

**UiT**

**THE ARCTIC  
UNIVERSITY  
OF NORWAY**

# **Red cell distribution width and risk of venous thromboembolism**

---

**Trygve Sølberg Ellingsen**

A dissertation for the degree of Philosophiae Doctor

March 2018

Faculty of Health Sciences, Department of Clinical Medicine

————— TREC —————

K.G. JEBSEN THROMBOSIS  
RESEARCH AND EXPERTISE CENTER

—————





## Table of contents

<b>Acknowledgements</b> .....	<b>3</b>
<b>List of papers</b> .....	<b>5</b>
<b>Summary</b> .....	<b>6</b>
<b>Sammendrag</b> .....	<b>7</b>
<b>Abbreviations</b> .....	<b>8</b>
<b>1. Introduction</b> .....	<b>9</b>
1.1 Epidemiology of venous thrombosis .....	9
1.2 Pathophysiology of venous thrombosis .....	12
<b>2. Risk factors</b> .....	<b>14</b>
2.1 Risk factors for VTE .....	14
2.1.1 <i>Inherited risk factors for VTE</i> .....	15
2.1.2 <i>Acquired risk factors for VTE</i> .....	17
2.2 Red cell distribution width and risk of diseases .....	21
2.2.1 <i>Iron metabolism and risk of VTE</i> .....	23
<b>3. Aims of the thesis</b> .....	<b>27</b>
<b>4. Study Population and Methods</b> .....	<b>28</b>
4.1 The Tromsø Study .....	28
4.2 Baseline measurements (Tromsø 4) .....	29
4.3 Outcome measures .....	30
4.3.1 <i>Identification and validation of venous thromboembolic events</i> .....	30
4.3.2 <i>Identification and validation of cancer</i> .....	31
4.3.3 <i>Identification and validation of myocardial infarction</i> .....	32
4.3.4 <i>Identification and validation of stroke</i> .....	33
<b>5. Main results</b> .....	<b>34</b>
5.1 Paper I .....	34
5.2 Paper II .....	35
5.3 Paper III .....	36
5.4 Paper IV .....	37
<b>6. General discussion</b> .....	<b>38</b>
6.1 Methodological considerations .....	38
6.1.1 <i>Study design</i> .....	38
6.1.2 <i>Information bias, misclassification and regression dilution bias</i> .....	40
6.1.3 <i>Confounding and effect modification</i> .....	43
6.1.4 <i>Multivariable regression models</i> .....	45
6.1.5 <i>Detection and validation of outcomes</i> .....	47

6.1.6 Selection bias, missing variables and external validity.....	47
<b>7. Discussion of main results.....</b>	<b>51</b>
7.1 RDW as a biomarker for risk of VTE, cancer and mortality.....	51
7.2 Possible underlying mechanisms for the association between RDW and VTE .....	53
7.2.1 Prothrombotic effects of RDW.....	54
7.2.2 Intermediate development of other diseases.....	54
7.2.3 Iron metabolism.....	55
7.2.4 Inflammation and oxidative stress.....	57
7.2.5 Other possible mechanisms .....	58
<b>8. Conclusions .....</b>	<b>59</b>
<b>9. Implications of results and future perspectives .....</b>	<b>60</b>
<b>10. References .....</b>	<b>62</b>

**Papers I-IV**



## Acknowledgements

The present work was carried out at the Hematological Research Group (HERG) and the K.G. Jebsen Thrombosis Research and Expertise Center (TREC), Department of Clinical Medicine, UiT – The Arctic University of Norway, from August 2013 to February 2018. During this period, I have been part of the MD PhD program for medical students (2013-2017), and for the last 7 months, I have worked as a PhD-student funded by the Northern Norway Health Authority.

First and foremost, I want to thank my main supervisor, Professor John-Bjarne Hansen. I am very grateful that you convinced me to join HERG when I started the MD/PhD program. Since then, you have provided tremendous support, and your scientific knowledge and capacity for hard work is truly impressive. Your contribution to the writing of all four manuscripts as well as this thesis has been invaluable. Also, I appreciate our discussions about the only thing more important than venous thrombosis, namely the (terrible) state of the Norwegian cross country ski team.

