V .................................................. 33
  Validation of potential VTE events – papers I to V ......................... 34
  Classification of cases according to VTE location – papers I to III .... 35
  Risk markers for VTE at the time of VTE diagnosis – paper I ........... 35
Measurements and definitions – papers II to V ................................. 36
  Height, weight and body mass index .............................................. 36
  Education level ............................................................................. 36
  Smoking ....................................................................................... 36
  Blood pressure and hypertension .................................................. 36
  Cancer .......................................................................................... 36
  Glucose levels and diabetes ............................................................ 37
  Alcohol consumption .................................................................. 37
  Alcohol dependence .................................................................... 38
  Leisure time physical activity ....................................................... 38
  Occupational physical activity ...................................................... 38
  Cambridge index ......................................................................... 38
Statistical analyses ........................................................................ 39
Ethical considerations .................................................................... 40

Results ............................................................................................... 42
  Accuracy of VTE diagnoses in administrative registries ................. 42
Incidence of first-time VTE in relation to age and sex ........................................ 45
Risk markers at first-time VTE diagnosis ............................................................... 46
The VEINS study population ................................................................................. 46
Risk marker profile at first-time VTE diagnosis ................................................... 50
Glucose levels, diabetes and risk of first-time VTE ............................................... 51
Alcohol intake, alcohol dependence and risk of first-time VTE ......................... 53
Leisure time physical activity, occupational physical activity and risk of first-time VTE .......................................................... 57
Interaction between sex and risk markers for VTE .............................................. 59

Discussion .............................................................................................................. 60
Main findings .......................................................................................................... 60
  Accuracy of VTE diagnoses in administrative registries ...................................... 60
  Incidence of first-time VTE ................................................................................. 61
  Risk markers for first-time VTE at the time of VTE diagnosis ....................... 62
  Glucose levels and risk of first-time VTE .......................................................... 63
  Diabetes and risk of first-time VTE ................................................................. 64
  Alcohol consumption and risk of first-time VTE ............................................. 64
  Alcohol dependence and risk of first-time VTE .............................................. 66
  Physical activity and risk of first-time VTE ..................................................... 66
Methodological considerations .............................................................................. 67
Internal validity ....................................................................................................... 67
  Exposure variables from medical records ........................................................ 67
  Exposure variables from health examinations ............................................... 68
  Risk of selection bias in the inclusion of study participants ............................. 70
  Risk of selection bias conferred by the treatment of missing data .................. 71
Confounders ........................................................................................................... 72
Outcome assessment .............................................................................................. 73
External validity ...................................................................................................... 75
  Causality ............................................................................................................ 75
  Gender perspective ............................................................................................ 76
  Clinical implications ......................................................................................... 77
  Future perspectives ......................................................................................... 78

Conclusions .......................................................................................................... 79

Acknowledgements ............................................................................................... 80

References .............................................................................................................. 82

Appendix ................................................................................................................ 1
  Corrections to paper I ......................................................................................... 1

Original papers
Abstract

Background
Venous thromboembolism (VTE) is a vascular disease with an incidence of approximately 140 cases per 100,000 person-years in adults. The incidence of VTE has increased over the last decades, and more than 20% of affected individuals die in the first year after diagnosis. To reduce the incidence of VTE, it is important to identify modifiable risk factors for the condition.

Aims
The aims of this thesis were a) To study the incidence of first-time VTE and the prevalence of risk markers for VTE at the time of VTE diagnosis, b) To determine the validity of diagnoses of deep vein thrombosis and pulmonary embolism in administrative registries, and c) To study the association between glucose levels, diabetes, alcohol consumption, physical activity and risk of first-time VTE.

Methods
To determine the incidence of first-time VTE and the prevalence of risk markers for VTE at the time of VTE diagnosis, a retrospective, population-based cohort study was conducted. The study included all adult residents of Västerbotten County during the year 2006.

All other aims were addressed in the prospective, population-based Venous thromboEmbolism In Northern Sweden (VEINS) cohort study. The VEINS cohort included 108,025 residents of Västerbotten County aged 30 to 60 years without previous VTE events. They were included from 1985 onwards and were followed until a VTE event, death, emigration, or the study end on September 5, 2014. All underwent a health examination within the Västerbotten Intervention Programme where weight, height, blood pressure and glucose levels were measured, and answered a questionnaire regarding smoking, education level, medication use, history of diabetes, alcohol intake and physical activity. VTE diagnoses were validated by review of medical records and radiology reports.

To study the validity of diagnoses of deep vein thrombosis and pulmonary embolism in administrative registries, a registry search for International Classification of Diseases diagnosis codes indicating pulmonary embolism and/or deep vein thrombosis events was made in the Swedish National Patient Registry and the Cause of Death Registry. An additional search using an
extended set of International Classification of Diseases diagnosis codes was performed in order to identify misclassified events.

**Results**

The incidence of first-time VTE was 137 (95% confidence interval [CI] 122–154) per 100,000 adults per year. The most common risk markers for VTE were recent hospitalization and concurrent malignancy.

The positive predictive value for a diagnosis of pulmonary embolism was 80.7% (95% CI 78.4–82.9), and that of deep vein thrombosis 59.2% (95% CI 56.7–61.7). Misclassification occurred in 1.1% (95% CI 0.4–1.7) of pulmonary embolism events and in 16.4% (95% CI 14.2–18.7) of deep vein thrombosis events.

In the VEINS cohort, a total of 2,054 participants experienced an objectively verified first-time VTE event during approximately 1.5 million person-years of follow-up. In univariable analysis, there were associations between fasting plasma glucose, oral glucose tolerance test two-hour post-load plasma glucose, diabetes and increased risk of first-time VTE. These associations were attenuated after adjustment for potential confounders, and were no longer significant.

There was an association between alcohol consumption and risk of first-time VTE in men ($P$ for trend 0.02 after adjustments for increased risk of first-time VTE over quartiles of weekly alcohol consumption). Alcohol dependence was associated with risk of first-time VTE in men (hazard ratio [HR] 1.30; 95% CI 1.07–1.59 after adjustments). In women, there were no significant associations between alcohol consumption and risk of first-time VTE.

Women who performed leisure time physical activity at least once a week had a lower risk of first-time VTE (HR 0.83; 95% CI 0.71–0.98 after adjustments) compared to women with less or no physical activity. Women with high occupational physical activity also had a lower risk of first-time VTE (HR 0.85; 95% CI 0.74–0.98 after adjustments). In men, there were no consistent association between either measure of physical activity and risk of first-time VTE.

**Conclusions**

VTE is a common vascular disease. Registry data on diagnoses of pulmonary embolism, but not deep vein thrombosis, is of acceptable quality and can be considered for use in registry-based studies. Glucose levels and diabetes are not
associated with risk of first-time VTE. Alcohol intake and alcohol dependence are associated with an increased risk of first-time VTE in men, whereas high leisure time physical activity and occupational physical activity are associated with a decreased risk of first-time VTE in women.
Sammanfattning på svenska

Bakgrund
Venös trombos är en vanlig och allvarlig kärlsjukdom. De vanligaste formerna av venös trombos är djup ventrombos och lungemboli. Förekomsten av venös trombos har ökat under de senaste decennierna. Mer än en av fem personer som drabbas av venös trombos dör under det första året efter diagnos. Hos ungefär en tredjedel av dem som insjuknar i venös trombos hittar man ingen uppenbar förklaring till att de får sjukdomen. För att minska insjuknandet i venös trombos är det viktigt att identifiera påverkbara riskfaktorer för tillståndet.

Syfte
Syftet med denna avhandling var att studera frekvensen av nyinsjuknande i venös trombos i befolkningen och förekomsten av bidragande faktorer till venös trombos vid insjuknandet. Vi ville också undersöka i vilken utsträckning personer med en diagnoskod för venös trombos i sjukvårdsregister faktiskt haft en venös trombos. Slutfinen ville vi studera om det finns ett samband mellan blodsockernivåer, diabetes, alkoholintag, fysisk aktivitet och risken för att senare i livet få en första venös trombos.

Metod
För att undersöka antalet personer som nyinsjuknar i venös trombos och förekomsten av riskmarkörer för venös trombos vid insjuknandet genomförde vi en retrospektiv (tillbakablickande) befolkningsbaserad studie. Vi registrerade alla vuxna i Västerbottens Län som fick venös trombos under år 2006 och vilka bidragande faktorer till venös trombos de hade vid insjuknandet.

För att studera i vilken utsträckning diagnoser för venös trombos i sjukvårdsregister var korrekt gjorde vi en sökning för att hitta diagnoskoder för lungemboli och djup ventrombos i Nationella Patientregistret och Dödsorsaksregistret. Vi gjorde ytterligare en sökning, efter ännu fler diagnoskoder, för att hitta insjuknanden i venös trombos som felklassificerats som andra sjukdomar.

Sambandet mellan blodsockernivåer, diabetes, alkoholintag, fysisk aktivitet och risken för att senare i livet få en första venös trombos undersökte vi i en prospektiv (framåtblickande), befolkningsbaserad studie som vi kallat Venous thromboEmbolism In Northern Sweden (VEINS)-studien. I VEINS-studien deltog de 108 025 invånare i Västerbottens Län som inte haft någon tidigare venös trombos och som genomgått en hälsoundersökning inom ramen för

Personer i studiegruppen som fått en venös trombos identifierades genom sökning efter diagnoskoder för venös trombos i diagnosregister. Riktigheten i diagnoser kontrollerades genom granskning av svar på röntgenundersökningar och medicinska journaler.

Resultat

I Västerbottens Län fick varje år 137 av 100 000 vuxna venös trombos för första gången. De vanligaste bidragande faktorerna till att personer fick venös trombos var att de nyligen lega på sjukhus eller hade en cancersjukdom.

Av alla personer med en diagnoskod för lungemboli i ett sjukvårdsregister var diagnosen korrekt hos 81%. För djup ventrombos var motsvarande siffra 59%. En av hundra personer med lungemboli hade inte en diagnoskod för lungemboli i ett sjukvårdsregister. För djup ventrombos var motsvarande siffra sexton av hundra.

I VEINS-studien insjuknade 2054 studiedeltagare i venös trombos för första gången. Totalt földes studiedeltagarna i 1,5 miljoner personår. Deltagare med högre fasteblodsocker, högre blodsockervärde taget två timmar efter intag av en sockerlösning och deltagare med diabetes hade högre risk för venös trombos än andra deltagare. Efter att dessa analyser korrigerats för andra faktorer som har samband med blodsockervärden och diabetes och som kan ge ökad risk för venös trombos (störfaktorer), till exempel ålder och övervikt, fanns det inte längre något samband mellan blodsockervärden, diabetes och risk att drabbas av venös trombos.

Kvinnor som idrottade på fritiden minst en gång i veckan hade en lägre risk för att få en venös trombos jämfört med kvinnor som mer sällan eller inte alls idrottade på fritiden. Kvinnor med ett fysiskt aktivt arbete hade också lägre risk för venös trombos jämfört med kvinnor med ett mer stillasittande arbete. Hos män såg vi inget tydligt samband mellan fysisk aktivitet på fritiden eller i arbetet och risk för venös trombos.

**Slutsatser**

Abbreviations

Alcohol depend., alcohol dependence

2HPG, oral glucose tolerance test two-hour post-load plasma glucose

BMI, body mass index

CAGE, Cut-down, Annoyance, Guilt, Eye-opener

CI, confidence interval

CTEPH, chronic thromboembolic pulmonary hypertension

CT, computed tomography

DVT, deep vein thrombosis

FPG, fasting plasma glucose

HR, hazard ratio

ICD, International Classification of Diseases

Leisure PA, leisure time physical activity

Occup. PA, occupational physical activity

OGTT, oral glucose tolerance test

PA, physical activity

PE, pulmonary embolism

SD, standard deviation

VEINS, Venous thromboembolism In Northern Sweden

VIP, Västerbotten Intervention Programme

VTE, venous thromboembolism
Original papers


Introduction

The formation of venous thrombi

Venous thromboembolism (VTE) is a condition where a thrombus (blood clot) is formed in a vein (1). The term VTE is usually restricted to thrombi located in deep veins, as opposed to superficial vein thrombosis. This definition of VTE is used throughout this thesis. Most venous thrombi are formed in the lower extremities. Venous thrombi differ from arterial thrombi; venous thrombi are primarily composed of fibrin and red blood cells, whereas the main content of arterial thrombi is platelets. Arterial thrombi form at or around ruptured atherosclerotic plaques, whereas venous thrombi can occur even if the endothelium is intact (2). The formation of a thrombus is facilitated by the presence of abnormalities of blood flow, vascular wall and blood clotting components. These three factors are known as Virchow’s triad (Figure 1) (3).

Figure 1 Abnormal blood flow, abnormalities in the blood vessel wall and hypercoagulability are the three components of Virchow’s triad.
The first component of Virchow’s triad is abnormal blood flow. The majority of venous thrombi have their origin in regions with slow blood flow, for example the large venous sinuses of the calf and thigh, or in the valve cusp pockets upstream of venous valves. Thrombi can also originate in bifurcations of the venous system where blood flow irregularities are present (4, 5). Blood pooling can cause activation of the coagulation system, which in turn leads to local hypercoagulability. Endothelial damage due to the distention of the vessel wall can potentially contribute to activation of the clotting system (4, 5). Examples of conditions that can result in venous stasis are obesity, pregnancy, heart failure, and tumors causing external compression of veins (5). The veno-muscular pumps, i.e. the pumping action of extremity skeletal muscles that promotes venous return, have an important role in preventing venous stasis. In situations where skeletal muscles, the calf muscles in particular, are less active, venous stasis can form. Examples of such situations are extremity fractures, orthopedic casts or restraints, paralysis and hospitalization with bed rest (4, 5).

Vascular wall abnormalities, the second component of Virchow’s triad, can occur in the form of endothelial damage in the setting of surgery, trauma or presence of indwelling venous catheters (4). Increased levels of markers of endothelial dysfunction have been found in persons with previous VTE compared to matched controls (6), but it is uncertain whether endothelial dysfunction is a cause or a consequence of the VTE event.

The third component of Virchow’s triad is hypercoagulability. The risk of VTE increases when the balance between the pro- and anticoagulant forces is shifted towards blood coagulation, as illustrated in Figure 2 (4).

Figure 2 In venous thromboembolism, the balance between pro- and anticoagulant forces is shifted towards blood coagulation.

This imbalance can be inherited or acquired. Examples of inherited conditions causing hypercoagulability are activated protein C resistance, prothrombin mutation and antithrombin deficiency (7), which are described in detail on pages 9 through 11. Such conditions, called hereditary thrombophilias, cause a
lifelong hypercoagulability. Acquired hypercoagulability can for example be caused by medication (e.g. estrogen therapy), pregnancy, cancer and autoimmune disorders. The duration of an acquired hypercoagulability depends on its cause (8).

**Deep vein thrombosis (DVT)**

DVTs of the lower extremities are predominantly located in the left lower extremity (9). Of patients with DVT, about one fourth of patients have a distal DVT (i.e. an isolated calf vein DVT or muscular vein DVT), half the patients have a proximal DVT involving femoropopliteal veins, but not veins above the inguinal ligament, and one fourth have a proximal DVT involving veins above the inguinal ligament (9). Common symptoms of lower extremity DVT are erythema, tenderness and swelling (10).

The gold standard for diagnosis of DVT has been venography, but this method is not recommended in current guidelines (11). Instead, the first-line method for identifying lower extremity DVTs is venous ultrasound, preceded by D-dimer testing in patients with low pre-test probability of DVT (12). The sensitivity and specificity of venous ultrasound for identifying proximal lower extremity DVTs is greater than 90%, whereas for distal lower extremity DVTs, the sensitivity is about 70% (13, 14). Due to the lower sensitivity of venous ultrasound in the diagnosis of distal lower extremity DVT, it is likely that the true prevalence of distal DVTs is underestimated. DVT of the upper extremities is a rare condition; only about 5% of DVTs affect the upper extremities (15). Symptoms of upper extremity DVT are similar to those caused by lower extremity DVT, and the condition can be diagnosed using venous ultrasound (16).

**Pulmonary embolism (PE)**

PEs are thought to be fragments of DVTs that move through the blood stream to the pulmonary vasculature. A lower extremity DVT can be identified in about 60% of the patients with PE (17, 18). The origin of the PEs in the remaining 40% of patients is not known. Right-sided cardiac thrombi and local thrombus formation in the pulmonary arteries have been suggested as explanations (19). Symptoms of PE include sudden onset dyspnea, chest pain, syncope and hemoptysis (20).

The gold standard diagnostic method for PE is pulmonary angiography, but this method is associated with a mortality of 0.5% (21) and is more costly compared to other diagnostic methods, e.g. CT angiography (22). For these reasons, it has fallen out of favor (23). Current guidelines regarding suspected PE in a patient without shock or hypotension advocate a strategy where the clinical probability
of PE is assessed using clinical judgment or a prediction rule. If the clinical probability of PE is low or intermediate, D-dimer testing is performed. If the D-dimer test is positive, the patient is recommended to undergo computed tomography (CT) angiography. In patients with a high clinical probability of PE, CT angiography is performed without previous D-dimer testing (23).

In patients with shock or hypotension where PE is suspected, a CT angiography is performed if immediately available. Otherwise, an echocardiography is performed. If right ventricular overload is present and the patient can be stabilized, a CT angiography can be made to confirm the diagnosis of PE. If right ventricular overload is present and no other diagnostic test for PE is available, or if the patient is unstable, PE-specific treatment can be initiated. If no right ventricular overload can be detected on echocardiography, other causes of hemodynamic instability should be considered (23).

**Rare thrombosis locations**

VTE events can also occur in other locations. Cerebral venous thrombosis is a rare type of stroke that can occur at any age. Headache is the most frequently reported symptom of cerebral venous thrombosis. Other symptoms and signs include seizures, focal neurological signs, impaired consciousness and papilledema (24, 25).

Abdominal VTE is another rare form of VTE. Persons experiencing an abdominal VTE are, in general, younger than persons with lower extremity DVT or PE. Abdominal VTE events can for example occur in the hepatic, portal, splenic, mesenteric or renal veins, or in the inferior vena cava. Patients with abdominal VTE can present with pain, edema, and symptoms and signs related to organ dysfunction (26).

**Incidence of VTE**

The incidence of first-time VTE is reported to be 62 to 143 cases per 100,000 person-years (27-32). The incidence increases steeply with age; in individuals aged 80 and up, the incidence of first-time VTE is approximately 800 cases per 100,000 person-years, compared to about 20 cases per 100,000 person-years in persons aged below 40 (31). At least one out of every twelve middle-aged adults will develop VTE during their remaining lifetime (33). The overall incidence of first-time VTE is similar in men and women, but the age-specific incidence differs between men and women (27, 29, 30). Women of childbearing age have a higher incidence of first-time VTE compared to men, whereas later in life, the incidence of first-time VTE is higher in men than in women (27). It has been suggested that the higher risk of VTE in young and middle-aged women is due
to exposure to reproductive risk factors (oral contraception, pregnancy/puerperium and postmenopausal hormone therapy). In a study that compared men and women without reproductive risk factors, the odds ratio for first-time VTE in men was 2.1 (95% confidence interval [CI] 1.9–2.4) (34).

The incidence of first-time DVT has been suggested to be 41–93 cases per 100,000 person-years, and the incidence of first-time PE has been suggested to be 46–66/100,000 person-years (27-31). Other forms of VTE are rare; the incidence of upper extremity DVT is 3.6 per 100,000 person-years (35), that of cerebral sinus thrombosis is 1.6 cases per 100,000 person-years (36), that of mesenteric vein thrombosis 2.7 per 100,000 person-years (37), and that of vena cava thrombosis 1.7 per 100,000 person-years (38).

**Accuracy of VTE diagnoses in administrative registries**

Diagnosis codes in administrative health care registries (e.g. the National Patient Registry in Sweden or the Medicare database in the United States) are increasingly used to conduct large-scale, cost-effective studies of various diseases. The accuracy of diagnosis codes in registries varies between different medical conditions; for example, the positive predictive value of a diagnosis of atrial fibrillation in the Swedish National Patient Registry is 96%, compared to 35% for herpes simplex encephalitis (39, 40). The positive predictive value of diagnosis codes for a disease determines whether unvalidated diagnosis codes from administrative registries can be used in research.

There are very few studies of the accuracy of VTE diagnoses in Swedish administrative registries, such as the Swedish National Patient Registry. One Swedish study showed a positive predictive value of more than 90% for VTE diagnoses in the Swedish National Patient Registry and the Swedish Cause of Death Registry. However, that study included men only, was conducted during a time period where modern diagnostic methods were not widely available, and during which only inpatient care was reported to the Swedish National Patient Registry (41). For VTE, inpatient diagnoses are generally more accurate than those coded in an outpatient setting. This has been shown in a Danish study and an American study of members of four healthcare delivery systems (42, 43). A Danish study showed that the positive predictive value for PE diagnosis codes was higher than that of diagnosis codes for DVT (42). It has not been investigated in a Swedish setting. The sensitivity of diagnosis codes for identifying VTE events (i.e. the proportion of VTE events identified by a diagnosis registry search) was higher for PE than for DVT in a French study (44).
Provoked and unprovoked VTE

VTE events can be classified as unprovoked, provoked by a transient risk factor, and provoked by a persistent risk factor. Unprovoked VTE events are not associated with transient or persistent provoking factors. Cancer is the most important persistent provoking risk factor for VTE (45). The proportion of unprovoked, provoked and cancer-related VTE events vary between different studies. About 18–30% of first-time VTE events occur in patients with active cancer, 32–50% are associated with other provoking factors and 20–50% are unprovoked (27, 46, 47). The distinction between provoked and unprovoked VTE events is important in a clinical setting where a patient’s risk of VTE recurrence is estimated and the duration of anticoagulant treatment is decided (12, 23).

Provoking factors

Institutionalization

The most common transient provoking factor for VTE is institutionalization. Of all cases of VTE, 35% can be attributed to nursing home stays or hospital stays without surgery, and 24% to hospital stays combined with surgery (48). Both hospital stays with and without immobilization are associated with increased risk of VTE (49). The age- and sex-adjusted incidence of VTE in hospitalized patients is more than 100 times that of community residents (50).

Surgery

Surgery is another transient risk factor for VTE. A case-control study showed an odds ratio of 22 for increased risk of VTE in patients who had undergone surgery requiring anesthesia during the past three months (51). The increased risk of VTE in persons who have undergone surgery decreases with time, but is detectable for at least one year after the procedure. The risk of VTE is higher in patients that undergo inpatient surgery compared to day surgery (52), and the duration of the surgical procedure is positively associated with risk of VTE (53). The risk of VTE also varies depending on the type of surgery performed. Patients undergoing major orthopedic surgery or cancer surgery are at particularly high risk of VTE (52, 54).

Trauma

Trauma requiring hospital admission is associated with an over 12-fold increase in risk of VTE (51). Injury severity score, number of operative procedures, pelvic injuries and concomitant medical conditions have been identified as markers of increased risk of VTE in trauma patients. In-hospital mortality in trauma patients with VTE is about double that of trauma patients without VTE (55). A
below-knee orthopedic cast seems to be a stronger risk factor for VTE when the indication for the cast is a traumatic injury (56). Minor injuries of the leg (not requiring surgery, orthopedic casts or extended bed rest) are also associated with an increased risk of VTE (57).

**Immobilization**

Immobilization is a further risk factor for VTE. For example, patients with a neurological disease with extremity paresis are at increased risk of VTE (51). In hemiparetic stroke, unilateral lower limb DVT usually affects the paretic limb (58). Prolonged travel is also associated with a small increase in risk of VTE; 27 cases of VTE occur every 1,000,000 flights (59). The risk of VTE is similar in persons who travel by air, train, car or bus (60). TV viewing, which can be used as a proxy for sedentary behavior, is associated with VTE in a dose-dependent manner (61).

**Central venous catheters**

Presence of central venous catheters or pacemaker electrodes is also associated with increased risk of VTE (51). In studies where routine diagnostic screening was used, the percentage of patients with a clinically manifest catheter-related DVT ranged between 0 and 12%. The risk of VTE in this setting varies depending on the catheter type and material (62). Patients with peripherally inserted central venous catheters seem to be at higher risk of VTE compared to patients with centrally inserted catheters (63).

**Pregnancy, puerperium and hormone therapy**

The risk of VTE is increased during pregnancy and puerperium. The incidence of VTE is 96 per 100,000 woman-years in pregnancy, and 511 per 100,000 woman-years during the first three months after delivery (64). During pregnancy clotting factor concentrations increase (65, 66), concentrations of endogenous anticoagulants decrease, and within the fibrinolytic system, both activator and inhibitor concentrations increase (65). The overall activity of the fibrinolytic system is reduced. These changes are most likely a physiologic response to maintain normal placental function and reduce the risk of massive hemorrhage at delivery (67). Use of combined oral contraceptives (68) and hormone replacement therapy (69) are also associated with increased risk of VTE, especially in women with a family history of VTE (70). Transdermal administration of hormone replacement therapy may be associated with a lower risk of VTE compared to oral administration (71). In contrast, high endogenous concentrations of sex hormones are not associated with increased risk of VTE (72).
Cancer
Active cancer is the most important persistent provoking risk factor for VTE (45), as cancer is a common condition (73), as well as a strong risk factor for VTE (51). The term “active cancer” is not well-defined, but a suggested definition includes cancer that has not received potentially curative therapy, cancer where there is evidence that the therapy has not been curative, and cancer where treatment is ongoing or the disease-free interval after treatment is too short to properly evaluate treatment effect (45).

There are many reasons for the increased risk of VTE in cancer patients. Tumor cells can have procoagulant activity. Aberrant tissue factor expression has been found in many human cancers (74), and some tumor cells produce cancer procoagulant which activates coagulation factor X independently of coagulation factor VII (75, 76). Circulating microparticles can have procoagulant properties, and tissue factor-bearing microparticles are associated with increased risk of VTE in persons with cancer (77). Both activating and inhibitory factors of the fibrinolytic pathway can be expressed on the surface of tumor cells (78). They can also secrete proinflammatory cytokines that can affect the endothelium in a procoagulant direction. Tumor cells that adhere to the vascular endothelium and/or extracellular matrix can promote localized clotting by releasing cytokines that attract other cells (79). Tumor cells can interact with the monocyte-macrophage system and induce these cells to express tissue factor (80). Furthermore, tumor cells can induce platelet activation and aggregation (81). Anti-cancer treatments such as surgery (82), hormonal therapy (83), and some anti-cancer drugs (84, 85), are associated with an increased risk of VTE. Long-term central venous catheters, commonly used in patients with cancer, are also associated with increased risk of VTE (51, 86).

Non-malignant persistent provoking factors
Non-malignant conditions considered to be persistent provoking factors are generally associated with at least a doubled risk of VTE recurrence after discontinuation of anticoagulant treatment for a first VTE event. For example, chronic inflammatory conditions such as inflammatory bowel disease (87) and autoimmune diseases (88, 89) are considered to be persistent provoking factors.

Risk factors for VTE
Many factors are associated with increased risk of VTE. The impact of these factors can vary between first-time and recurrent VTE events (90, 91). As the primary focus of this thesis is first-time VTE, the following summary is focused on risk factors for first-time VTE.
Risk factors for VTE can be hereditary or acquired. Hereditary risk factors are, as a rule, non-modifiable, but interactions between hereditary and acquired risk factors mean that the impact of a hereditary risk factor on a person’s risk of VTE can be lowered by addressing a modifiable risk factor (92). Acquired risk factors can be non-modifiable (e.g. advanced age) or modifiable (e.g. obesity).

**Hereditary risk factors for VTE**

*Family history*
Persons with a history of VTE in a first-degree relative have an approximately 2.5-fold higher risk of VTE (93-95). Studies of risk of VTE in adoptees (96), extended families (97), and spouses and siblings with varying age-differences (97, 98) have indicated that the increased risk of VTE in persons with family history of VTE is due to hereditary risk factors, rather than shared environmental risk factors. Family history of VTE is a risk factor for VTE, both in persons with and without an identifiable hereditary thrombophilia. In patients with VTE, the positive predictive value of family history of VTE for hereditary thrombophilia is 30%, and the negative predictive value is 78%, which means that the value of family history of VTE to rule out or rule in presence of hereditary thrombophilia is limited (93).

*Ethnicity*
The risk of VTE may vary between persons with different ethnicities. In the US, an increased risk of VTE was seen in blacks compared to whites in some study cohorts and regions (99). Asians and Pacific Islanders seem to have a lower risk of VTE compared to persons of other ethnicities (100, 101).

*ABO blood type and coagulation factor VIII*
ABO blood type is the most common hereditary risk factor for VTE. Persons with non-O blood type (62% of the Swedish population (102)) have an approximately doubled risk of VTE compared to persons with O blood type (103). This could possibly be explained by the fact that persons with non-O blood type have higher levels of von Willebrand factor (104), which is important for primary hemostasis, as well as serving as a carrier protein for factor VIII, thereby preventing its degradation (105).

High levels of coagulation factor VIII are associated with increased risk of VTE (106). The levels of coagulation factor VIII are only partly genetically determined; 60% of variations in factor VIII coagulant activity depend on non-genetic factors (107). Age, obesity, diabetes and malignancy are all associated with increased levels of coagulation factor VIII (108-111). The association
between high levels of coagulation factor VIII and risk of VTE may be partially mediated through an acquired activated protein C resistance, and partially through an increased rate of formation of thrombin and fibrin (112).

**Antithrombin deficiency**
Antithrombin deficiency is an autosomal dominant disorder that encompasses both quantitative and qualitative deficiencies in antithrombin (113). The prevalence of antithrombin deficiency is 0.02 to 0.2% (114, 115). Antithrombin is primarily an endogenous anticoagulant (113), and antithrombin deficiency is associated with increased risk of VTE with an OR of 14 (116). In prospective studies, the yearly risk of VTE in individuals with antithrombin deficiency exceeds 2% (116).

**Protein C and protein S deficiency**
Protein C and protein S deficiency are other rare hereditary thrombophilias. Activated protein C inactivates the active forms of coagulation factors V and VIII. This process is enhanced by a cofactor – protein S. A large number of mutations in the genes for proteins C and S are associated with increased risk of VTE (7). In a Scotch population, the prevalence of protein C deficiency was approximately 0.2% (117), and that of protein S deficiency between 0.03 and 0.13% (118). Protein C deficiency is associated with an about 7 to 8-fold increase in risk of VTE (119, 120), and protein S deficiency is associated with an approximately 8-fold increase in risk of VTE (120).

**Activated protein C resistance**
Activated protein C resistance is a form of hereditary thrombophilia in which active coagulation factor V is more resistant to inactivation by activated protein C. This leads to increased thrombin generation as well as decreased activated protein C anticoagulant activity. The most common cause of activated protein C resistance is a mutation in the factor V gene that causes one amino acid to replace another in the gene product, also known as factor V Leiden (7, 121). Heterozygotes for factor V Leiden have a doubled risk of DVT, but no increase in risk of PE (91), whereas homozygotes have an increased risk of both DVT and PE (91, 122). About 10% of the Swedish population is heterozygous, and 0.3% homozygous for factor V Leiden (91). Activated protein C resistance can also be an acquired hypercoagulable state, for example in users of oral contraceptives (123).

**Prothrombin mutation**
Prothrombin is a protein with procoagulant, anticoagulant and antifibrinolytic activity (7). A mutation termed G20210A in the gene that codes for prothrombin
is associated with an almost 3-fold increase in risk of VTE (124). The prevalence of this mutation is 1.8% in Sweden (125). The prevalence varies between different ethnic groups (126).

**Acquired risk factors for VTE**

**Age**

Age is a strong risk factor for VTE. As previously described, the incidence of VTE increases markedly with age (31). About 60% of all VTE events occur in persons aged 70 years and above (27). One reason for the association between advancing age and risk of VTE is that elderly persons tend to have a higher burden of provoking factors and risk factors for VTE when compared to younger persons. Institutionalization, surgery and cancer are all more common in older persons compared to younger persons (127-129). Age is also associated with changes in coagulation factors levels, for example levels of coagulation factor VIII and activated coagulation factor VII increase with age (109, 130). Muscle strength declines with age (131), and it is probable that this overall decline also affects the function of the veno-muscular pumps. Venous compliance decreases with increasing age (132, 133). Moreover, venous valve thickness increases with age. This could be one mediator of the association between age and risk of VTE (134, 135).

**Antiphospholipid antibodies**

Lupus anticoagulants, anticardiolipin antibodies and anti-beta 2-glycoprotein I antibodies are collectively known as antiphospholipid antibodies. The presence of antiphospholipid antibodies can be associated with increased risk of VTE. Antiphospholipid antibodies can occur in conjunction with autoimmune conditions, such as systemic lupus erythematosus, or as an isolated phenomenon (136).

In patients with systemic lupus erythematosus, there is an association between presence of antiphospholipid antibodies and increased risk of VTE with an OR of 5.6 for the association between lupus anticoagulants and risk of VTE. The association between anticardiolipin antibodies and risk of VTE is weaker (OR 2.2) (137).

In a meta-analysis where the vast majority of study participants did not have systemic lupus erythematosus, there was an association between lupus anticoagulants and risk of VTE with ORs between 5 and 16 in different studies (138). The association between anticardiolipin antibodies and risk of VTE in persons without systemic lupus erythematosus is debated. The aforementioned
meta-analysis concluded that there is a possible association between anticardiolipin immunoglobulin G antibodies in medium or high titers and risk of VTE (138), whereas two prospective population-based studies found no association between anticardiolipin antibodies and risk of VTE (139, 140).

**Smoking**
Several studies have reported an association between smoking and increased risk of VTE (141, 142). The risk of VTE seems to increase with exposure to smoking, measured as number of pack-years (143). The smoking-related increase in risk of VTE is thought to be driven by an association between smoking and an increased risk of provoked VTE events (144). For example, smoking is associated with an increased risk of cancer in many different organs, and cancer is an important provoking factor for VTE (51, 145).

**Obesity**
Obesity is associated with increased risk of death, as well as increased risk of a range of cardiovascular events such as myocardial infarction, stroke, heart failure and atrial fibrillation (146). A higher body mass index is also associated with increased risk of VTE (144). The association between the metabolic syndrome and risk of VTE seen in some studies is attributable to obesity (147-149). Possible mechanisms for the association between obesity and increased risk of VTE could be the association between obesity and development of an inflammatory and prothrombotic state (150, 151), increased intra-abdominal pressure (152) and altered venous hemodynamics of the lower limbs (153). A Mendelian randomization study showed an association between an obesity-specific genetic locus and DVT complicated by PE. These findings indicate that there may be a causal association between obesity and risk of VTE (154). Body height also influences venous pressure physiology (155), and there seems to be a synergistic effect of tall stature and obesity on the risk of VTE (156).

**Diabetes and hyperglycemia**
Diabetes, with a prevalence of 6.8% in Sweden (157), is a major public health concern. The diabetes prevalence is projected to continue to rise to about 10–12% in 2050 (157). About 90% of adults with diabetes have type 2 diabetes (158), a metabolic condition encompassing insulin resistance and progressive beta-cell dysfunction (159).

One of the features of diabetes is a prothrombotic state with increased levels of tissue factor procoagulant activity and levels of coagulation factor VII (160, 161). Furthermore, diabetes is associated with enhanced thrombin generation (160). Fibrinogen levels are higher in persons with type 2 diabetes compared to
persons without diabetes (161). Worse glycemic control in patients with diabetes type 1 is associated with reduced fibrinolytic activity, reflected by increased levels of plasminogen activator inhibitor type-1 and decreased tissue plasminogen activator activity (162). In persons without diabetes with normal glucose levels, increased levels of plasminogen activator inhibitor type-1 are associated with incident diabetes type 2 (163). In persons without diabetes, coagulation is stimulated by hyperglycemia and fibrinolysis is impaired by hyperinsulinemia (164).

There is an increased risk of arterial thromboembolic events in persons with diabetes (165, 166), as well as in persons with increased glucose levels that do not fulfill the diagnostic criteria for diabetes (167). It is not certain whether there is an increased risk of VTE in persons with increased glucose levels or diabetes. In an individual participant data meta-analysis of prospective studies, there were an association between diabetes and risk of VTE (hazard ratio [HR] 1.50; 95% CI 1.26–1.80), which was attenuated after adjustment for age, sex and body mass index (HR 1.01; 95% CI 0.89–1.15) (144). Similarly, an American case-control study showed that the association between diabetes and risk of VTE seen in univariable analysis was attenuated after adjustment for body mass index (168). This suggests that the association between diabetes and risk of VTE is partly due to a higher prevalence of obesity in persons with diabetes. After adjustment for hospitalization and nursing home confinement, the association between diabetes and risk of VTE was attenuated even further, giving rise to the theory that the seemingly increased risk of VTE in persons with diabetes is due to the fact that they are more often hospitalized or confined to a nursing home. A recent individual participant data meta-analysis showed an association between diabetes and decreased risk of VTE in a cohort that included mostly non-fatal VTE events, and between diabetes and increased risk of VTE in a cohort regarding only fatal VTE events as outcomes (141). The reason for this discrepancy is currently unknown.

Neither high levels of plasma glucose (142, 144, 169-171) nor glycated hemoglobin A (172, 173) have been shown to be associated with risk of VTE. One exception is a study that showed an association between increasing levels of fasting plasma glucose and risk of fatal VTE events (141). The oral glucose tolerance test, where a glucose load of 75 grams is ingested and a plasma glucose value is measured two hours later, can be used to identify persons with type 2 diabetes or impaired glucose tolerance (174). Impaired glucose tolerance is a prediabetic condition with reduced insulin sensitivity, whereas prediabetic fasting glucose values predominantly reflect an impaired insulin secretion (175). Thus, fasting plasma glucose and the oral glucose tolerance test measure different alterations of glucose homeostasis. The association between oral
glucose tolerance test two-hour post-load plasma glucose and risk of VTE has not been studied.

Alcohol use
Alcohol use is a widespread recreational habit; more than four in five Swedes have consumed alcohol during the past year (176). Alcohol consumption differs between men and women. Women are, in general, more often teetotalers and consume alcohol less frequently compared to men (177). Hazardous alcohol consumption is less common among women (176).

Alcohol use has multiple health effects; for example, there is an association between alcohol use and increased risk of cancer and injury (178). Women develop alcohol-related medical conditions, such as liver disease, at a lower cumulative alcohol intake compared to men (177, 179). High alcohol consumption is associated with an increase in mortality (178, 180). In contrast, several studies have shown that moderate alcohol consumption is associated with a decreased risk of arterial cardiovascular disease (181, 182). These epidemiological findings are in line with known associations between alcohol intake and markers of hemostasis. Light to moderate alcohol consumption is associated with lower levels of coagulation factors, whereas high alcohol intake is associated with an impairment of the fibrinolytic capacity (183).

Study participants frequently underreport their alcohol consumption when answering questionnaires. Only 59% of the total alcohol consumption in Swedish residents aged 15 and up is accounted for when questionnaires administered in cohort studies are used to estimate alcohol intake (184). The underreporting differs between age groups, between men and women, and between persons with different levels of alcohol consumption (185). Recent recall methods, such as reporting the amount of alcohol consumed the day before the survey, seem to yield a more accurate estimate of alcohol consumption, but are not adequate for capturing long-term drinking patterns (186). Biomarkers, such as phosphatidylethanol, can also be used to estimate alcohol consumption (187, 188).

It is not clear whether there is an association between alcohol intake and risk of VTE. Previous studies have had diverging results (189-191). Many studies have combined men and women when evaluating the association between alcohol consumption and risk of VTE (141, 190-192), or have only included persons of one sex (193-195).

In Sweden, 6.4% of men and 3.8% of women aged 15 and up suffer from alcohol dependence (196). The Cut-down, Annoyance, Guilt, Eye-opener (CAGE)
screening questionnaire can be used to identify persons with alcohol abuse or dependence (197, 198). Each of the four CAGE items scores one point. A CAGE score of two or more has a sensitivity of 75% and a specificity of 91% for identifying alcohol abuse or dependence in medicine outpatients (197). The CAGE questionnaire is also an appropriate screening test to detect alcohol problems in primary care outpatients and in the general population (199). The association between alcohol dependence and risk of VTE has not been evaluated, but one study has shown an association between alcohol use disorders, defined by International Classification of Diseases (ICD) diagnosis codes, and risk of VTE (200).

**Physical activity**

Physical activity is defined as “any bodily movement produced by skeletal muscles that results in energy expenditure” (201). The two principal categories of physical activity are leisure time physical activity (physical activity performed during an individual’s free time in which the individual participates due to personal interest or need) and occupational physical activity (physical activity performed while working). Leisure time physical activity has a variable duration, whereas occupational physical activity is usually defined as having an eight-hour duration (one working day) (202).

Over thirty different methods of measuring physical activity have been used in research and none of these fulfill the criteria of being accurate, affordable, and not influencing the participant’s behavior (203). A questionnaire is a pragmatic way of quantifying physical activity in large-scale studies (203). However, there is a risk that participants answering questionnaires overestimate their physical activity compared to physical activity measured by an accelerometer. The accordance between the two measures depend on the participant’s age, sex and education level (204). Factors such as seasonal changes in activity patterns, perceived social desirability of certain answers and the interindividual variability in the perceived intensity of an activity limits the usefulness of questionnaire data as a measure of physical activity (205).

Approximately 60% of Swedish adults perform regular leisure time physical activity (206). Persons with higher education levels are more physically active than persons with lower education levels. The differences in physical activity habits by academic achievement have increased over time (207). There are differences in leisure time physical activity between men and women. Men spend more time in moderate and vigorous physical activity compared to women (208), and performing physical activity aimed at increasing muscle strength is more common among men (209). Women spend more time in light physical activity (210), and a higher proportion of women take walks for
pleasure (211). Among Swedes with more than secondary school education, almost 50% of men, and about 30% of women described their working conditions as sedentary in 1999. Among men and women with secondary school education or less, about 20% described their work as sedentary (212).

Physical activity has many health effects. Higher physical activity is associated with lower risk of mortality, cardiovascular disease and cancer (213, 214). When treating occupational physical activity as a separate entity, health benefits are less clear. There is no consistent association between occupational physical activity and risk of myocardial infarction (215, 216), and one study showed that a high level of occupational physical activity was associated with a higher mortality in men, but not in women (217).

The health effects of physical activity may involve several pathways. Single episodes of physical activity induce an anti-inflammatory environment, possibly by inducing a release of interleukin-6 from skeletal muscles. Interleukin-6 then affects levels of other inflammatory mediators in an anti-inflammatory direction. The exercise-induced release of interleukin-6 depends on exercise duration and intensity, the amount of skeletal muscle activated, and, inversely, on the glycogen status of the activated muscles. Regular moderate to vigorous intensity physical activity can reduce chronic inflammation (218).

Physical activity also affects hemostasis. Regarding primary hemostasis, acute strenuous exercise is associated with an increase in platelet count as well as in platelet reactivity and aggregation. These effects are balanced by an adaptive exhaustion of platelet function during resting conditions in persons performing regular physical activity (219).

Acute strenuous exercise, such as long-distance running, is associated with an increased activity of the secondary hemostasis. Notably, thrombin generation increases (220), the activated partial thromboplastin time is shortened (221-223), and both coagulation factor VIII activity (223) and von Willebrand factor levels rise (224). The effects of regular physical activity on secondary hemostasis at rest are incompletely understood. Some studies have shown no effect of regular physical activity on levels of coagulation factors (225, 226), whereas others have shown an association between higher physical activity and lower levels of fibrinogen, von Willebrand factor, and coagulation factor VIII at rest (227-229).

Acute strenuous exercise increases fibrinolytic activity, which can be measured as a shortened euglobin clot lysis time and increased levels of tissue plasminogen activator (223, 230-232). Plasminogen activator inhibitor type-1 has been shown to either decrease (223, 230, 232) or increase (231) after acute
strenuous exercise. D-dimer levels and levels of plasma fibrinogen degradation products increase after acute exercise (220-222, 231). The effects of regular physical activity on the fibrinolytic activity at rest vary between different studies. One study showed that endurance athletes at rest had an increased fibrinolytic activity (225), whereas another study showed no difference in D-dimer levels between endurance athletes and sedentary controls (233).

Studies of the association between physical activity and risk of VTE have had diverging results, showing a positive (193), negative (234, 235), or no (236) association between higher levels of physical activity and risk of VTE. A review article summarizing the available evidence on the association between physical activity and risk of VTE concluded that physical activity may have a small protective effect on the risk of VTE, but the effect does not seem to be dose dependent (237). It is possible that the association between higher physical activity and lower risk of VTE is more pronounced in persons aged 65 and above (234). A prospective study showed no association between occupational physical activity and risk of VTE (236).

**Treatment of VTE**

VTE is treated with anticoagulants. The first anticoagulant used to treat VTE was unfractionated heparin, which has been used for this purpose since the 1930s (238). Vitamin K antagonists were introduced in the 1940s. Sequential VTE treatment with first unfractionated heparin and then oral vitamin K antagonists was the norm (239, 240). In the 1990s, initial outpatient treatment of DVT with low molecular weight heparins was shown to be as safe and effective as inpatient treatment with unfractionated heparin. Vitamin K antagonists continued to be used after the initial treatment phase (241). From 2009 to 2013, the results of four clinical trials investigating the efficacy and safety of direct oral anticoagulants in the treatment of VTE were published (242-244). Direct oral anticoagulants are at least as effective as, and most likely safer than, sequential treatment with heparin or low molecular heparin and a vitamin K antagonist (245). Direct oral anticoagulants are recommended as first-line treatment for proximal DVT in non-cancer patients (12) and can also be used in patients with PE (23). Optimal treatment duration for a VTE event depends on what type of VTE event the patient has experienced (DVT versus PE), thrombus location (distal versus proximal), estimated risk of VTE recurrence, bleeding risk and patient preferences. Vena cava filters can be used in patients with proximal DVT or PE when anticoagulation is not possible (12, 23).

In selected patients with extensive, proximal DVT, catheter-directed thrombolysis, pharmacomechanical thrombolysis or surgical thrombectomy can
be considered as a means of reducing the risk of post-thrombotic syndrome (246). However, a recent American study that randomized patients with proximal DVT to treatment with anticoagulation alone or to treatment with pharmacomechanical thrombolysis in addition to anticoagulation showed no reduction in the risk of post-thrombotic syndrome and an increased risk of major bleeding in participants treated with pharmacomechanical thrombolysis (247).

Patients with PE and shock or hypotension should be considered for primary reperfusion therapy. Administration of systemic thrombolytic agents is the treatment of choice. In patients with contraindications to systemic thrombolysis or in whom systemic thrombolysis is unsuccessful, surgical embolectomy or percutaneous catheter-directed treatment can be considered (23). Studies that randomized hemodynamically stable patients with PE to receive or to not receive systemic thrombolysis showed that thrombolysis did not reduce the risk of death, but was associated with an increased risk of bleeding complications (248).

**Prognosis of VTE**

The prognosis of untreated PE is dismal. In a paper published in 1960, Barrit et al. report a trial where patients were randomized to receive or to not receive anticoagulants. The randomized part of the trial was stopped after an interim analysis showed that of 19 untreated patients, five had died and five had non-fatal recurrence of PE. Among 16 patients treated with anticoagulants, one patient had died from a cause other than PE, and no patient had had recurrent PE (249). More recently, the introduction of systemic thrombolysis has changed the prognosis of massive PE. A study from the 1990s of patients with massive PE was terminated prematurely as 100% of patients (n=4) randomized to systemic thrombolysis survived compared to 0% (n=4) of patients randomized to treatment with heparin alone (250).

In a setting where contemporary treatment for VTE was used, the 30-day case-fatality rate in PE was 10%, compared to 5% in DVT. The difference in case-fatality rate between PE and DVT decreased over time – the one-year case-fatality rate in PE was 23% compared to 21% in DVT (27). During the first three months after a PE, 21% of patients died, compared to 1.3% in a control group matched for age, sex and residency (251). The mortality in persons with a cancer-related VTE is much higher than that of persons with a non-cancer-related VTE (27, 46). In one study, the one-year case-fatality rate in patients with a first-time VTE was 13% in patients without cancer, compared to 63% in patients with cancer (27).
An important complication of VTE is chronic thromboembolic pulmonary hypertension (CTEPH). Theoretically, the cause of CTEPH is one or more pulmonary thrombi that do not resolve, giving rise to chronic obstruction of the pulmonary arteries with fibro-thrombotic material, changes in distal pulmonary vessels, high pulmonary vascular resistance and subsequent pulmonary hypertension and right-sided heart failure (252). The incidence of CTEPH after PE is 0.6% in all patients with PE, and about 3% in patients who survive at least three months after PE diagnosis (253). One study showed that a diagnosis of DVT also carries a risk of CTEPH, albeit lower than the risk seen in patients with PE (254). Symptoms of CTEPH are non-specific, but include dyspnea, edema of the lower extremities, fatigue and chest pain (252). The treatment of choice for CTEPH is surgery, but medical treatment with drugs used for pulmonary arterial hypertension can be considered in inoperable patients and patients with residual pulmonary arterial hypertension after surgical intervention (255).

The post-thrombotic syndrome is a complication of DVT, which is associated with a lower quality of life (256, 257). Symptoms of the post-thrombotic syndrome include lower extremity pains, cramps, heaviness, pruritus, paresthesia and venous ulcers (258).

The incidence of post-thrombotic syndrome varies depending on which definition of post-thrombotic syndrome that is used. A randomized controlled trial of the use of compression stockings to prevent post-thrombotic syndrome in persons with proximal DVT showed a cumulative incidence of post-thrombotic syndrome during 750 days of follow-up of about 13% using the Ginsberg criteria, and above 50% using the Villalta score for post-thrombotic syndrome (259). The Ginsberg criteria for post-thrombotic syndrome include pain and swelling of at least one month’s duration of a character typical for post-thrombotic syndrome (260), whereas the Villalta score is a point-based scoring system including both symptoms and clinical signs indicative of post-thrombotic syndrome. According to the Villalta score, persons with a venous ulcer are always defined as having post-thrombotic syndrome (261). Examples of predictors of a high Villalta score are older age, female sex, proximal location of the index DVT, previous DVT in the same extremity, and high body mass index (258).

It was previously thought that compression stockings could prevent post-thrombotic syndrome in patients with VTE (262, 263), but when this hypothesis was tested in a randomized placebo-controlled trial, no effect of compression stockings was seen (259). Compression stockings can be used to reduce symptomatic swelling in patients with proximal DVT. Exercise training does not aggravate the post-thrombotic syndrome, and may be beneficial. Surgical or
endovascular interventions can be considered for severely symptomatic patients (246).

**Recurrence of VTE**

The risk of VTE recurrence is highest during the first year after diagnosis, when it has been reported to be 8 per 100 person-years. After the first year, the VTE recurrence rate was 3 per 100 person-years. In that study, nine out of ten patients discontinued anticoagulant therapy at the latest by 12 months after the initial VTE event (46). The risk of VTE recurrence once anticoagulant treatment has been discontinued does not seem to be affected by treatment duration, as long as the patient is treated for at least three months (264). Extended treatment with anticoagulants reduces the risk of VTE recurrence (265, 266). If a person’s first episode of VTE is a DVT, a recurrent episode is likely to be a DVT too, and vice versa for PE (46). The likelihood of VTE recurrence varies depending on the context of the first episode of VTE, as depicted in Figure 3. Over 10 years of follow-up, the recurrence rate is highest in persons with cancer-related VTE, followed by unprovoked VTE and provoked VTE (267). The risk of VTE recurrence after a first episode of unprovoked VTE is about twice as high in men as in women, whereas the recurrence risk after a first episode of provoked VTE is equal in men and women (268). Due to this finding, male sex is a component of several clinical decision rules regarding indications for long-term treatment with anticoagulants to reduce the risk of VTE recurrence in patients with unprovoked VTE (269-271).

![Figure 3](image_url)

*Figure 3 The risk of venous thromboembolism (VTE) recurrence is lowest in patients with a transient provoking factor at the first VTE event, intermediate in patients with unprovoked VTE and highest in patients with VTE associated with a persistent provoking factor.*

**Risk factors and risk markers**

The concept of risk factors was popularized by the Framingham study, an early study of risk factors for coronary heart disease (272). There is no universally acknowledged definition of the term “risk factor” (273). Risk is defined as the
probability of an outcome within a population (274). A correlate is defined as a measurable characteristic associated with the risk of an outcome. Correlates are measured in subjects. Subjects are usually persons or animals, but can also be e.g. classes, workplaces, or communities. A characteristic can be a trait of the subject itself (age, blood pressure etc.) or of the context of the subject (e.g. number of VTE events in the subject’s first-grade relatives). The correlate can be present before, at the same time as, or after the outcome with which it is associated (274). Risk factors and risk markers are subtypes of correlates. A risk factor or risk marker is always present before the outcome occurs. The other two types of correlates are concomitants (the studied factor is present at the same time as the outcome) and consequences (the studied factor is present after the outcome) (274). These concepts are shown in Figure 4.

To distinguish risk factors or risk markers from concomitants and consequences, it is essential to measure the studied exposure before the outcome. Consequently, studies of risk factors or risk markers are best performed as prospective studies. The difference between risk markers and risk factors lies in whether or not they are part of the causal chain between the exposure and the outcome. A risk marker is not known to be part of the causal chain between the exposure and the outcome. Risk factors, on the other hand, are part of the causal chain or expose the subject to the causal chain (275). Risk factors may be modifiable (e.g. weight or blood pressure) or fixed (e.g. age or genetic factors).
Causality

Observational studies are well suited to detect associations, but are less appropriate to determine whether or not the observed associations are causal (276). In 1965, Sir Austin Bradford Hill asked the question “In what circumstances can we pass from this observed association to a verdict of causation? Upon what basis should we proceed to do so?” (277). He proposed nine criteria to help determine whether observed epidemiologic associations are causal. These criteria have later been questioned. They cannot be used to conclusively separate causal from non-causal relationships, as the complete causal mechanism of a disease includes many factors, some of which interact with each other. Most exposures are, by themselves, neither necessary, nor sufficient to cause a disease (278).

The only inarguable condition that must be fulfilled is that of temporality. The exposure must be present before the outcome, otherwise a causal relationship between the exposure and the outcome is impossible (278). In prospective
studies the exposure is always measured before the outcome occurs. A prospective study design thus ensures that the temporality criterion is met.

**Study designs**

One of the principles of evidence-based medicine is that all evidence is not of equal value. The evidence produced by different study designs is traditionally classified using a hierarchical system (279, 280), depicted in Figure 5.

![Figure 5 Level of evidence produced by different study designs.](Image)

Evidence belonging to a certain tier of the pyramid can be rated up or down using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system (281). It has been argued that meta-analyses and systematic reviews should be regarded as a lens through which other studies can be appraised, rather than automatically constituting the highest level of evidence (279). The quality of the studies included in a meta-analysis inevitably affect the value of the evidence presented in the meta-analysis.
The most reliable evidence supporting cause-effect relationships is that gained from randomized controlled trials (RCTs), a study design where one or more factors are altered and the effects of this alteration are examined. Study participants are randomly allocated to an intervention group or a control group (282). However, RCTs are not adapted to answer all research questions. Intervention studies may be unnecessary if the effect of an intervention is so pronounced that the likelihood of unknown confounding factors being responsible for the effect seen in observational studies is negligible. If the studied outcome is rare, RCTs are often not large enough to reliably detect the outcome. When studying long-term outcomes, RCTs are both costly and fraught with practical difficulties. For ethical reasons, RCTs cannot be conducted if the exposure studied is harmful. It would, for example, be unethical to randomize study participants to smoking ten cigarettes daily to study the effects of smoking on health. Even if the intervention studied is not yet known to be harmful or beneficial, physicians may be unwilling to recruit patients to an RCT if they, themselves, have a strong view on the subject. This means that conducting an RCT on an intervention that has already been introduced into clinical practice can be practically impossible because of low recruitment rates (283). RCTs may not always be the best method to evaluate treatments that require a high level of patient participation if the study cannot be blinded. If a study participant has a preference for one of the treatments, he or she is probably not motivated to participate fully if randomized to the other study treatment (284).

Some RCTs may have a low external validity. Highly experienced practitioners or specialist centers are more likely to participate in a trial of new procedures. Restrictive exclusion criteria mean that the patients eligible for study participation may only represent a small proportion of the patients treated for the condition (285). Among patients with atrial fibrillation deemed suitable for anticoagulant treatment, only 42% would have been eligible for participation in a trial of a novel oral anticoagulant (286). Even among eligible patients, those choosing to participate in an RCT probably differ from those who decline to participate. Participants in both the intervention group and the control group may receive better care compared to the care given in a non-study setting (287). Observational studies often have a higher external validity compared to RCTs, and observational methods can be used as an alternative to an RCT when an RCT is deemed inappropriate or impossible to conduct (285).

**Cohort studies**

Among observational studies, cohort studies generally provide the highest level of evidence. In a prospective cohort study, persons free of the study outcome are recruited to participate in the study and exposure to the studied variables is
measured. The participants are then followed for a period of time and all occurrences of the study outcome are recorded. This is shown in Figure 6.

Figure 6 Cohort study design. Participants free of the study outcome are shown in blue and participants with the study outcome are shown in orange.

Both the unexposed and the exposed group should be drawn from the same study population. As the exposure variables are measured before the outcome occurs there is no risk of confusing causes and effects. More than one outcome can be studied in the same cohort, in contrast to case-control studies where only one outcome is studied (the outcome on the basis of which the cases were included). If the outcome is rare, a large cohort and/or a long follow-up period is required. In such cases, a case-control design can be a better choice (288).
**Cohort effects**

In epidemiological research, a cohort effect is seen as an interaction between a period effect and age. A period effect is a widespread change in exposure on the population level. These changes in exposure can affect persons born in different years (i.e. persons of different ages at the time of the change in exposure) in distinct ways. The degree of change in exposure may either vary between persons of different ages, or the change in exposure may affect persons in a certain age group to a higher degree (289). In another interpretation of the cohort effect, being a member of a specific birth cohort can, in itself, affect the lives of the cohort members and lead to long-term health consequences (289).

When including participants with a wide range of birth years and/or conducting a study of long duration it is important to consider cohort effects when interpreting the study results.

**Reverse causality**

It is important to consider the possibility of reverse causality when interpreting the results of observational studies of the association between a risk marker and a disease. Reverse causality occurs when a premorbid condition alters the levels of a risk marker so that the risk marker seems to be associated with an adverse outcome (290). One approach to handle the issue of reverse causality could be to include only apparently healthy persons in the study. This has the disadvantage of lowering the study’s external validity.

**Confounding**

One of the limitations of observational studies is the risk of confounding. A confounding factor obscures the real relationship between the exposure and the outcome. Instead of evaluating the true association between an exposure and an outcome, the effect of a third factor, a confounding factor, is measured. Confounding occurs when the unexposed and exposed groups differ in other ways, apart from the difference in exposure to the studied exposure variable. For a factor to be a confounder, the factor must be associated with both the studied exposure variable and the outcome, and the factor must not be caused by the studied exposure (291). The distribution of a confounding factor differs between the exposed and the unexposed groups. Statistical tests, such as the $\chi^2$ test or the $t$ test, can be used to determine whether there are imbalances in the distribution of confounding factors between the exposed and the unexposed groups. However, the results of such tests must be interpreted with caution. In large study populations, small, irrelevant differences between the groups can be statistically significant. When the study groups are small, important differences between the groups may be statistically non-significant (292).
There are approaches to reduce confounding in observational studies. Restriction is one of these approaches. If cigarette smoking is thought to be a confounder for the studied association, the study population can be restricted to non-smokers. However, this lowers the external validity of the study results. Matching is a method used to reduce the effects of confounders in case-control studies. For example, if sex is a potential confounder, female controls are matched to female cases. Unfortunately, the effect of the matched variables on the outcome cannot be examined and it can be difficult to find suitable controls if the matching procedure is performed for many variables (293, 294).

Propensity score matching is a balancing method that can be used to reduce confounding. Each participant is given a propensity score that is based on baseline variables. Thus, in participants with the same propensity score the distribution of measured baseline variables is the same. In one propensity score method, pairs of exposed and unexposed participants with the same propensity score are formed and the association between the exposure variable and the outcome is assessed (295).

After the study is completed, stratification and multivariable techniques can be used to correct for confounding. In stratification, participants with and without, or with different levels of a confounding factor are analyzed separately. If the categories used for stratification are too broad, residual confounding can be an issue. For example, in a study setting where age is thought to be an important confounder and participants of all ages are included, stratification for age above or below 50 years can result in substantial residual confounding. Regression analyses are examples of multivariable techniques. An advantage of multivariable techniques, as opposed to stratification, is that many confounding factors can be handled simultaneously. All the aforementioned approaches to reduce confounding require that the confounder is known and measured. The effect of unknown or unmeasured confounders cannot be eliminated (293, 294).

An elegant example of the effect of different approaches to control for confounding is presented by Ziff et al. in a meta-analysis regarding the safety and efficacy of digoxin (296). In observational studies where no adjustment was made for confounding factors, treatment with digoxin was associated with death (relative risk 1.76). In observational studies where multivariable adjustment was made, the relative risk was 1.61. In propensity score matched studies the relative risk was 1.18, and in RCTs there was no increased risk of death in patients treated with digoxin (relative risk 0.99). The authors maintain that because digoxin is not first line treatment for any cardiac condition, digoxin is prescribed to sicker patients where the initial treatment has been unsuccessful. This phenomenon is called confounding by indication. Even in the adjusted and propensity score matched analyses it is likely that unmeasured confounders
cause an increase in all-cause mortality in digoxin-treated patients, and that this increase is then falsely attributed to the treatment with digoxin. In the RCTs, randomization ensures that unknown and unmeasured confounders are distributed equally in the treated and non-treated groups, and no association between treatment with digoxin and all-cause mortality is seen.

**Selection bias**

Selection bias is another limitation of observational studies. It is due to a systematic difference between the analyzed study population and the source population, i.e. the population to which the study results are to be applied. The association between the studied exposure and the outcome can differ between the analyzed subjects and the source population (297). These systematic differences can be introduced at different stages of a cohort study as shown in Figure 7 (288, 298).

![Diagram of study stages]

*Figure 7 Stages of a study during which selection bias can be introduced.*

In a case-control study selection bias can occur if there are systematic, unintended differences between cases and controls, other than the difference in outcome (276).
**Information bias**

Information bias is due to errors in the measurement of exposure or outcome. Such errors are referred to as misclassification (276). If the misclassification is non-differential, i.e. random, it does not result in information bias, but usually leads to bias towards the null. Differential misclassification occurs when the extent or direction of misclassification is dependent on the exposure or outcome status (i.e. the exposure and/or outcome is not measured in the same way for the exposed and the unexposed groups). Differential misclassification can lead to information bias causing either an under- or overestimation of the true association between the exposure and the outcome (299).
Objectives

The objectives of this thesis were: a) To describe the incidence of first-time VTE and the risk marker pattern at the time of VTE diagnosis, b) To determine the accuracy of VTE diagnoses in administrative registries, and c) To evaluate the association between diabetes, glucose levels, alcohol intake, physical activity and risk of first-time VTE.

The specific aims of each paper were:

I. To describe the incidence of first-time VTE in relation to age and sex, and to describe the risk marker pattern at the time of VTE diagnosis.

II. To estimate the positive predictive value of ICD diagnosis codes for first-time PE and DVT, and to estimate the proportion of valid PE and DVT events that are misclassified as other diseases.

III. To investigate the association between glucose levels, in the forms of diabetes, fasting plasma glucose and oral glucose tolerance, and risk of first-time VTE.

IV. To investigate the association between alcohol consumption, alcohol dependence and risk of first-time VTE in men and women separately.

V. To investigate the association between leisure time physical activity, occupational physical activity and risk of first-time VTE in men and women separately.
Materials and Methods

All studies were performed as cohort studies in Västerbotten County in northern Sweden. Västerbotten County had a population of 262,175 inhabitants in 2014 and is a sparsely populated region (4.8 inhabitants per square kilometer) (300). Healthcare is provided by one teaching hospital, two district general hospitals, and associated primary health care centers.

Study population – paper I

The study population in paper I consisted of all residents of Västerbotten County aged 18 and up (n=204,836 in 2006). The study participants were followed during the year 2006. Verified VTE events during the study period were recorded.

Study population – papers II to V

The Västerbotten Intervention Programme (VIP)

Since 1985, inhabitants of Västerbotten County have been invited to participate in a community intervention program – the VIP – aimed at reducing morbidity and mortality from cardiovascular disease and diabetes. The VIP started in the municipality of Norsjö in 1985, and was then successively implemented throughout the county. Västerbotten County inhabitants are invited to participate in the VIP when they reach 30, 40, 50, and 60 years of age (or, since 1996, 40, 50, and 60 years of age) (301).

The program offers participants a health examination and individual counseling about healthy lifestyle habits. The VIP is carried out at the participants’ primary health care centers by specially trained nurses (301). All VIP participants are asked if they are willing to participate in research. Blood samples and health examination data from participants who have given informed consent to participate in research are stored at the Department of Biobank Research at Umeå University, Umeå, Sweden.

The Venous thromboEmbolism In Northern Sweden (VEINS) study

Participants in the VIP from January 1, 1985 to September 5, 2014 who had given informed consent to participate in research were included in the VEINS study, the results of which are presented in papers II to V. Persons who had experienced a VTE event before participating in the VIP were excluded. VEINS participants were followed as a cohort from the health examination date until a first-time VTE event, emigration, death or the termination of follow-up on
September 5, 2014. Health examination data from each participant’s first health examination was used. A flow chart of the VEINS study population is shown in Figure 8.

Figure 8 Flow chart showing the Venous thromboEmbolism In Northern Sweden study population. VTE denotes venous thromboembolism.
Cases

Case identification – paper I
Potential VTE events were identified by a Swedish National Patient Registry search for ICD-10 diagnosis codes registered during inpatient, outpatient and emergency department visits in all three hospitals in Västerbotten County during the year 2006. The codes used were I26, I27.8, I27.9, I67.6, I80, I81, I82, O08.2, O08.7, O22.2, O22.3, O22.5, O22.8, O22.9, O87, and O88. We also searched the Swedish Cause of Death Registry to identify individuals whose death certificates listed the cause of death as a VTE event. To quantify the number of VTE cases that were not detected by the search method described above, we also searched the anticoagulant registry, the radiology registry and the diagnosis registry (ICD-10 diagnosis codes I74.3 and K55.0) at one of the three hospitals in the area.

In addition, a diagnosis registry search for a limited number of ICD-10 diagnosis codes (I26 and I80.1–I80.9) was made.

Case identification – papers II to V
Potential VTE events between January 1, 1985 and September 5, 2014 were identified by a diagnosis registry search for ICD diagnosis codes registered during inpatient, outpatient, emergency department, and primary health care visits in Västerbotten County. In paper II, ICD diagnosis codes registered at primary health care visits were not included. The Cause of Death Registry was searched for a subset of the ICD codes. The ICD diagnosis codes used are listed in Table 1.
Table 1. International Classification of Diseases (ICD) diagnosis codes used for identification of potential venous thromboembolism events in papers II to V

| ICD-8 | 321.00*; 321.09*  
|       | 426.02; 426.08; 426.09; 438.00; 438.99; 440.20; 450.01*†;  
|       | 450.02*; 450.03*†; 450.09*†; 451.00*; 451.98*‡; 451.99*‡;  
|       | 452.99*; 453.00; 453.99*  
|       | 631.00; 631.10; 631.11; 631.20; 634.50; 634.99; 642.00;  
|       | 642.20; 643.00; 643.20; 644.00; 644.20; 671.00; 671.01*‡;  
|       | 671.02*‡; 671.08*; 671.09*; 673.00; 673.10; 673.98*†; 673.99*;  
|       | 674.99; 677.98; 677.99  
| ICD-9 | 325X  
|       | 415A*; 415B*†; 416A; 416B; 416W; 416X; 437G*; 444C; 451A;  
|       | 451B*‡; 451C*‡; 451W*‡; 451X*‡; 452X*; 453A*; 453B; 453C*;  
|       | 453D*; 453W*; 453X*  
|       | 557A  
|       | 634G; 634H; 634W; 635G; 635H; 635W; 636G; 636H; 636W;  
|       | 637G; 637H; 637W; 638G; 638H; 638W; 639G; 639W; 639X;  
|       | 671C; 671D*‡; 671E*‡; 671F; 671W; 671X; 673A; 673B; 673C*‡;  
|       | 673D; 673W; 674A  
| ICD-10 | I26.0*‡; I26.9*‡; I27.8; I27.9; I27.6*; I74.3; I80.0; I80.1*‡;  
|       | I80.2*‡; I80.3*‡; I80.8*‡; I80.9*‡; I81.9*; I82.0*; I82.1; I82.2*;  
|       | I82.3*; I82.8*; I82.9*  
|       | K55.0  
|       | O08.2*; O08.7; O22.2; O22.3*‡; O22.5*; O22.8; O22.9; O87.0;  
|       | O87.1*‡; O87.2; O87.3*; O87.8; O87.9; O88.0; O88.1; O88.2*‡;  
|       | O88.3; O88.8  

*ICD diagnosis codes used for identification of potential venous thromboembolism cases in the Cause of Death Registry  
†ICD diagnosis codes used for calculation of positive predictive value for pulmonary embolism in paper II  
‡ICD diagnosis codes used for calculation of positive predictive value for deep vein thrombosis in paper II

Only first-time VTE events were included. To avoid inclusion of persons with recurrent VTE events, VIP participants with one or more of the ICD diagnosis codes used to identify potential VTE events registered before the VIP health examination were excluded from the study.

Validation of potential VTE events – papers I to V

All cases identified as potential VTE events were validated by review of medical records and/or radiology reports. A PE was considered to be verified if confirmed by pulmonary angiography, CT, magnetic resonance imaging, high probability ventilation-perfusion scan, or autopsy. A DVT of the lower or upper extremity, or an abdominal VTE was considered to be verified if confirmed by venography, ultrasonography, CT, magnetic resonance imaging, or autopsy. A
cerebral venous thrombosis was considered to be verified if confirmed by CT, magnetic resonance imaging or autopsy. For each individual only the first verified VTE event during the study period was considered. In paper I, a recurrence of VTE required objective evidence on an appropriate imaging scan of a new thrombosis not identified at previous imaging. In papers II to V, participants who had a VTE event registered as the main cause of death, and who underwent autopsy, were regarded as having a verified VTE.

**Classification of cases according to VTE location – papers I to III**

Individuals were classified according to VTE location. The groups used were PE, lower extremity DVT, upper extremity DVT (defined as thrombosis of the deep upper extremity veins, the axillary vein, the superior vena cava, the internal jugular vein, the subclavian vein, the innominate vein or the azygos vein), VTE in the veins of the abdomen and cerebral venous thrombosis. Individuals with a verified DVT and symptoms of PE were classified into the PE group. Individuals with multiple concurrent thrombosis locations were classified hierarchically into one of the following groups in descending order: 1: PE, 2: lower extremity DVT, 3: upper extremity DVT, 4: VTE in the veins of the abdomen or cerebral venous thrombosis.

A lower extremity DVT was classified as *iliac* when located at or above the inguinal ligament, as *femoropopliteal* when located in or above the popliteal vein, but distal to the inguinal ligament, and as a *calf DVT* when located below the popliteal vein. An event was classified as a *muscular vein thrombosis* if located in a muscular vein or a perforating vein.

**Risk markers for VTE at the time of VTE diagnosis – paper I**

All cases were characterized according to presence of absence of clinical risk markers for VTE at the time of VTE diagnosis. Information on risk markers was extracted from the participants’ medical records. The risk markers considered were: use of central venous lines, systemic hormone therapy (including contraceptives), recent hospitalization, recent surgery, recent immobilization for more than 48 hours, recent orthopedic cast or restraint, pregnancy, postpartum (defined as the period within 60 days of delivery), recent trauma and recent travel for more than eight hours. Recent events were defined as events occurring 60 days or less before the onset of VTE symptoms. Previous VTE events, family history of VTE in a first degree relative, presence of a coagulation disorder and use of pharmacological VTE prophylaxis were also recorded. Data on cancer in participants was extracted from the Swedish Cancer Registry.
Measurements and definitions – papers II to V

VIP participants answered a questionnaire regarding age, sex, education level, history of diabetes and cardiovascular disease, medication use, smoking, alcohol use and physical activity habits. The questionnaire was written in Swedish, the first language of the vast majority of the study participants. The questions were phrased so as to be easily understood and were similar to those used in the World Health Organization Monitoring Trends and Determinants in Cardiovascular Disease study. The questionnaire was optically read to minimize the risk of human error (302). The participants also underwent a health examination including measurements of height, weight, blood pressure, serum lipids, fasting plasma glucose and an oral glucose tolerance test.

Height, weight and body mass index

Height and weight in light clothing were measured. Body mass index was calculated as the weight in kilograms divided by the square of the height in meters.

Education level

Education level was categorized as more than secondary school education and secondary school education or less.

Smoking

Smoking habits were dichotomized into ever smokers, a category that included both current and previous smokers, and never smokers.

Blood pressure and hypertension

Blood pressure measurements were taken on the right mid-arm at the level of the heart with the participant in a supine position after at least five minutes rest. From September 1, 2009, all blood pressure measurements were made while the participant was in a sitting position. Blood pressure measurements taken with the participant in a sitting position were recalculated to make them comparable to measurement taken in a supine position (303). Hypertension was defined as systolic blood pressure ≥140 mmHg, diastolic blood pressure ≥90 mmHg, or self-reported treatment with antihypertensives.

Cancer

Data on cancer was extracted from the Swedish Cancer Registry which has a coverage rate of 96% (304). To account for delays in the registration of malignant diseases, cancer at health examination was defined as cancers registered before or up to two years after the health examination date.
Glucose levels and diabetes

Fasting plasma glucose and oral glucose tolerance test two-hour post-load plasma glucose were measured in capillary samples using point-of-care equipment that was calibrated regularly using a calibration scheme provided by the External Quality Assurance in Laboratory Medicine in Sweden (301).

Fasting plasma glucose was measured in all participants. An oral glucose tolerance test was performed in participants without known diabetes and fasting plasma glucose <7.0 mmol/L. The oral glucose tolerance test was performed according to World Health Organization recommendations with an oral glucose load of 75 grams (174). The participants were instructed to fast overnight before sampling. Fasting plasma glucose and oral glucose tolerance test two-hour post-load plasma glucose were assessed as continuous variables (per standard deviation) and as categorical variables using quartile boundaries as cut-off points. Diabetes was defined as self-reported diabetes, self-reported treatment with oral antidiabetics or insulin, fasting plasma glucose ≥7.0 mmol/L, or oral glucose tolerance test two-hour post-load plasma glucose ≥12.2 mmol/L.

Alcohol consumption

Alcohol consumption was assessed using a food frequency questionnaire. The participants answered the question: "How often do you consume the following products? Choose the category that matches your average consumption during the past year". The participant could choose one of nine different categories ranging from "never" to "four times per day or more" for each of the beverages light beer (1.8 weight percent of ethanol), medium strong beer (2.8 weight percent of ethanol), strong beer (4.5 weight percent of ethanol), wine (9.9 weight percent of ethanol) and liquor (32 weight percent of ethanol). The reported consumption frequencies were then converted to number of intakes per week and multiplied by an age- and sex-specific portion size to derive weekly intake of each beverage in grams (305). The weekly intake of each beverage in grams was multiplied with the weight percent of ethanol in that specific beverage to derive grams of ethanol consumed per week originating from that beverage. The ethanol intake from the five studied beverages was added up to get a total weekly ethanol intake. The total weekly alcohol intake in grams was then divided by 12 to get alcohol consumption in the unit "standard drinks per week". One standard drink was defined as 12 grams of ethanol. An alcohol intake of more than 70 standard drinks weekly was deemed improbable and intake in these participants (n=7) was set at 70 standard drinks weekly. Alcohol consumption was analyzed both as a continuous variable (number of standard drinks per week), and in categories with number of standard drinks in sex-specific quartiles. In an exploratory analysis, alcohol consumption was analyzed as a binary variable with three standard drinks per weeks as cut-off.
**Alcohol dependence**

The CAGE questionnaire was used to assess the presence of alcohol dependence (198). The questionnaire includes four questions: "Have you ever felt you should cut down on your drinking?", "Have people annoyed you by criticizing your drinking?", "Have you felt bad or guilty about your drinking?", and "Have you had a drink first thing in the morning after drinking the day before (eye-opener)?". Alcohol dependence was defined as two or more affirmative responses to the CAGE questions. All other participants were defined as not having alcohol dependence. Teetotalers were defined as persons with an affirmative response to the question "Are you a teetotaler?", or a person who answered "never" to the question "How often do you drink alcohol?". Teetotalers were categorized into the CAGE 0 group. Alcohol dependence was analyzed both as a binary variable and as an ordinal variable where the number of affirmative answers to the CAGE questionnaire was used to form categories.

**Leisure time physical activity**

Leisure time physical activity was assessed using the question: "How often have you trained or exercised in training clothes during the last three months to improve your fitness and/or to feel healthy?". Available categories were "never", "sometimes", "1–2 times per week" (expressed as "1 time per week" in some versions of the questionnaire), "2–3 times per week", and "more than 3 times per week". In our analyses, the categories "2–3 times per week" and "more than 3 times per week" were combined to form the category "more than 1–2 times per week". In an additional analysis, leisure time physical activity was dichotomized to form the categories "<1 time per week" and "1 time per week or more".

**Occupational physical activity**

Occupational physical activity was assessed using the question: "Choose the alternative that best describes your occupation". Available categories were "sedentary or standing", "light and partially active", "light and active", "sometimes involving strenuous physical activity” and ”involving strenuous physical activity for the majority of the time". In our analyses the categories "sedentary or standing” and "light and partially active” were combined to form the category "low occupational physical activity”. The categories “light and active”, ”sometimes involving strenuous physical activity” and ”involving strenuous physical activity for the majority of the time” were combined to form the category "high occupational physical activity”.

**Cambridge index**

A validated physical activity index, the Cambridge index, was used as a combined measure of leisure time physical activity and occupational physical
activity (306). Participants were categorized as inactive, moderately inactive, moderately active, and active. As the Cambridge index requires data on both leisure time physical activity and occupational physical activity, the number of participants with missing data was expected to be high. Therefore, a sensitivity analysis where participants who had answered only one of the physical activity questions were included in the analysis of the association between physical activity according to the Cambridge index and risk of VTE. These participants were given the lowest intensity score for the missing variable (either leisure time physical activity or occupational physical activity). Participants with missing data on both physical activity variables were not included in the sensitivity analysis.

**Statistical analyses**

In all publications the number and proportion and the mean and standard deviation were used to describe baseline characteristics. The $\chi^2$ test was used to test for differences between groups in the distribution of categorical variables. For proportions, 95% CIs were calculated using the normal approximation to the binomial distribution.

In paper II, the positive predictive value was calculated as the number of participants with a validated VTE diagnosis divided by the number of participants with an ICD code (listed in Table 1) indicating PE or DVT. The proportion of misclassified VTE events was calculated as the number of valid PE or DVT events not identified by the ICD diagnosis codes indicating PE or DVT divided by the total number of valid PE or DVT events.

In papers III to V, Spearman’s rank-order correlation was used to evaluate relationships between variables. Cox proportional hazards regression was used to calculate HRs and 95% CIs for the association between the studied risk markers and VTE. For categorical variables, the proportional hazards assumption was tested using Kaplan-Meier plots and/or plots of the log-minus-log of the survival function with separate curves for each level of the categorical variables. For continuous variables, the proportional hazards assumption was tested by creating plots of the partial residuals versus survival time. The different Cox regression models used are shown in Table 2.
Table 2. Cox regression models used in papers III to V

<table>
<thead>
<tr>
<th></th>
<th>Paper III</th>
<th>Paper IV</th>
<th>Paper V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univariable model</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model adjusted for age and sex</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model stratified by sex, adjusted for age</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Model adjusted for age, sex, body mass index, hypertension, smoking, education level and cancer</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Model stratified by sex, adjusted for age, body mass index, hypertension, smoking, education level and cancer</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Only participants with complete data on all adjustment variables included in the fully adjusted model were included in the univariable and age- and/or sex-adjusted models. Tests for trend were performed by entering ordinal categorical variables as continuous variables in the regression models. Two-sided P-values < 0.05 were considered significant. Multiplicative and additive interaction analyses were performed. As a sensitivity analysis, a nested, matched case-control analysis of the association between alcohol intake, alcohol dependence and risk of VTE was performed. Four controls from the VEINS study population were matched to each case. Cases and controls were matched for age, sex, health examination date and primary health care center.

A power calculation showed that given a sample size of 108,000 participants (54,000 men and 54,000 women), a cumulative incidence of VTE of 1.9% during the study period, a power of 80% and a significance level of 0.05 (two-sided), a HR of ≥1.3 for increased risk of VTE or ≤0.8 for decreased risk of VTE could be detected when categorizing participants into four groups of equal size and analyzing men and women separately.

Ethical considerations

At the time of the VIP health examination all participants provided written informed consent for participation in future research. Persons included in the VEINS study were not subjected to any additional medical procedures. However, validation of VTE events required that the study investigator accessed the participants’ medical and radiology records. This can be considered a violation of the participants’ right to privacy and could potentially cause the participants distress. All living individuals with VTE included in the studies received a letter explaining the planned studies in detail. In the letter, contact information for study investigators was provided and the participants could consult the investigators if they had any questions or concerns about the study,
or if they did not consent to participate. They were also informed about the right to have inaccurate information corrected and each individual’s right to get a copy of the data concerning him/her.

In the datasets used, each participant with VTE was given a code number. The key linking code numbers to personal identity numbers was kept in a separate, secure location. Neither datasets nor the key linking code numbers to personal identity numbers could be accessed by unauthorized persons. All results were presented on group level and participants’ identities were kept anonymous.

The studies were approved by the Regional Ethics Review Board, Umeå, Sweden (approval number 06-162M §157/06 with amendments).
Results

Accuracy of VTE diagnoses in administrative registries

In paper I, the diagnosis registry ICD codes used for case finding had a positive predictive value of 61% for valid VTE diagnoses. Other methods (a search for VTE diagnoses in the Swedish Cause of Death Registry, the anticoagulant registry, the radiology registry and a diagnosis registry search for additional ICD diagnosis codes – I74.3 and K55.0) identified 30 individuals with VTE in addition to the 313 identified by the diagnosis registry search. Accordingly, the diagnosis registry search had a false negative rate of 9%. The limited diagnosis registry search, including only ICD-10 diagnosis codes I26 and I80.1–I80.9, resulted in a positive predictive value of 64% and a 22% false negative rate for identifying patients with PE and/or DVT of the upper or lower extremities.

In paper II, the accuracy of VTE diagnoses in the Swedish National Patient Registry and Cause of Death Registry was evaluated. The VEINS study population was used in this paper. Thus, individuals below the age of 30 were not included. A search for a limited set of ICD diagnosis codes (indicated in Table 1) was used for calculations of the positive predictive value for PE and DVT diagnoses in these registries. The remaining ICD diagnosis codes, listed in Table 1, were used to identify additional DVT and PE events, termed “misclassified” events. Flow-charts of the study population are shown in Figure 9.
Figure 9 Flow charts showing the study population in Paper II. The flow chart for the analyses regarding the positive predictive value of pulmonary embolism (PE) and deep vein thrombosis (DVT) diagnoses in administrative registries is shown in blue. The flow chart regarding the analyses of misclassified PE and DVT events is shown in green.
Table 3 shows the positive predictive value of diagnoses of PE and DVT in the Swedish National Patient Registry and Cause of Death Registry.

**Table 3.** Positive predictive value of diagnoses of pulmonary embolism (PE) and deep vein thrombosis (DVT) in the Swedish National Patient Registry and Cause of Death Registry during the years 1985 to 2014

<table>
<thead>
<tr>
<th></th>
<th>Number of valid VTE cases</th>
<th>Participants with diagnosis codes indicating a VTE event</th>
<th>Positive predictive value (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE or DVT</td>
<td>1,771</td>
<td>2,450</td>
<td>72.3 (70.3–74.1)</td>
</tr>
<tr>
<td>PE</td>
<td>934</td>
<td>1,158</td>
<td>80.7 (78.4–82.9)</td>
</tr>
<tr>
<td>DVT</td>
<td>885</td>
<td>1,495</td>
<td>59.2 (56.7–61.7)</td>
</tr>
</tbody>
</table>

VTE, venous thromboembolism.

After restricting the analysis to the years 2009 to 2014 the positive predictive value for PE was higher (85.8%; 95% CI 83.1–88.5). For DVT the positive predictive value was slightly lower (54.1%; 95% CI 50.5–57.7), although the CI overlapped with that for the period 1985 to 2014.

The proportion of misclassified VTE events in the Swedish National Patient Registry and the Cause of Death Registry is shown in Table 4.

**Table 4.** Percentage of misclassified events of pulmonary embolism (PE) and deep vein thrombosis (DVT) in relation to the total number of valid PE or DVT events in the Swedish National Patient Registry and Cause of Death Registry during the years 1985 to 2014

<table>
<thead>
<tr>
<th></th>
<th>Misclassified events</th>
<th>Total number of valid events</th>
<th>Percent misclassified (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE or DVT</td>
<td>180</td>
<td>1,951</td>
<td>9.2 (7.9–10.5)</td>
</tr>
<tr>
<td>PE</td>
<td>10</td>
<td>944</td>
<td>1.1 (0.4–1.7)</td>
</tr>
<tr>
<td>DVT</td>
<td>174</td>
<td>1,059</td>
<td>16.4 (14.2–18.7)</td>
</tr>
</tbody>
</table>

The proportions of misclassified PE and DVT events were slightly lower after restricting the analysis to the years 2009 to 2014 (0.9%; 95% CI 0.0–1.7 for PE and 13.2%; 95% CI 10.0–16.3 for DVT).
Incidence of first-time VTE in relation to age and sex

As described in paper I, a first-time VTE event was verified for 281 residents of Västerbotten County during the year 2006. The resulting incidence of first-time VTE was 137 (95% CI 122–154) per 100,000 person-years. The mean age at first-time VTE was 70 years (95% CI 68–72 years) and 46% were male. The incidence of first-time VTE in men and women in different age groups is shown in Figure 10.

Figure 10 Incidence of venous thromboembolism in different age groups. The incidence in men is shown in blue and the incidence in women is shown in orange.
Risk markers at first-time VTE diagnosis

Patient characteristics for patients with first-time VTE are shown in Table 5.

Table 5. Risk markers for venous thromboembolism (VTE) in 281 patients, 128 men and 153 women, with first-time VTE

<table>
<thead>
<tr>
<th>Risk marker</th>
<th>Number of participants</th>
<th>Percentage of participants (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recent hospitalizationa</td>
<td>120</td>
<td>43 (37–49)</td>
</tr>
<tr>
<td>Recent surgerya</td>
<td>42</td>
<td>15 (11–19)</td>
</tr>
<tr>
<td>Recent immobilization &gt;48 hoursa</td>
<td>42</td>
<td>15 (11–19)</td>
</tr>
<tr>
<td>Recent traumaa</td>
<td>15</td>
<td>5 (3–8)</td>
</tr>
<tr>
<td>Recent orthopedic casta</td>
<td>5</td>
<td>2 (0.2–3)</td>
</tr>
<tr>
<td>Recent travel &gt;8 hoursa</td>
<td>4</td>
<td>1 (0–3)</td>
</tr>
<tr>
<td>Concurrent malignancy</td>
<td>88</td>
<td>31 (26–37)</td>
</tr>
<tr>
<td>Malignancy diagnosed ≤2 years after the VTE event</td>
<td>7</td>
<td>2 (1–4)</td>
</tr>
<tr>
<td>Central venous line</td>
<td>27</td>
<td>10 (6–13)</td>
</tr>
<tr>
<td>VTE in first-degree relative</td>
<td>7</td>
<td>2 (1–4)</td>
</tr>
<tr>
<td>Coagulation disorder</td>
<td>7</td>
<td>2 (1–4)</td>
</tr>
<tr>
<td>Hormone therapy</td>
<td>7</td>
<td>5 (1–8)b</td>
</tr>
<tr>
<td>Pregnancy</td>
<td>0</td>
<td>0b</td>
</tr>
<tr>
<td>Postpartum</td>
<td>1</td>
<td>1 (0–2)b</td>
</tr>
</tbody>
</table>

\(^{a}≤60\) days before the VTE event \(^{b}\)Percentage of female participants

Among the participants with first-time VTE, 93 (33%; 95% CI 28–39) had none of the studied risk markers for VTE at diagnosis. The proportion of participants without any of the studied risk markers for VTE increased with age. The percentage of participants with first-time VTE without risk markers for VTE at VTE diagnosis was 26% (95% CI 17–35) among participants aged 18 to 64 years, 28% (95% CI 19–37) among participants aged 65 to 79 years and 46% (95% CI 36–56) among participants aged 80 years and above.

The VEINS study population

A total of 108,025 individuals were included in the VEINS cohort and were followed for 1,496,669 person-years. Women composed 51% of the study population. Baseline characteristics for the entire cohort are shown in Table 6.
**Table 6.** Baseline characteristics of the Venous thromboEmbolism In Northern Sweden study cohort (n=108,025)

<table>
<thead>
<tr>
<th></th>
<th>Men (n=53,393)</th>
<th>Women (n=54,632)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>46.3 (9.3)</td>
<td>46.3 (9.2)</td>
</tr>
<tr>
<td>Body mass index$^a$, kg/m$^2$</td>
<td>26.3 (3.8)</td>
<td>25.4 (4.6)</td>
</tr>
<tr>
<td>Hypertension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No hypertension</td>
<td>36,065 (67.5)</td>
<td>39,799 (72.8)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>16,522 (30.9)</td>
<td>13,866 (25.4)</td>
</tr>
<tr>
<td>Missing</td>
<td>806 (1.5)</td>
<td>967 (1.8)</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never smoker</td>
<td>25,389 (47.6)</td>
<td>26,730 (48.9)</td>
</tr>
<tr>
<td>Ever smoker</td>
<td>26,860 (50.3)</td>
<td>27,011 (49.4)</td>
</tr>
<tr>
<td>Missing</td>
<td>1,144 (2.1)</td>
<td>891 (1.6)</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary school or less</td>
<td>40,117 (75.1)</td>
<td>36,513 (66.8)</td>
</tr>
<tr>
<td>Above secondary school</td>
<td>12,368 (23.2)</td>
<td>17,123 (31.3)</td>
</tr>
<tr>
<td>Missing</td>
<td>908 (1.7)</td>
<td>996 (1.8)</td>
</tr>
<tr>
<td>Cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No cancer</td>
<td>52,392 (98.1)</td>
<td>52,871 (96.8)</td>
</tr>
<tr>
<td>Cancer</td>
<td>1,001 (1.9)</td>
<td>1,761 (3.2)</td>
</tr>
<tr>
<td>Fasting plasma glucose$^b$, mmol/L</td>
<td>5.5 (1.1)</td>
<td>5.4 (0.9)</td>
</tr>
<tr>
<td>2HPG$^c$, mmol/L</td>
<td>6.3 (1.6)</td>
<td>6.8 (1.5)</td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No diabetes</td>
<td>50,061 (93.8)</td>
<td>51,892 (95.0)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>2,585 (4.8)</td>
<td>1,905 (3.5)</td>
</tr>
<tr>
<td>Missing</td>
<td>747 (1.4)</td>
<td>835 (1.5)</td>
</tr>
<tr>
<td>Alcohol intake$^d$, standard drinks per week</td>
<td>3.5 (3.6)</td>
<td>1.5 (1.8)</td>
</tr>
<tr>
<td>Alcohol dependence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No alcohol dependence</td>
<td>45,142 (84.5)</td>
<td>49,686 (90.9)</td>
</tr>
<tr>
<td>Alcohol dependence</td>
<td>5,329 (10.0)</td>
<td>1,749 (3.2)</td>
</tr>
<tr>
<td>Missing</td>
<td>2,922 (5.5)</td>
<td>3,197 (5.9)</td>
</tr>
<tr>
<td>Leisure time physical activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>20,163 (37.8)</td>
<td>20,037 (36.7)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>12,676 (23.7)</td>
<td>12,423 (22.7)</td>
</tr>
<tr>
<td>1–2 times/week</td>
<td>7,442 (13.9)</td>
<td>9,105 (16.7)</td>
</tr>
<tr>
<td>&gt;1–2 times/week</td>
<td>9,500 (17.8)</td>
<td>9,399 (17.2)</td>
</tr>
<tr>
<td>Missing</td>
<td>3,612 (6.8)</td>
<td>3,668 (6.7)</td>
</tr>
<tr>
<td>Occupational physical activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>21,785 (40.8)</td>
<td>20,361 (37.3)</td>
</tr>
<tr>
<td>High</td>
<td>26,456 (49.5)</td>
<td>28,621 (52.4)</td>
</tr>
<tr>
<td>Missing</td>
<td>5,152 (9.6)</td>
<td>5,650 (10.3)</td>
</tr>
<tr>
<td>Cambridge index</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>7,576 (14.2)</td>
<td>7,164 (13.1)</td>
</tr>
<tr>
<td>Moderately inactive</td>
<td>15,038 (28.2)</td>
<td>15,634 (28.6)</td>
</tr>
<tr>
<td>Moderately active</td>
<td>13,925 (26.1)</td>
<td>13,988 (25.6)</td>
</tr>
<tr>
<td>Active</td>
<td>11,152 (20.9)</td>
<td>11,651 (21.3)</td>
</tr>
<tr>
<td>Missing</td>
<td>5,702 (10.7)</td>
<td>6,195 (11.3)</td>
</tr>
</tbody>
</table>

Values are shown as mean (standard deviation) or numbers (percentages)

2HPG, Oral glucose tolerance test two-hour post-load plasma glucose

$^a$516 men and 657 women had missing data for body mass index, $^b$859 men and 928 women had missing data for fasting plasma glucose, $^c$Data on 2HPG refer to the 51,072 men and 53,058 women without self-reported diabetes and with fasting plasma glucose <7.0 mmol/L. Among these
participants, 2,260 men and 2,749 women had missing data for 2HPG. 4,976 men and 5,250 women had missing data for alcohol intake.

Correlations between the studied risk markers for VTE are shown in table 7.

Table 7. Correlations between the studied risk markers

<table>
<thead>
<tr>
<th></th>
<th>FPG</th>
<th>2HPG</th>
<th>Diabetes</th>
<th>Alcohol intake</th>
<th>Alcohol depend.</th>
<th>Leisure PA</th>
<th>Occup. PA</th>
<th>Cambridge index</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPG</td>
<td>1</td>
<td></td>
<td></td>
<td>0.29&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.29&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.01&lt;sup&gt;*&lt;/sup&gt;</td>
<td>0.01&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.09&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>2HPG</td>
<td></td>
<td>1</td>
<td></td>
<td>0.16&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.13&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.05&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.08&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.02&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Diabetes</td>
<td>0.29&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>1</td>
<td>-0.04&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.004</td>
<td>0.05&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.001</td>
<td>-0.03&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Alcohol intake</td>
<td>0.01&lt;sup&gt;*&lt;/sup&gt;</td>
<td>-0.13&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.04&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1</td>
<td>0.23&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.09&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.12&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.03&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Alcohol depend.</td>
<td>0.01&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.05&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.004</td>
<td>0.23&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1</td>
<td>0.005</td>
<td>-0.02&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.004</td>
</tr>
<tr>
<td>Leisure PA</td>
<td>-0.09&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.08&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.05&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.09&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.005</td>
<td>1</td>
<td>-0.07&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.55&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Occup. PA</td>
<td>0.01&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.02&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.001</td>
<td>-0.12&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.02&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.07&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1</td>
<td>0.49&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cambridge index</td>
<td>-0.04&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.08&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.03&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.03&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.004</td>
<td>0.55&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.49&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1</td>
</tr>
</tbody>
</table>

FPG, fasting plasma glucose
2HPG, oral glucose tolerance test two-hour post-load plasma glucose
Alcohol depend., alcohol dependence
Leisure PA, leisure time physical activity
Occup. PA, occupational physical activity
*P <0.05, bP <0.01

During the follow-up period, 2,054 participants had a first-time VTE event with a median time-to-event of 12.6 years. Locations of the VTE events in the 2,054 participants with first-time VTE are shown in Figure 11. There was no difference between men and women in the distribution of thrombosis locations (p=0.25).
Figure 11 Thrombosis locations in 2,054 participants of the Venous thromboEmbolism In Northern Sweden study with first-time venous thromboembolism (VTE). DVT denotes deep vein thrombosis.

For comparison, the distribution of VTE locations in patients with first-time VTE in Paper I is shown in Figure 12.

Figure 12 Location of venous thromboembolism (VTE) events in 281 patients with first-time VTE in 2006. A concurrent verified lower extremity deep vein thrombosis (DVT) was found in 19 patients with first-time pulmonary embolism and two patients with verified pulmonary embolism had symptoms and clinical signs of a lower extremity DVT.
In the VEINS study, over half of the patients with PE had bilateral emboli. Among the 1,004 participants with PE, 161 participants (16.0%) had a verified lower extremity DVT, and a further 33 participants (3.3%) had symptoms or clinical signs of lower extremity DVT. Among the participants with PE, 14 (1.4%) had a verified upper extremity DVT and 15 (1.5%) had an abdominal VTE.

Among the 874 participants with lower extremity DVT without concurrent PE, the most common location (54.4%) was a unilateral DVT in the left lower extremity. The DVT was iliac in 201 (23.0%) cases, femoropopliteal in 448 (51.3%) cases and located in the calf in 169 (19.3%) cases. An isolated muscular vein thrombosis was found in 52 (5.9%) cases, and in 4 cases, data on thrombus location was missing. Among the participants with lower extremity DVT, 15 (1.7%) had a concurrent abdominal VTE, most commonly due to an extension of the lower extremity DVT into the inferior vena cava.

**Risk marker profile at first-time VTE diagnosis**

Data on risk markers for VTE at VTE diagnosis was collected for the 1,447 participants diagnosed with first-time VTE between 2006 and 2014. This equals 70.4% of the total number of cases with first-time VTE in the VEINS study. The risk marker profile at VTE diagnosis is shown in Table 8.

**Table 8.** Risk markers for venous thromboembolism (VTE) in 1,447 patients, 771 men and 676 women, with first-time VTE diagnosed between 2006 and 2014

<table>
<thead>
<tr>
<th>Risk Marker</th>
<th>Number of Participants</th>
<th>Percentage of Participants (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recent hospitalization&lt;sup&gt;a&lt;/sup&gt;</td>
<td>627</td>
<td>43 (41–46)</td>
</tr>
<tr>
<td>Recent surgery&lt;sup&gt;a&lt;/sup&gt;</td>
<td>313</td>
<td>22 (20–24)</td>
</tr>
<tr>
<td>Recent immobilization &gt;48 hours&lt;sup&gt;a&lt;/sup&gt;</td>
<td>188</td>
<td>13 (11–15)</td>
</tr>
<tr>
<td>Recent trauma&lt;sup&gt;a&lt;/sup&gt;</td>
<td>102</td>
<td>7 (6–8)</td>
</tr>
<tr>
<td>Recent orthopedic cast&lt;sup&gt;a&lt;/sup&gt;</td>
<td>49</td>
<td>3 (3–4)</td>
</tr>
<tr>
<td>Recent travel &gt;8 hours&lt;sup&gt;a&lt;/sup&gt;</td>
<td>31</td>
<td>2 (1–3)</td>
</tr>
<tr>
<td>Malignancy diagnosed before VTE</td>
<td>566</td>
<td>39 (37–42)</td>
</tr>
<tr>
<td>Malignancy diagnosed at, or ≤2 years after, the VTE event</td>
<td>79</td>
<td>5 (4–7)</td>
</tr>
<tr>
<td>Central venous line</td>
<td>299</td>
<td>21 (19–23)</td>
</tr>
<tr>
<td>VTE in first-degree relative</td>
<td>78</td>
<td>5 (4–7)</td>
</tr>
<tr>
<td>Coagulation disorder</td>
<td>101</td>
<td>7 (6–8)</td>
</tr>
<tr>
<td>Hormone therapy</td>
<td>30</td>
<td>4&lt;sup&gt;b&lt;/sup&gt; (3–6)</td>
</tr>
<tr>
<td>Pregnancy</td>
<td>1</td>
<td>0.1&lt;sup&gt;b&lt;/sup&gt; (0–0.4)</td>
</tr>
<tr>
<td>Postpartum</td>
<td>2</td>
<td>0.3&lt;sup&gt;b&lt;/sup&gt; (0–0.7)</td>
</tr>
</tbody>
</table>

<sup>a</sup>≤60 days before the VTE event, <sup>b</sup>Percentage of female participants
Among the participants with first-time VTE, 332 (23%; 95% CI 21–25%) had none of the studied risk markers for VTE listed in Table 8.

**Glucose levels, diabetes and risk of first-time VTE**

In univariable analysis, there was an association between diabetes (HR 1.71; 95% CI 1.41–2.07), fasting plasma glucose ($P$ for trend <0.001 over quartiles of fasting plasma glucose), oral glucose tolerance test two-hour post-load plasma glucose ($P$ for trend 0.003 over quartiles) and risk of VTE. These associations were attenuated after adjustment for age and sex. After adjustment for a range of potential confounders, including body mass index, there were no significant associations between diabetes, fasting plasma glucose, oral glucose tolerance test two-hour post-load plasma glucose and risk of VTE (data not shown). The results of sex-stratified Cox regression analyses of the associations between diabetes, fasting plasma glucose, oral glucose tolerance test two-hour post-load plasma glucose and VTE are shown in Table 9.
Table 9. Associations between diabetes, fasting plasma glucose (FPG) (in mmol/L), oral glucose tolerance test two-hour post-load plasma glucose (2HPG) (in mmol/L) and the risk of venous thromboembolism in men and women

<table>
<thead>
<tr>
<th></th>
<th>Events</th>
<th>Model 1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Model 2&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No diabetes</td>
<td>959</td>
<td>1 (ref.)</td>
<td>1 (ref.)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>59</td>
<td>1.02 (0.78–1.32)</td>
<td>0.93 (0.71–1.21)</td>
</tr>
<tr>
<td>FPG&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;4.90</td>
<td>179</td>
<td>1 (ref.)</td>
<td>1 (ref.)</td>
</tr>
<tr>
<td>4.90–5.29</td>
<td>185</td>
<td>0.87 (0.71–1.07)</td>
<td>0.85 (0.69–1.05)</td>
</tr>
<tr>
<td>5.30–5.69</td>
<td>255</td>
<td>1.08 (0.89–1.31)</td>
<td>1.04 (0.86–1.27)</td>
</tr>
<tr>
<td>≥5.70</td>
<td>367</td>
<td>1.16 (0.97–1.39)</td>
<td>1.07 (0.89–1.29)</td>
</tr>
<tr>
<td>P for trend</td>
<td></td>
<td>0.01</td>
<td>0.11</td>
</tr>
<tr>
<td>FPG per SD&lt;sup&gt;d&lt;/sup&gt;</td>
<td>986</td>
<td>1.02 (0.96–1.08)</td>
<td>0.99 (0.94–1.05)</td>
</tr>
<tr>
<td>2HPG&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;5.60</td>
<td>290</td>
<td>1 (ref.)</td>
<td>1 (ref.)</td>
</tr>
<tr>
<td>5.60–6.39</td>
<td>209</td>
<td>0.89 (0.74–1.06)</td>
<td>0.90 (0.76–1.08)</td>
</tr>
<tr>
<td>6.40–7.29</td>
<td>216</td>
<td>0.89 (0.75–1.06)</td>
<td>0.90 (0.76–1.08)</td>
</tr>
<tr>
<td>≥7.30</td>
<td>218</td>
<td>0.88 (0.73–1.05)</td>
<td>0.85 (0.71–1.02)</td>
</tr>
<tr>
<td>P for trend</td>
<td></td>
<td>0.16</td>
<td>0.09</td>
</tr>
<tr>
<td>2HPG per SD&lt;sup&gt;d&lt;/sup&gt;</td>
<td>933</td>
<td>0.96 (0.90–1.02)</td>
<td>0.95 (0.89–1.01)</td>
</tr>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No diabetes</td>
<td>825</td>
<td>1 (ref.)</td>
<td>1 (ref.)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>52</td>
<td>1.37 (1.03–1.81)</td>
<td>1.11 (0.83–1.48)</td>
</tr>
<tr>
<td>FPG&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;4.90</td>
<td>182</td>
<td>1 (ref.)</td>
<td>1 (ref.)</td>
</tr>
<tr>
<td>4.90–5.29</td>
<td>203</td>
<td>1.02 (0.83–1.24)</td>
<td>1.00 (0.82–1.22)</td>
</tr>
<tr>
<td>5.30–5.69</td>
<td>208</td>
<td>1.02 (0.83–1.24)</td>
<td>0.97 (0.79–1.18)</td>
</tr>
<tr>
<td>≥5.70</td>
<td>266</td>
<td>1.22 (1.01–1.48)</td>
<td>1.08 (0.89–1.31)</td>
</tr>
<tr>
<td>P for trend</td>
<td></td>
<td>0.04</td>
<td>0.49</td>
</tr>
<tr>
<td>FPG per SD&lt;sup&gt;d&lt;/sup&gt;</td>
<td>859</td>
<td>1.08 (1.02–1.15)</td>
<td>1.03 (0.97–1.10)</td>
</tr>
<tr>
<td>2HPG&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;5.60</td>
<td>118</td>
<td>1 (ref.)</td>
<td>1 (ref.)</td>
</tr>
<tr>
<td>5.60–6.39</td>
<td>164</td>
<td>0.91 (0.72–1.16)</td>
<td>0.93 (0.73–1.18)</td>
</tr>
<tr>
<td>6.40–7.29</td>
<td>213</td>
<td>0.86 (0.69–1.08)</td>
<td>0.87 (0.69–1.09)</td>
</tr>
<tr>
<td>≥7.30</td>
<td>300</td>
<td>0.98 (0.79–1.22)</td>
<td>0.92 (0.74–1.15)</td>
</tr>
<tr>
<td>P for trend</td>
<td></td>
<td>0.96</td>
<td>0.52</td>
</tr>
<tr>
<td>2HPG per SD&lt;sup&gt;d&lt;/sup&gt;</td>
<td>795</td>
<td>0.99 (0.93–1.06)</td>
<td>0.96 (0.90–1.03)</td>
</tr>
</tbody>
</table>

SD, standard deviation

Only participants with complete data on all adjustment variables are presented. Associations are shown as hazard ratios with 95% confidence intervals.

<sup>a</sup>Model adjusted for age.
<sup>b</sup>Model adjusted for age, body mass index, hypertension, smoking, education level and cancer.
<sup>c</sup>Participants with self-reported diabetes were excluded.
<sup>d</sup>One SD of FPG was 0.81 mmol/L.
<sup>e</sup>Participants with self-reported diabetes or FPG ≥ 7.0 were excluded.
<sup>f</sup>One SD of 2HPG was 1.54 mmol/L.
Alcohol intake, alcohol dependence and risk of first-time VTE

The types of alcoholic beverages consumed differed between men and women, as shown in Figure 13. For men, beer constituted more than half of the total alcohol intake. For women, wine was the main source of alcohol.

![Pie chart showing alcohol intake by sex](image)

*Figure 13 The percentage of total alcohol intake consumed as beer, wine and liquor. Analysis stratified by sex.*
The distribution of VEINS participant characteristics according to sex-specific quartiles of alcohol intake is shown in Table 10.

**Table 10.** Distribution of participant characteristics according to sex-specific quartiles of alcohol intake

<table>
<thead>
<tr>
<th></th>
<th>Quartile 1&lt;sup&gt;a&lt;/sup&gt; (n=24,756)</th>
<th>Quartile 2&lt;sup&gt;b&lt;/sup&gt; (n=24,466)</th>
<th>Quartile 3&lt;sup&gt;c&lt;/sup&gt; (n=24,348)</th>
<th>Quartile 4&lt;sup&gt;d&lt;/sup&gt; (n=24,229)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>47.9 (9.6)</td>
<td>46.1 (9.1)</td>
<td>45.6 (8.7)</td>
<td>45.9 (9.0)</td>
</tr>
<tr>
<td>BMI, kg/m&lt;sup&gt;2&lt;/sup&gt;</td>
<td>26.4 (4.7)</td>
<td>26.1 (4.3)</td>
<td>25.7 (4.0)</td>
<td>25.3 (3.8)</td>
</tr>
<tr>
<td>Hypertensive</td>
<td>8,183 (33.1)</td>
<td>6,800 (27.8)</td>
<td>6,150 (25.3)</td>
<td>6,015 (24.8)</td>
</tr>
<tr>
<td>Ever smoker</td>
<td>9,708 (39.2)</td>
<td>12,049 (49.2)</td>
<td>12,834 (52.7)</td>
<td>13,770 (56.8)</td>
</tr>
<tr>
<td>&gt; secondary school</td>
<td>5,167 (20.9)</td>
<td>5,831 (23.8)</td>
<td>7,363 (30.2)</td>
<td>9,446 (39.0)</td>
</tr>
<tr>
<td>education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>778 (3.1)</td>
<td>619 (2.5)</td>
<td>573 (2.4)</td>
<td>567 (2.3)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>1,487 (6.0)</td>
<td>1,061 (4.3)</td>
<td>828 (3.4)</td>
<td>756 (3.1)</td>
</tr>
<tr>
<td>Leisure time PA ≥1</td>
<td>6,962 (28.1)</td>
<td>7,994 (32.7)</td>
<td>9,179 (37.7)</td>
<td>9,564 (39.5)</td>
</tr>
<tr>
<td>time/week High</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>occupational PA</td>
<td>14,347 (58.0)</td>
<td>14,035 (57.4)</td>
<td>12,820 (52.7)</td>
<td>11,296 (46.6)</td>
</tr>
</tbody>
</table>

BMI, body mass index
PA, physical activity
Values are shown as mean (standard deviation) or numbers (percentages)

<sup>a</sup> <0.94 standard drinks/week in men, <0.09 standard drinks/week in women, b0.94–2.85 standard drinks/week in men, 0.09–1.09 standard drinks/week in women, c2.85–4.84 standard drinks/week in men, 1.09–2.13 standard drinks/week in women, d≥4.84 standard drinks/week in men, ≥2.13 standard drinks/week in women

Mean age, body mass index and percentage of participants with hypertension, cancer, diabetes and high occupational physical activity decreased with increasing alcohol intake. The percentage of participants that were ever smokers, had more than secondary school education, and performed leisure
time physical activity at least once a week increased with increasing alcohol intake.

The association between alcohol consumption, alcohol dependence and risk of first-time VTE in a multivariable model adjusted for age, body mass index, hypertension, smoking, education level and cancer is shown in Table 11. There was an association between alcohol consumption and risk of VTE in men ($P$ for trend $0.02$ over quartiles of alcohol consumption). Men in the highest quartile of alcohol consumption had a higher risk of VTE (HR $1.22$; 95% CI $1.01$–$1.47$) compared to men in the lowest quartile of alcohol consumption. There was also an association between alcohol dependence and risk of VTE in men (HR $1.30$; 95% CI $1.07$–$1.59$). There were no significant associations between alcohol consumption, alcohol dependence and risk of VTE in women. As part of a sensitivity analysis, we also performed a matched case-control analysis of the association between alcohol consumption and risk of VTE. The aim of this analysis was to account for any changes in alcohol consumption on the population level during the study period by matching cases and controls for health examination date. The results of this analysis are shown in Table 11.
Table 11. Alcohol consumption, alcohol dependence and risk of first-time venous thromboembolism in men and women

<table>
<thead>
<tr>
<th></th>
<th>Events</th>
<th>Cox regression model$^a$</th>
<th>Matched case-control model$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol dependence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>846</td>
<td>1 (ref.)</td>
<td>1 (ref.)</td>
</tr>
<tr>
<td>Yes</td>
<td>113</td>
<td>1.30 (1.07–1.59)</td>
<td>1.22 (0.96–1.54)</td>
</tr>
<tr>
<td>Standard drinks per week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;0.94</td>
<td>223</td>
<td>1 (ref.)</td>
<td>1 (ref.)</td>
</tr>
<tr>
<td>0.94–2.85</td>
<td>201</td>
<td>1.05 (0.87–1.27)</td>
<td>1.01 (0.81–1.26)</td>
</tr>
<tr>
<td>2.85–4.84</td>
<td>220</td>
<td>1.18 (0.98–1.42)</td>
<td>1.08 (0.87–1.34)</td>
</tr>
<tr>
<td>≥4.84</td>
<td>230</td>
<td>1.22 (1.01–1.47)</td>
<td>1.12 (0.90–1.39)</td>
</tr>
<tr>
<td>$P$ for trend</td>
<td></td>
<td>0.02</td>
<td>0.26</td>
</tr>
<tr>
<td>Continuous</td>
<td>884</td>
<td>1.02 (0.99–1.04)</td>
<td>1.01 (0.99–1.03)</td>
</tr>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol dependence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>782</td>
<td>1 (ref.)</td>
<td>1 (ref.)</td>
</tr>
<tr>
<td>Yes</td>
<td>21</td>
<td>1.19 (0.77–1.85)</td>
<td>1.26 (0.76–2.09)</td>
</tr>
<tr>
<td>Standard drinks per week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;0.09</td>
<td>250</td>
<td>1 (ref.)</td>
<td>1 (ref.)</td>
</tr>
<tr>
<td>0.09–1.09</td>
<td>181</td>
<td>0.83 (0.68–1.00)</td>
<td>0.81 (0.64–1.01)</td>
</tr>
<tr>
<td>1.09–2.13</td>
<td>164</td>
<td>0.92 (0.75–1.13)</td>
<td>0.90 (0.71–1.15)</td>
</tr>
<tr>
<td>≥2.13</td>
<td>157</td>
<td>0.92 (0.75–1.14)</td>
<td>0.94 (0.73–1.20)</td>
</tr>
<tr>
<td>$P$ for trend</td>
<td></td>
<td>0.58</td>
<td>0.73</td>
</tr>
<tr>
<td>Continuous</td>
<td>752</td>
<td>1.01 (0.97–1.06)</td>
<td>1.03 (0.98–1.09)</td>
</tr>
</tbody>
</table>

$^a$Model adjusted for age, body mass index, hypertension, smoking, education level and cancer. Associations are shown as hazard ratios with 95% confidence intervals.

$^b$Cases and controls matched for age, sex, health examination date and primary health care center. Four controls from the Venous thromboEmbolism In Northern Sweden study population were matched to each case (nested case-control study). Model adjusted for body mass index, hypertension, smoking, education level and cancer. Associations are shown as odds ratios with 95% confidence intervals.
Leisure time physical activity, occupational physical activity and risk of first-time VTE

The association between leisure time physical activity, occupational physical activity and risk of first-time VTE is depicted in Figure 14. Women who performed leisure time physical activity at last once a week had a lower risk of VTE compared to women who performed leisure time physical activity less often (HR 0.83; 95% CI 0.71–0.98). Women with high occupational physical activity had a lower risk of VTE (HR 0.85; 95% CI 0.74–0.98) compared to women with low occupational physical activity. There was no significant association between performing leisure time physical activity at least once a week and risk of VTE in men. No association was seen between occupational physical activity and risk of VTE in men.

We also categorized frequency of leisure time physical activity into four groups. The associations between leisure time physical activity and risk of VTE are shown in Table 12, as is the association between physical activity according to the Cambridge index and risk of VTE.

Figure 14 The association between leisure time physical activity (PA), occupational PA and risk of first-time venous thromboembolism in men and women. Associations are shown as hazard ratios with 95% confidence intervals and are adjusted for age, body mass index, hypertension, smoking, education level and cancer.
Table 12. The association between leisure time physical activity, physical activity according to the Cambridge index and risk of first-time venous thromboembolism

<table>
<thead>
<tr>
<th></th>
<th>Events</th>
<th>Age-adjusted model</th>
<th>Multivariable modela</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leisure time physical activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>455</td>
<td>1 (ref.)</td>
<td>1 (ref.)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>230</td>
<td>0.83 (0.71–0.97)</td>
<td>0.81 (0.69–0.95)</td>
</tr>
<tr>
<td>1–2 times/week</td>
<td>126</td>
<td>0.85 (0.70–1.04)</td>
<td>0.84 (0.69–1.03)</td>
</tr>
<tr>
<td>&gt;1–2 times/week</td>
<td>140</td>
<td>1.00 (0.83–1.21)</td>
<td>1.01 (0.83–1.22)</td>
</tr>
<tr>
<td>P for trend</td>
<td>0.48</td>
<td></td>
<td>0.48</td>
</tr>
<tr>
<td>Cambridge index</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>136</td>
<td>1 (ref.)</td>
<td>1 (ref.)</td>
</tr>
<tr>
<td>Moderately inactive</td>
<td>332</td>
<td>1.03 (0.84–1.26)</td>
<td>1.08 (0.89–1.32)</td>
</tr>
<tr>
<td>Moderately active</td>
<td>273</td>
<td>1.06 (0.86–1.30)</td>
<td>1.15 (0.93–1.41)</td>
</tr>
<tr>
<td>Active</td>
<td>159</td>
<td>0.97 (0.77–1.22)</td>
<td>1.06 (0.84–1.34)</td>
</tr>
<tr>
<td>P for trend</td>
<td>0.88</td>
<td></td>
<td>0.50</td>
</tr>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leisure time physical activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>397</td>
<td>1 (ref.)</td>
<td>1 (ref.)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>207</td>
<td>0.92 (0.78–1.09)</td>
<td>0.95 (0.80–1.12)</td>
</tr>
<tr>
<td>1–2 times/week</td>
<td>124</td>
<td>0.78 (0.63–0.95)</td>
<td>0.82 (0.67–1.01)</td>
</tr>
<tr>
<td>&gt;1–2 times/week</td>
<td>78</td>
<td>0.75 (0.59–0.96)</td>
<td>0.80 (0.63–1.03)</td>
</tr>
<tr>
<td>P for trend</td>
<td>0.003</td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td>Cambridge index</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>114</td>
<td>1 (ref.)</td>
<td>1 (ref.)</td>
</tr>
<tr>
<td>Moderately inactive</td>
<td>296</td>
<td>1.07 (0.86–1.32)</td>
<td>1.08 (0.87–1.34)</td>
</tr>
<tr>
<td>Moderately active</td>
<td>204</td>
<td>0.90 (0.71–1.13)</td>
<td>0.90 (0.72–1.14)</td>
</tr>
<tr>
<td>Active</td>
<td>131</td>
<td>0.87 (0.68–1.12)</td>
<td>0.88 (0.69–1.14)</td>
</tr>
<tr>
<td>P for trend</td>
<td>0.06</td>
<td></td>
<td>0.08</td>
</tr>
</tbody>
</table>

Associations are shown as hazard ratios with 95% confidence intervals.
aModel adjusted for age, body mass index, hypertension, smoking, education level and cancer.

There was an association between more frequent leisure time physical activity and lower risk of VTE in women (P for trend 0.02 in a fully adjusted model), but not in men. Men who sometimes performed leisure time physical activity had a lower risk of VTE compared to men who never performed leisure time physical activity (HR 0.81; 95% CI 0.69–0.95 in a fully adjusted model). There was no significant association between increasing levels of physical activity according to the Cambridge index and risk of VTE in women (P for trend 0.08 in a fully adjusted model). In a sensitivity analysis, participants with missing data for one of the components of the index (leisure time physical activity or occupational
physical activity) were given the lowest intensity score for the missing component. In that analysis, an association between higher levels of physical activity according to the Cambridge index and lower risk of VTE in women was seen ($P$ for trend $0.03$ in a fully adjusted model, data not shown). In men, there was no association between physical activity according to the Cambridge index and risk of VTE in any model.

**Interaction between sex and risk markers for VTE**

A multiplicative interaction analysis for the interaction between sex and the studied risk markers for VTE was performed. There was no significant interaction between sex and fasting plasma glucose, oral glucose tolerance test two-hour post-load plasma glucose, diabetes, alcohol consumption, alcohol dependence, leisure time physical activity or occupational physical activity. The multiplicative interaction term for the interactions between leisure time physical activity (dichotomized), occupational physical activity and sex had the lowest $P$ values in a fully adjusted model ($0.07$ and $0.15$ respectively). An additive interaction analysis was performed for the interaction between leisure time physical activity, occupational physical activity and sex in relation to risk of first-time VTE. The results of this analysis are shown in Tables 13 and 14.

**Table 13.** Additive interaction analysis between sex and leisure time physical activity in relation to risk of first-time venous thromboembolism

<table>
<thead>
<tr>
<th>Leisure time physical activity</th>
<th>Female sex</th>
<th>Male sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥1 time/week</td>
<td>1 (ref.)</td>
<td>1.54 (1.28–1.85)</td>
</tr>
<tr>
<td>&lt;1 time/week</td>
<td>1.21 (1.03–1.42)</td>
<td>1.53 (1.30–1.79)</td>
</tr>
</tbody>
</table>

Associations are shown as hazard ratios with 95% confidence intervals. Model adjusted for age, body mass index, hypertension, smoking, education level and cancer.

**Table 14.** Additive interaction analysis between sex and occupational physical activity in relation to risk of first-time venous thromboembolism

<table>
<thead>
<tr>
<th>Occupational physical activity</th>
<th>Female sex</th>
<th>Male sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>1 (ref.)</td>
<td>1.44 (1.26–1.64)</td>
</tr>
<tr>
<td>Low</td>
<td>1.16 (1.004–1.34)</td>
<td>1.45 (1.26–1.66)</td>
</tr>
</tbody>
</table>

Associations are shown as hazard ratios with 95% confidence intervals. Model adjusted for age, body mass index, hypertension, smoking, education level and cancer.
Discussion

Main findings

Accuracy of VTE diagnoses in administrative registries

In paper II, we found a higher positive predictive value for first-time PE (81%) as compared to first-time DVT (59%). The proportion of misclassified first-time PE events (1%) was lower than that of first-time DVT (16%).

The positive predictive values for PE and DVT in our study was higher than those reported in a study evaluating the accuracy of VTE diagnosis codes in the Danish National Patient Registry (42). In that study, the positive predictive value was 67% for PE and 55% for DVT. In Denmark, visits to emergency departments and outpatient clinics became reportable to the Danish National Patient Registry in 1995. In Sweden, emergency department visits and outpatient visits were not included in the Swedish National Patient Registry until 2001. Consequently, during the earlier period of our study, only diagnoses registered during inpatient stays were included. In the Danish study, diagnoses registered in emergency departments were included during the whole study period. Diagnoses registered at emergency departments have low positive predictive values for VTE; 30 to 56% for PE and 32 to 42% for DVT (42, 307). When restricting our analyses to the period 2009 to 2014, a period during which diagnoses registered during emergency department and outpatient visits were included in the Swedish National Patient Registry, the positive predictive value of PE diagnoses was higher than in the Danish study, but that of DVT was comparable to the positive predictive value reported in the Danish study (42).

There are several strategies for increasing the accuracy of outcome assessment in registry-based studies of VTE. The positive predictive value of ICD diagnosis codes for VTE depends on the population studied. A systematic review of studies conducted in the United States found a higher positive predictive value for ICD diagnosis codes for VTE in studies investigating persons at high risk of VTE, such as postsurgical patients (308). Thus, if possible, it is preferable to restrict the study population to high-risk individuals. Moreover, as shown in our study as well as in a Danish study, the positive predictive value for diagnoses of PE is higher than that of DVT (42). Registry-based studies should ideally be focused on PE only. VTE diagnoses registered in the primary diagnosis field have been shown to have a higher positive predictive value for VTE (42, 309, 310). In our study, we had no information on whether the registered diagnoses were primary or secondary. Another strategy is to limit the diagnosis codes used to define VTE to those with the highest positive predictive value. An American study of
pregnancy-related VTE in the 1990s showed a wide range of positive predictive values for VTE depending on the diagnosis codes used to define a VTE event (310). As described above, restricting the search to inpatient diagnoses alone can increase the positive predictive value of VTE diagnoses (42, 43). Alternatively, an additional requirement can be added to the definition of VTE events. A Swedish study of peripartum VTE suggested that a definition of VTE requiring both a VTE diagnosis code in the National Patient Registry and an anticoagulant prescription could lead to a lower number of false positive VTE cases (311). Similarly, a Canadian study showed that a combination of ICD diagnosis codes and imaging procedure codes had a high positive predictive value for VTE (309).

**Incidence of first-time VTE**
The incidence of first-time VTE in our study was 137 (95% CI 122–154) per 100,000 adults per year, similar to that of two Norwegian studies (27, 28), a French study (29) and an American study by Huang et al. (31), but higher than that reported in an Australian study (32) and an American study by Silverstein et al. (30). In our study, the incidence of VTE increased markedly with age, consistent with published data (27-32). Additionally, in our study, the incidence of first-time VTE was numerically higher in women than in men but the CIs overlapped.

Differences in incidence between studies can have several explanations. First, the studies can have been conducted during different time periods. There is some evidence that the incidence of VTE is increasing over time. One study showed that the incidence of VTE events with a higher diagnostic certainty increased slightly between 1966 and 1990 (30). In another study, conducted between 1985 and 2009, the incidence of first-time VTE nearly doubled and the incidence of first-time PE almost tripled (31). Our study was conducted in 2006, and is one of the most recent studies on the incidence of VTE. The incidence of first-time VTE in our study was similar to that reported for the year 2009 in the study by Huang et al. (31).

Secondly, some studies included children and adolescents in the study population (29-32), and some did not (27, 28). As the incidence of VTE in children and adolescents is very low (29), the inclusion of these groups can lower the overall incidence of VTE. In our study, only individuals aged 18 and above were included.

Thirdly, the studied populations may be composed of persons with different ethnicities. Ethnicity is suggested to have a modest impact on risk of VTE. For example, persons of Asian or Pacific Islander ancestry seem to have a lower risk
of VTE compared to persons of other ethnicities (100, 101). The vast majority of the participants of our study were of European ancestry. The lower incidence of VTE found in the Australian study may partly be explained by a higher proportion of persons with Asian ancestry residing in the study area (32).

Fourthly, different case-identification strategies, such as physician-reported VTE cases (29), search of ICD diagnosis codes for VTE or VTE-related diagnostic procedures in registries (27), or a combination of prospective patient inclusion by treating physicians, retrospective registry searches and outreach methods such as advertising (32) have been employed in different studies.

Lastly, there are differences in the rigorousness of the definition of a verified VTE event between the studies. Some studies required the VTE event to be symptomatic and treated with anticoagulants, thrombolytics or vascular surgery (28), whereas others, like ours, did not (27, 29, 30). Some (28, 30), but not all (27, 29), studies included VTE events diagnosed at autopsy. Our study did include VTE events identified at autopsy, but as the autopsy rate in Västerbotten County was low, only 17% in 2006, it is probable that some fatal VTE cases were not identified. A few studies (27, 30), made a distinction between possible, probable or definite VTE events. The incidence of VTE in these studies increased markedly when possible VTE events were included in the incidence calculations.

**Risk markers for first-time VTE at the time of VTE diagnosis**

In our study, the most common risk markers for VTE at the time of diagnosis of a first-time VTE event were recent hospitalization and concurrent malignancy. One reason for the differences between studies in the proportion of patients with various risk markers for VTE is selection bias. We identified cases by a diagnosis registry search, as well as a search for fatal VTE events in the Cause of Death Registry. Information about presence of risk markers for VTE was obtained from medical records. This approach led to inclusion of a wide range of VTE patients, including patients with fatal VTE events, very elderly patients, and patients with severe comorbidities such as metastatic cancer. Such patients may be underrepresented in studies where patients are included at, for example, anticoagulation clinics.

The percentage of patients with VTE with concurrent malignancy was 31% in our study and 30% in an American study where ICD diagnosis codes were used to identify potential cases (47). In contrast, the percentage of patients with concurrent malignancy was 12% in both the Malmö Thrombophilia Study, a study conducted at a Swedish teaching hospital where a research nurse
identified patients with VTE by screening hospital records for VTE cases (312), and a Dutch study where cases were identified at anticoagulation clinics (313).

Another reason for variations in the reported percentage of patients with risk markers for VTE is different definitions of the risk markers. For example, our study and the Malmö Thrombophilia Study defined all patients with a diagnosis of cancer registered before the VTE event as having concurrent malignancy (312). In contrast, a Norwegian study only considered active cancer as a risk factor for VTE (27). The time frame used in the definition of provoking factors for VTE also differs between studies. In our study, we considered events such as hospitalization and surgery to be risk markers for a VTE event if they occurred within 60 days before the onset of symptoms of VTE. Other studies have used a time period of three months, or varied the time limit according to the risk marker in question (27, 28).

**Glucose levels and risk of first-time VTE**

We found an association between fasting plasma glucose and risk of VTE in an unadjusted model. The association was attenuated after adjustment for age and sex. After further adjustment for potential confounders, including body mass index, the association was no longer significant. This underscores the importance of adjusting for potential confounders when studying the association between fasting plasma glucose and risk of VTE. Our results are in agreement with the results of most previous studies of the association between fasting plasma glucose and the risk of VTE (142, 144, 169-171). In contrast, one study showed an association between higher levels of fasting plasma glucose and increased risk of fatal VTE events (141). It seems unlikely that there should be a causal association between fasting plasma glucose and fatal VTE events, and no association between fasting plasma glucose levels and the risk of fatal and non-fatal VTE events as a combined endpoint. To our knowledge, no other study has investigated the association between oral glucose tolerance test two-hour post-load plasma glucose and risk of VTE.

Glycated hemoglobin A, which measures glycemia over a time period of 8 to 12 weeks, has a lower within-person variability compared to fasting plasma glucose (314). Consequently, it may be a better marker of glycemic status in studies such as ours, where glycemic status is evaluated at one time point only. It would have been valuable to have information on levels of glycated hemoglobin A, in addition to fasting plasma glucose and oral glucose tolerance test two-hour post-load plasma glucose. However, levels of glycated hemoglobin A have not been shown to be associated with risk of VTE in other studies (172, 173).
Diabetes and risk of first-time VTE
In univariable analysis, we found an association between diabetes and risk of VTE. This association was attenuated after adjustment for potential confounders and was no longer significant. Similarly, in an individual participant data meta-analysis of prospective studies, there was an association between diabetes and risk of VTE, which was no longer significant after adjustment for age, sex and body mass index (144). In another individual participant data meta-analysis, there was an association between diabetes and decreased risk of VTE in the United Kingdom Biobank cohort that included mostly non-fatal VTE events, and between diabetes and increased risk of VTE in the Emerging Risk Factors Collaboration cohort, where only fatal VTE events were registered as outcomes (141). The Emerging Risk Factors Collaboration cohort consisted of 75 different subcohorts with different inclusion criteria. The associations between diabetes and risk of VTE in the different subcohorts are not presented separately, but it is possible that the strength and direction of the association differ substantially across subcohorts.

Alcohol consumption and risk of first-time VTE
One previous study showed an association between higher alcohol consumption and increased risk of VTE. Separate analyses for men and women were not presented (190). Other prospective studies have not shown an association between alcohol intake and risk of VTE (191, 315), except for one study that showed an association between higher alcohol intake and lower risk of VTE in elderly persons (192). An individual participant data meta-analysis found that persons currently consuming alcohol had a lower risk of VTE compared to persons with "other" alcohol consumption (141). If a high proportion of the persons with "other" alcohol consumption were former drinkers, as opposed to lifetime abstainers, the association between current alcohol consumption could be explained by reverse causality. The former drinkers may be sick quitters, i.e. persons with preexisting medical conditions who have stopped drinking (316). As persons with a range of medical conditions are at increased risk of VTE (87, 88, 317), this could explain the association between being a current drinker and having a lower risk of VTE. However, this theory is contradicted by the fact that the inverse association between alcohol consumption and risk of VTE was also seen when only current drinkers were included in the analysis (141).

Our finding of an association between alcohol consumption and risk of VTE in men differs from the findings of previous studies. To date, the largest study of the association between alcohol consumption and risk of VTE in men was a retrospective case-control study (189). That study showed no significant association between moderate alcohol consumption and risk of VTE in men. Similarly, two Danish population-based studies and two American studies of
male health care professionals found no association between alcohol consumption and risk of VTE in men (193, 236, 315, 318).

We found no association between alcohol consumption and risk of VTE in women. This is in agreement with the results of two Danish population-based studies and a study of female nurses (236, 315, 318). On the other hand, two other studies have shown an association between moderate alcohol consumption, compared to no alcohol consumption, and lower risk of VTE in women (189, 194). One of the studies is a retrospective case-control study, a study design that can be subject to recall bias (189). An American study showed an inverse association between alcohol consumption and risk of VTE in women (195). In that study, the participants had a mean age of 66 years at the start of the follow-up period, compared to 46 years in our study. Among elderly women, those who drink alcohol have better self-perceived health status and lower rates of hospitalizations (319), that is, elderly women that don’t drink alcohol tend to be sick quitters. This phenomenon could explain the association between higher alcohol consumption and lower risk of VTE in the American study.

We can only speculate about the reason for our finding of an association between alcohol consumption and risk of VTE in men, but not in women. One possible explanation is lack of power. As fewer women consume large amounts of alcohol, it is possible that a true association between alcohol consumption and higher risk of VTE in women could not be detected. However, when analyzing the association between alcohol consumption in quartiles and risk of VTE in women, there was a non-significant association between higher alcohol consumption and lower risk of VTE. Confounding could also contribute towards the differences between men and women in the association between alcohol consumption and risk of VTE. Many measured participant characteristics, such as age, smoking and level of physical activity, differ between participants with different levels of alcohol consumption, demonstrated in Table 10. It is likely that the distribution of unmeasured variables also differs between participants with different levels of alcohol intake, and if these unmeasured variables affect the risk of VTE, they act as confounders of the association between alcohol intake and risk of VTE. The impact of such confounders could differ between men and women. Differences in exposure could also contribute towards the observed differences between men and women in the association between alcohol consumption and risk of VTE. As shown in Figure 13, more than half of the total amount of alcohol consumed by men was consumed as beer, whereas for women, wine was the predominant alcoholic beverage. It is possible that health effects differ between different alcoholic beverages. Lastly, there may be an inherent difference between men and women in how the body is affected by alcohol.
**Alcohol dependence and risk of first-time VTE**

We found a significant association between alcohol dependence, defined as two or more affirmative answers to the CAGE questionnaire, and risk of VTE in men, but not in women. Few women with alcohol dependence participated in the VEINS study. Thus, the lack of association between alcohol dependence and risk of VTE in women could be explained by a lack of statistical power. In men, our results were similar to those of a Swedish study of the association between alcohol use disorders, with and without somatic complications, and risk of VTE in men and women combined (200). A diagnosis of an alcohol use disorder probably reflects a more extensive misuse when compared to our definition of alcohol dependence.

**Physical activity and risk of first-time VTE**

We found an association between physical activity and a lower risk of VTE in women. We found no consistent association between different measures of physical activity and risk of VTE in men.

Our study showed an association between higher leisure time physical activity, occupational physical activity and decreased risk of VTE in women. Five previous studies regarding the association between physical activity and risk of VTE in women have shown an association between higher physical activity and lower risk of VTE (194, 235, 320-322). Three studies showed no association between leisure time physical activity and risk of VTE in women (236, 323, 324). One study showed that women who performed strenuous activities up to two to three times per week had a decreased risk of VTE compared to inactive women, but women who performed strenuous physical activity more often did not have a decreased risk of VTE (325). Only one previous prospective study has presented results for the association between occupational physical activity and risk of VTE in women separately (236). In that study, there was no significant association between occupational physical activity and risk of VTE in women.

We found no consistent association between leisure time physical activity, occupational physical activity and risk of VTE in men. This is in agreement with the results of two Scandinavian population-based studies (236, 323). Other studies have shown an association between physical activity and either a lower (235, 321), or a higher (193) risk of VTE in men.

The reason for the differences between men and women in the association between physical activity and risk of VTE may be explained by differences in exposure to physical activity. On group level, both the type and the intensity of a leisure time physical activity differ between men and women. For example, men spend more time than women performing moderate and vigorous physical
activity, whereas women spend more time in light physical activity (208, 210). It is possible that the effects of physical activity on the coagulation system and the fibrinolytic system differ between different types of physical activity. Men and women who report the same level of occupational physical activity can have different occupations, encompassing different types of physical activity, such as walking as opposed to lifting heavy objects. Moreover, men and women with the same occupation report different exposure patterns, for instance regarding the performance of repetitive tasks and work in tiring positions (326). The biological response to physical activity may also differ between men and women.

**Methodological considerations**

Study I is a retrospective cohort study. The VEINS study, results of which are reported in papers II through V, is a prospective cohort study. One advantage of cohort studies is that study participants with and without the study outcome, VTE, are drawn from the same population. In the papers III through V, which investigated risk markers for VTE, the presence of risk markers was assessed before the event, thus eliminating the possibility of recall bias. Recall bias is a common issue in retrospective case-control studies of the association between a risk marker and an outcome. Cases tend to search their memories to identify possible exposure to various risk markers more thoroughly than controls, as the cases are motivated by a wish to find an explanation for their condition (293).

**Internal validity**

Internal validity in a cohort study is defined as the extent to which the results of a study are valid for the study cohort (327). A high accuracy in the measurements of exposure variables and outcome assessment, a low level of selection bias and an appropriate treatment of potential confounding factors are all important for a high internal validity (276, 297, 328).

**Exposure variables from medical records**

In paper I, information about risk markers for VTE at the time of VTE diagnosis was obtained from medical records. In some cases, information was sparse, and it is possible that, in some participants, risk markers for VTE were present but not recorded in the medical records. Thus, we may have underestimated the proportion of participants with VTE and risk markers for VTE at the time of VTE diagnosis. However, our finding that 33% of participants with VTE had no risk markers for VTE at VTE diagnosis is well in line with the findings of other studies (27, 46, 47).
Exposure variables from health examinations

VIP health examinations were conducted at the primary health care centers where the participants were registered. The familiar environment and staff were thought to make the participants feel less anxious in the health examination situation, thus increasing the probability of correct resting blood pressure values, questionnaire answers et cetera. On the other hand, the fact that the health examination results and questionnaire answers were evaluated by a primary health care nurse, and that the health examination was followed by counseling about healthy lifestyle habits, might make some participants reluctant to admit to perceived unhealthy lifestyle habits when answering the questionnaire.

Questionnaires, such as those used in the VEINS study, tend to underestimate true alcohol intake quantified by indirect measures, such as sales statistics, and to overestimate true physical activity, as measured by accelerometer (184, 204). This introduces an error in the measurement of these exposures, leading to misclassification of participants. We assume that, on the whole, the degree of misclassification of alcohol intake or physical activity habits in the VEINS study was not dependent on outcome status. That is, the misclassification was non-differential. Non-differential misclassification normally leads to bias towards the null, which means that the magnitude of any association is reduced (205). However, the accuracy of questionnaire data regarding alcohol consumption and physical activity depends on the age and sex of the person answering the questionnaire (185, 204). As age and sex are associated with risk of VTE, there is a risk of differential misclassification, which can lead to an under- or overestimation of the association between the studied risk marker and the outcome (299). We have tried to minimize this issue by stratifying our analyses for sex, and by performing matched case-control sensitivity analyses where cases and controls were matched for a range of variables including sex and age at health examination.

The use of a food frequency questionnaire to assess alcohol consumption meant that we were unable to study the association between specific drinking patterns, such as binge drinking, and risk of VTE. Binge drinking has not been significantly associated with VTE in previous studies (191, 315). As the calibration coefficient between the food frequency questionnaire and 24-hour diet recalls was close to 1.0 (305), we consider the answers to the food frequency questionnaire used in paper IV to be a relatively accurate representation of the participants' actual alcohol consumption.

The CAGE questionnaire was used to detect alcohol dependence. This questionnaire was designed to be used as a screening instrument, rather than a diagnostic instrument (198, 329). A clinical diagnosis of alcohol dependence
cannot be based on the CAGE questionnaire only. In the latest edition of the Diagnostic and Statistical Manual of Mental Disorders, alcohol dependence has been replaced by the term "alcohol use disorder" (330), a diagnosis which encompasses features from both alcohol dependence and alcohol abuse. The ICD system still treats alcohol dependence as a separate entity. It would have been preferable if we had access to questionnaire or interview data with information on all items included in the diagnostic criteria for alcohol use disorder or alcohol dependence. Unfortunately, this was not the case. Nevertheless, the CAGE questionnaire performs well at detecting alcohol problems. A study comparing the CAGE questionnaire and the Alcohol Use Disorders Identification Test (331) found that the CAGE questionnaire had an area under the receiving operating characteristic curve of 0.82 for detecting active alcohol abuse or dependence, compared to 0.78 for the Alcohol Use Disorders Identification Test (332). In contrast, the Alcohol Use Disorders Identification Test was superior to the CAGE questionnaire at identifying heavy drinkers.

One criticism against the original version of the CAGE questionnaire is that it aims to detect lifetime alcohol dependence. For instance, participants are asked if they ever have felt they should cut down on their drinking. This means that the questionnaire cannot distinguish between active and past alcohol dependence (332). In our study, the timeframe of the CAGE questionnaire varied between different versions of the questionnaires used during the 30 years of VIP health examinations. In different versions of the questionnaire, the participants were asked to answer the questions according to their behavior during the last year, the last five years, or during their lifetime. It is unlikely that these changes in timeframe have lead to differential misclassification. It may indicate that a higher proportion of the participants in the VEINS study with two or more affirmative answers to the CAGE questionnaire had an active or recent alcohol dependence when compared to the proportion in other studies.

The physical activity questionnaire lacked information on the duration and intensity of the leisure time physical activity performed. Also, some types of leisure time physical activity, for example walking or bicycling in street clothes, was not captured by the questionnaire. This means that energy expenditure could not be estimated. However, it has been shown that a questionnaire on physical activity patterns gave a more accurate representation of total physical activity, as measured by accelerometer, compared to a more complex questionnaire that assessed energy expenditure (333). Regarding occupational physical activity, we had no information on the type of occupational physical activity performed (for instance walking, lifting, carrying or kneeling). It would also have been useful to have information on the participants’ occupations to be
able to assess the association between different types of occupations and risk of VTE.

In the VEINS study, exposure variables were measured at one time-point only. True within-person biological variability in, for example, glucose levels and errors in the glucose level measurements can cause non-differential misclassification. This is a cause of regression dilution bias, a statistical phenomenon where the association between an exposure variable and an outcome is attenuated due to random measurement errors. Measuring exposure variables as precisely as possible and/or obtaining multiple measurements per study participant leads to better estimates of the true exposure levels (334). The number of repeated measurements needed depends on the degree of within-person variability in the studied exposure. A study using accelerometers to estimate habitual physical activity found that a mean measurement period of 182 days was needed to correctly estimate the amount of vigorous physical activity, but only 2.4 days was needed to correctly estimate the amount of sedentary behavior (335). To lower the risk of regression dilution bias in the VEINS study, it would have been preferable to measure exposure variables at two or more time-points.

Longitudinal changes in behavior may also dilute the association between an exposure and an outcome in studies such as ours where the exposure is measured at a single time-point, in many cases years before the outcome occurs. It is reasonable to believe that the exposure during the follow-up period also affects the individual’s risk of experiencing an outcome, assuming a causal association between the exposure and the outcome. In the VEINS study, participants were followed for a mean period of 13.9 years and received individualized counseling about healthy lifestyle habits as an integral part of the VIP. Such counseling can trigger a change in the participants’ behavior and thus make the laboratory measurements taken and questionnaire answers given at the time of the health examination visit less representative of the participants’ biomarker levels and behavior during the follow-up period. In the VIP study population, there were significant temporal shifts in the distribution of cardiovascular risk factors between 1991 and 2010 (336). It would have been preferable to obtain repeated measurements of the exposure variables to account for these longitudinal changes in biomarker levels and behavior.

**Risk of selection bias in the inclusion of study participants**

All VEINS study participants had participated in the VIP health examinations. The risk of selection bias in the VEINS study consequently depends on that of the VIP. The VIP participation rates in Västerbotten County increased over time, independently of age, sex, socioeconomic status and history of medical
conditions; the overall VIP participation rate was 66% during the years 1990 to 2006 (337). The participation rate among persons with low income or single marital status was approximately ten percentage points lower than among persons with middle or high income or persons who were married. There were only minimal differences in education level between participants and non-participants. The involvement of primary health care centers is thought to contribute to the high participation rates.

Among the VIP participants, the vast majority consented to participate in research at their health examination and were considered for participation in the VEINS study. In papers II through V, only persons with no previous VTE events were included. To avoid inclusion of persons with previous VTE events, VIP participants with one or more of the ICD diagnosis codes used to identify potential VTE events registered before the VIP health examination were excluded from the study (n=2,212). As shown in Papers I and II, the positive predictive value of a VTE diagnosis code in Swedish administrative registries is low. Thus, it is reasonable to believe that some individuals without previous VTE events, but with other medical conditions that had necessitated a health care visit, were excluded from the VEINS study. This means that the VEINS study population may be healthier than the inhabitants of Västerbotten County in general.

Among the persons eligible for inclusion in the VEINS study, 17 individuals declined to participate, and medical records could not be obtained for 33 individuals. The exclusion of these individuals (n=50) likely confers a minimal risk of selection bias.

**Risk of selection bias conferred by the treatment of missing data**
In the VEINS study, we used a complete case methodology where only participants with complete data on the studied exposure and all adjustment variables were included in the regression analyses. This approach is most suitable if data is missing completely at random. It is a simple approach that ensures comparability across different analyses. However, if data is not missing completely at random, estimates of associations can be biased. It has been shown that study participants with missing data sometimes have a higher risk of adverse outcomes (338). Furthermore, omitting participants from analyses leads to a reduction in sample size. In turn, this leads to a reduced statistical power and a reduced lack of precision in the study results. Use of a complete case methodology can be justified when analyzing a large dataset with few missing observations (339). In our study, only 5% of study participants had missing data for any adjustment variable. It is unlikely that omitting participants with missing data for adjustment variables can have introduced...
more than minimal bias in our estimates. Among the studied exposures, the percentage of missing data was highest for Cambridge index, where 11% of participants had missing data. For this variable, a sensitivity analysis, where participants who had answered only one of the physical activity questions included in the Cambridge index were given the lowest intensity score for the missing variable, was made. The association estimates from the sensitivity analysis were similar to those of the complete case analysis, except that the estimates yielded by the sensitivity analysis had a slightly higher precision.

Multiple imputation is the choice method to deal with missing data in many branches of research (339). The aim of multiple imputation is to generate unbiased, valid estimates of associations based on the observed data, not to estimate the missing values themselves (340). Data is assumed to be missing at random, but the method can also be used to handle data missing not at random. Multiple imputation involves replacing the missing values with values generated based on the participant’s values for other variables in the dataset (339). In the VEINS study, we did not use multiple imputation to handle missing values as we considered the percentage of participants with missing data to be sufficiently low to justify a complete case methodology. We deemed that the exclusion of participants with missing data from regression analyses would not greatly affect our ability to detect clinically relevant associations.

**Confounders**

As it is known that there are associations between age, body mass index, blood pressure, education level, smoking, cancer and risk of VTE (31, 51, 144, 236), as well as associations between these factors and the exposure variables studied in papers III through V (plasma glucose, diabetes, alcohol intake, and physical activity) (178, 341-346), we considered these factors to be possible confounders. To adjust for the impact of these known confounders, multivariable analysis was used in papers III through V. Furthermore, as there is an association between sex and the studied exposure variables as well as with VTE, analyses were either stratified by sex or adjusted for sex (34, 177, 209, 347).

There is a risk of residual confounding when known confounding factors cannot be measured with enough precision (348). A potential source of residual confounding in our studies is the limited data on cancer, which is an important risk factor for VTE (51). There was no information on cancer types or stages, or on whether or not participants were subject to cancer treatment at the time of the health examination. The cancer type, stage and ongoing cancer treatment can affect the association between cancer and risk of VTE (349, 350). However, as only 2.6% of our study population was diagnosed with cancer at the time of
the health examination, the lack of more detailed information on cancer should not affect the main conclusions of this thesis.

There is also a risk of confounding from unmeasured or unknown confounders. For example, it would have been of importance to identify the presence of preexisting medical conditions at baseline. Preexisting medical conditions may in themselves be associated with an increase in risk of VTE (87, 88, 317) and/or expose the individual to known provoking factors for incident VTE, such as hospitalizations (351). At the same time, medical conditions are associated with other factors, such as a lower level of physical activity (352). This can lead to a risk of reverse causation, i.e. a false conclusion that a low level of physical activity is a risk factor for VTE.

As the mean age at inclusion in the VEINS study was 46 years, it is unlikely that a large proportion of the study participants had medical conditions of sufficient severity to affect behaviors like alcohol consumption and physical activity. This assumption is supported by the fact that, among the VEINS study participants, only 6% reported that their health had been "pretty bad" or "bad" during the year preceding the health examination. As all studied risk markers were measured before the outcome, there is no risk that the VTE event itself could cause a difference in behavior between participants with and without VTE. Thereby, an important cause of reverse causation is eliminated.

Only a randomized controlled trial of sufficient size can cancel out the effects of unknown confounders.

**Outcome assessment**

Potential VTE events were identified by a diagnosis registry search for a wide range of ICD diagnosis codes. The Cause of Death Registry was searched for a subset of the ICD codes. In paper I, we found a false negative rate of 9% for VTE. In order to lower the rate of false negative VTE cases, we revised the list of ICD diagnosis codes before commencing the case-finding process for papers II through V.

A diagnosis registry search is a commonly used method for identifying VTE events in prospective studies of first-time VTE events (236, 353). Some studies have used a combination of ICD diagnosis codes and information from registries of radiology procedures to decrease the proportion of false negative VTE cases (27, 28). Other studies have assigned the task of identifying patients with VTE to study coordinators active at the department managing patients with VTE at each participating hospital (322), or identified patients diagnosed with VTE by screening hospital records for VTE cases (312).
Case finding was based on diagnosis codes registered in Västerbotten County. Our case finding strategy would not capture VTE events diagnosed and treated exclusively in other parts of Sweden or abroad. Additionally, our case finding strategy would not identify VTE events where the person affected did not seek medical care, or VTE events where none of the diagnosis codes used for case finding was registered during the health care visit. Thus, there is a risk of false negative VTE cases. However, we think that our search strategy identified the vast majority of VTE events diagnosed in VEINS participants. Regarding VTE events in VEINS participants initially diagnosed outside of Västerbotten County, VTE-related diagnosis codes registered at follow-up visits in Västerbotten County would have been detected during case finding.

All potential VTE events were validated by review of medical records and/or radiology reports. If it was still unclear whether a VTE event was valid or not, the issue was resolved by internal discussion. A Canadian study showed very good inter-observer agreement for validation of VTE diagnosis codes (307). The researchers performing the validation had no knowledge of the participants’ health examination results. Thus, it is unlikely that exposure status influenced the likelihood of a potential VTE event being considered valid.

The high proportion of false positive VTE events among unvalidated VTE diagnosis codes reported in both papers I and II, and in several other studies (42, 310), suggest that manual validation is necessary to safeguard internal validity. If manual validation is not feasible, an additional requirement on the VTE definition, such as an anticoagulant prescription, might decrease the number of false positive VTE cases (311). If random misclassification of outcomes results in a high proportion of false positive outcome events, the association between an exposure and an outcome can be attenuated. However, misclassification due to incorrect diagnosis codes is unlikely to be completely random. If there is a systematic misclassification of other diseases, for instance peripheral arterial thromboembolic events, as VTE events, there is a risk of under- or overestimating the association between the studied exposure and VTE. The direction and size of bias depends on the association between the exposure and the other conditions misclassified as VTE events.

For a VTE event to be considered valid, we only required that the VTE event was objectively verified using an adequate diagnostic method. Other studies have required signs and symptoms consistent with VTE, and in some cases that VTE-specific treatment was given in patients not diagnosed at autopsy (28, 353). This means that a wider range of VTE events, including incidentally detected events, which must not be confused with asymptomatic VTE events, could be included in our study. An incidental finding of a PE on a chest CT scan is not exceedingly rare; an incidental PE is detected on 3.6% of chest CT scans in cancer patients.
and 1.1% of coronary CT scans (354). During the study period, there were no ongoing screening programs for DVT in Västerbotten County, and no cases of lower or upper extremity DVT were identified at autopsy only. Thus, we believe the number of incidentally detected DVT events in the VEINS study to be very low. It is possible that our inclusion of incidental VTE events means that the proportion of patients with a low thrombotic burden is higher in our study than in studies that did not include incidentally detected VTE events. If the association between our studied risk markers and VTE is stronger for VTE events with a high thrombotic burden, this could have diluted our results. However, as current recommendations advocate similar treatment for incidentally detected PE and non-incidently detected PE events, at least for patients with cancer, we think that the inclusion of incidentally detected VTE events in the VEINS study can be justified (23, 354).

**External validity**

External validity is defined as the degree to which the results of a study can be generalized to the target population – the extended population to which the results of the study are to be applied. Systematic differences between the source population and the target population can introduce bias and pose a threat to the external validity of the study (297).

The studies described in this thesis were conducted in Västerbotten County, northern Sweden. The vast majority of the inhabitants in the study area are of European ancestry. The generalizability of our results to other ethnic groups is thus uncertain. The diabetes prevalence, alcohol intake and physical activity habits in Västerbotten County is comparable to the rest of Sweden (176, 206, 300, 355), and the incidence of first-time VTE is comparable to that presented in other studies (29, 31).

As persons below age 30 were not included in the VIP, our results may not be generalizable to VTE in young adults. Similarly, as persons were only included up to the age of 60 years and the mean study follow-up time was 13.9 years, our study includes few VTE events in the very elderly.

We think that the results of our studies can be generalized to persons of comparable age groups in other parts of Sweden and to other countries with similar populations.

**Causality**

In papers III through V, the associations between glucose levels, diabetes, alcohol consumption, alcohol dependence, physical activity and risk of VTE are
explored. Due to the observational study design, only circumstantial evidence of causality can be gained from our studies.

In the VEINS study, all exposure variables were measured at the time of the VIP health examinations. Thus, it is certain that the exposure preceded the outcome, which is a requirement for causality.

The consistency of our findings regarding the association between alcohol consumption and alcohol dependence and risk of VTE in men supports the theory of a causal association between alcohol intake and risk of VTE in men. In women, we found a consistent association between higher levels of leisure time physical activity, occupational physical activity and lower risk of VTE.

We found evidence of a dose-response relationship in the association between alcohol consumption and risk of VTE in men. This strengthens the theory of a causal relationship. The risk of VTE also increased by the number of affirmative answers to the CAGE questions. In women, there was a dose-response relationship between increasing levels of leisure time physical activity and lower risk of VTE.

There is no evidence regarding the underlying biological mechanisms for the associations between alcohol consumption, alcohol dependence, physical activity and risk of VTE. Such evidence would strengthen the theory of causal associations between these risk markers and risk of VTE. Our studies were not adapted to yield information about the biological mechanisms responsible for the detected associations. This could be the subject of future studies.

In summary, there is some circumstantial evidence that the associations we found between alcohol consumption, alcohol dependence, physical activity and risk of VTE are indeed causal associations. However, we cannot rule out that these associations are, at least in part, due to confounding.

**Gender perspective**

In paper I, all adult residents of Västerbotten County were included. Among the participants diagnosed with first-time VTE, 54% were women. As the incidence of VTE increases with age, the slightly higher proportion of women among participants diagnosed with first-time VTE could partly be explained by the higher proportion of women among the elderly. In 2006, 62% of persons in Västerbotten County aged 80 years and above were women (300). However, the sex-specific incidence of first-time VTE also seems to be higher in elderly women than in men (Figure 10). This could be due to biological differences between elderly women and men, or to differences in care-seeking behavior.
In papers II to V, the study population was drawn from the participants of the VIP. In the VIP, the participation rates were higher among women than men, but these differences decreased over time (337). Other European health examination programs have also reported higher participation rates among women compared to men (356). An interest in one’s own health can be a motivating factor leading to participation in a health examination. Higher health examination participation rates in women may reflect the fact that women, in general, are more interested in health-related issues compared to men (357).

In papers IV and V, we found differences between men and women in the association between alcohol consumption, alcohol dependence, physical activity and risk of VTE. Possible causes of these differences are discussed in detail above.

**Clinical implications**

We found that, in Sweden, unvalidated VTE diagnosis codes from administrative registries are not of sufficient quality for use in research. This means that stakeholders wishing to estimate the incidence of VTE in different settings need to use methods that combine data from diagnosis registries with data from other sources. We also estimated the incidence of VTE in an adult population with access to contemporary diagnostic methods for and preventive measures against VTE. Our incidence figures can be helpful in estimating the VTE-related burden of disease in different age groups and for planning expenditure of health care resources for the treatment of VTE. Ultimately, our incidence figures can be used to guide resource allocation within the health care system.

The associations we found between high alcohol consumption, alcohol dependence and increased risk of VTE in men, as well as the associations between high levels of physical activity and lower risk of VTE in women, could be used in a clinical setting where a patient’s pre-test probability of VTE is estimated. However, it must be stressed that until our findings are confirmed by other studies and have been included in a validated prediction model, determination of a patient’s pre-test probability of VTE should be based on established risk scores.

If the associations between high alcohol consumption, alcohol dependence and increased risk of VTE in men are causal, our findings could have implications for recommendations regarding alcohol intake. Our findings could help to motivate the public to keep alcohol intake at a moderate level, and be used by medical professionals when advising patients about alcohol use. Our
observational study design did not, however, allow us to draw conclusions about whether or not the associations we found were indeed causal.

Similarly, if the associations between high levels of physical activity and lower risk of VTE in women are causal, preventive measures against VTE could include efforts to increase the level of physical activity at the population level. It is possible that this would lead to a more pronounced reduction in the incidence of VTE in women compared to men. An increase in the level of physical activity would also have other positive health effects (358).

**Future perspectives**

It would be valuable to perform an updated study to investigate the incidence of VTE in Västerbotten County in a few years’ time. It could then be determined if the finding of an increase in the incidence of VTE, noted by an American study (31), is valid under Swedish conditions.

The best way to corroborate our findings regarding the associations between alcohol consumption, alcohol dependence, physical activity and risk of VTE would be to conduct randomized controlled trials where participants were randomized to different levels of physical activity or alcohol intake. However, to detect associations of similar magnitudes to those found in our studies, such studies would either have to include a very large number of participants or to follow them for a long period of time. This type of approach would be both time-consuming and costly. Furthermore, it would be unethical to randomize participants to harmful interventions, such as high alcohol consumption.

It would be interesting to attempt to confirm our findings regarding the association between alcohol consumption and risk of VTE using a biomarker, such as phosphatidylethanol. That would give a more precise estimate of actual alcohol consumption and would eliminate the risk of bias induced by sex- and age-related differences in the accuracy of questionnaires. One way of determining whether there is a causal association between exposure to alcohol and risk of VTE could be a Mendelian randomization study. This approach has been used in studies of alcohol intake as a risk factor for cancer (359, 360).

It would also be valuable to perform a study of the association between physical activity and risk of VTE where physical activity was measured by accelerometer. That approach would give a combined estimate of leisure time physical activity, including activities not performed in training clothes, and occupational physical activity.
Conclusions

- The incidence of first-time VTE was 137 (95% CI 122–154) per 100,000 adults per year.
- The most common risk markers for VTE at the time of first-time VTE diagnosis were recent hospitalization and concurrent malignancy.
- Registry data on diagnoses of PE, but not DVT, was of acceptable quality and can be considered for use in registry-based studies.
- Fasting plasma glucose, oral glucose tolerance test two-hour post-load plasma glucose and diabetes were not associated with risk of first-time VTE.
- Alcohol intake and alcohol dependence were associated with an increased risk of first-time VTE in men.
- High leisure time physical activity and occupational physical activity were associated with a decreased risk of first-time VTE in women.
Acknowledgements

I would like to express my sincere gratitude to everyone who made this thesis possible:

My supervisor, Lars Johansson, for your advice, support and guidance. Thank you for believing we could do this together, and for never letting me down. Our discussions have taught me that a question can have more than one right answer, and that sometimes no answer is the right one. Thank you for your ideas and your enthusiasm, and for passing some of that enthusiasm on to me.

My co-supervisor, Jan-Håkan Jansson, for your insightful comments and invaluable input on this thesis. Thank you for letting me share your experience in ethics, your knowledge of epidemiology and your wisdom in things great and small. And thank you for your attention to detail. May you have many multi-page tables and freshly prepared ski tracks to enjoy in the future!

My co-supervisor, Marcus Lind, for your support and patience, and for bringing laughter to our meeting discussions. Thank you for your all-important input on this thesis and for your ideas about the next chapter of the VEINS project. Thank you for helping me grow as a person, a doctor and a researcher.

My co-authors Ludvig Öhman, Eva Fhärm, Maria Wennberg and Patrik Wennberg, for your input and thoughts on the papers of this thesis.

My clinical supervisor, Christina Sipola, for your guidance during my residency in internal medicine. Thank you for your unwavering confidence in me, and for always being there for me when I need you the most.

Thomas Suh, for your skillful proofreading.

All my colleagues at the Department of Internal Medicine and Geriatrics at Skellefteå County Hospital, for your support and understanding.

My friends and colleagues at Skellefteå Research Unit: Birgitta Stenmark, Marie Tjernqvist, Mona Olofsson, Kurt Boman, Bengt Norrfors, Lars-Micael Lundmark, Jonas Andersson, Ulf Forsberg, Jenny Hernestähl-Boman and Anette Broberg, for all your help and encouragement.

My family for loving me, no matter what. Without you, none of this would have been possible.
All those who have participated in, performed, planned and funded the Västerbotten Intervention Programme.

The Northern Sweden Diet Database and the funds supporting it, including the Swedish Research Council and the Swedish Research Council for Health, Working Life and Welfare.

The funding sources that made this thesis possible: the County Council of Västerbotten, the Foundation for Medical Research in Skellefteå, and the Joint Committee of Northern Sweden Health Care Region.
References


56. van Adrichem RA, Debeij J, Nelissen RG, Schipper IB, Rosendaal FR, Cannegieter SC. Below-knee cast immobilization and the risk of venous


100


# Appendix

## Corrections to paper I

All corrections are written in bold text.

<table>
<thead>
<tr>
<th>Page, Section</th>
<th>Published text</th>
<th>Corrected text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page 2, Methods</td>
<td>To include all potential VTE events, we performed an extensive search for VTE cases in the National Patient Registry that included ICD-10 diagnosis codes I26, I27.8, I27.9, I67.6, I80, I81, I82, O08.2, O08.7, O22.5, O22.8, O22.9 and O22.2, for inpatients, outpatients, and emergency department visits in all three hospitals in the County of Västerbotten for the year 2006.</td>
<td>To include all potential VTE events, we performed an extensive search for VTE cases in the National Patient Registry that included ICD-10 diagnosis codes I26, I27.8, I27.9, I67.6, I80, I81, I82, O08.2, O08.7, O22.5, O22.8, O22.9, O22.2, <strong>O22.3, O87 and O88</strong>, for inpatients, outpatients, and emergency department visits in all three hospitals in the County of Västerbotten for the year 2006.</td>
</tr>
<tr>
<td>Page 3, Results</td>
<td>The overall proportions of women were 54% for those under age 80 years and 70% for those 80 and over.</td>
<td>The overall proportions of women were <strong>46%</strong> for those under age 80 years and 70% for those 80 and over.</td>
</tr>
<tr>
<td>Page 3, Results</td>
<td>The proportion of first-time VTE cases with none of the registered risk factors increased with age, and it comprised 58% of the oldest age group.</td>
<td>The proportion of first-time VTE cases with none of the registered risk factors increased with age, and it comprised <strong>46%</strong> of the oldest age group.</td>
</tr>
<tr>
<td>Page 6, Discussion</td>
<td>Of the total population of patients with first-time VTE, we could not identify any risk factor in 41%.</td>
<td>Of the total population of patients with first-time VTE, we could not identify any risk factor in <strong>33%</strong>.</td>
</tr>
</tbody>
</table>
In Table 1, "Patient characteristics and risk factors in 281 patients with first-time VTE, grouped by age", the row concerning "No risk factors" reads as follows in the published version:

<table>
<thead>
<tr>
<th></th>
<th>All (n=281)</th>
<th>18-64 years (n=92)</th>
<th>65-79 years (n=100)</th>
<th>80 years and over (n=89)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No risk factors</strong></td>
<td>116 (41.3)</td>
<td>29 (31.5)</td>
<td>35 (35.0)</td>
<td>52 (58.4)</td>
</tr>
</tbody>
</table>

The corrected version reads as follows:

<table>
<thead>
<tr>
<th></th>
<th>All (n=281)</th>
<th>18-64 years (n=92)</th>
<th>65-79 years (n=100)</th>
<th>80 years and over (n=89)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No risk factors</strong></td>
<td><strong>93 (33.1)</strong></td>
<td><strong>24 (26.1)</strong></td>
<td><strong>28 (28.0)</strong></td>
<td><strong>41 (46.1)</strong></td>
</tr>
</tbody>
</table>