Second, I would like to express my gratitude to my co-supervisor, Associate Professor Sigrid K. Brækkan. Thank you for patiently teaching me all there is to know, at least in my mind, about statistics and epidemiology. Despite your busy schedule, you always find time to help me when I come knocking on your door, and you somehow manage to keep a cheerful and smiling attitude even if it is the eight time that hour. You and Ina Høiland also deserve recognition for keeping all of us fit and healthy with your weekly TRECexercise.

I would also like to thank my co-authors Jostein Lappegård, Tove Skjelbakken, Thor Ueland, Pål Aukrust, Ellisiv Mathiesen and Inger Njølstad for valuable contributions along the way. A special thanks to my officemate Jostein for providing excellent company and

friendship throughout our PhD journeys. I value our scientific and non-scientific discussions, and I really, really enjoy the few times I beat you in table tennis.

All past and present HERG/TREC members deserve a big thank you for creating a wonderful social and scientific environment. Caroline, Gunhild and the original members of the basement office (Lars, Hilde, Kristine, Trond and Trond) deserve special thanks for guiding and supporting me through my first steps in the somewhat scary world of STATA and medical statistics.

I also want to express my gratitude to the staff and participants of the Tromsø Study for making this research possible.

Finally, I want to thank my family and friends for enduring my endless talk about red cell sizes and methodological challenges you clearly have no interest in hearing about. My “non-medical” friends deserve praise for keeping a part of me active outside the world of medicine and research. A special thanks to my parents for valuable scientific advices despite your very limited medical knowledge.

Trygve

Tromsø, March 2018

## List of papers

- I. Red cell distribution width is associated with incident venous thromboembolism (VTE) and case-fatality after VTE in a general population.  
Ellingsen TS, Lappegård J, Skjelbakken T, Brækkan SK, Hansen JB.  
*Thrombosis and Haemostasis*. 2015 January;113(1) 193-200.
  
- II. Plasma hepcidin is associated with future risk of venous thromboembolism.  
Ellingsen TS, Lappegård J, Ueland T, Aukrust P, Brækkan SK, Hansen JB.  
*Manuscript*
  
- III. Impact of red cell distribution width on future risk of cancer and all-cause mortality among cancer patients – the Tromsø Study  
Ellingsen TS, Lappegård J, Skjelbakken T, Brækkan SK, Hansen JB.  
*Haematologica*. 2015 October; 100: e387-e389
  
- IV. The association between red cell distribution width and venous thromboembolism is not explained by myocardial infarction, stroke, or cancer.  
Ellingsen TS, Lappegård J, Skjelbakken T, Mathiesen EB, Njølstad I, Brækkan SK, Hansen JB.  
*Research and Practice in Thrombosis and Haemostasis*. 2018 January; 00:1–7.

## Summary

Venous thromboembolism (VTE) is a collective term for deep vein thrombosis and pulmonary embolism. VTE is the third most common cardiovascular disease, causing significant morbidity and mortality. Despite preventive strategies, the incidence of VTE has been stable or slightly increasing during the last decades, affecting 1-2 per 1000 individuals each year. Up to half of all VTE patients have no obvious cause of the disease. Thus, identifying novel biomarkers and unraveling underlying disease mechanisms might help diminish the health burden of VTE. Red cell distribution width (RDW) is a measure of the variability of size of the circulating red blood cells. It can increase due to various conditions that alter the bone marrow's production of red blood cells, such as iron deficiency. Recent years, RDW has been associated with risk of several diseases. The first aim of this thesis was to investigate whether RDW is associated with future risk of VTE and mortality among VTE patients. Secondly, we aimed to investigate whether the association could be explained by underlying iron deficiency or intermediate development of other diseases.

All papers in this thesis use data from the Tromsø Study, a large population-based cohort study. Our study populations were recruited from the fourth survey. For paper I, subjects were followed from date of inclusion in 1994/95 through January 1, 2012. In paper III and IV, study participants were followed from date of inclusion through December 31, 2010. Paper II is a nested case-control study with cases and controls selected among participants in Tromsø 4. Validated events of VTE, cancer, myocardial infarction (MI) and stroke were registered throughout the study periods.

We found that RDW was associated with future risk of VTE and all-cause mortality among VTE-patients. The association could not be explained by underlying iron deficiency. Further, RDW was associated with future risk of cancer, but intermediate development of cancer, stroke or MI did not explain the association between RDW and VTE.

## Sammendrag

Venøs blodpropp eller venøs tromboembolisme (VTE) er en fellesbetegnelse på tilstandene dyp venetrombose og lungeemboli. VTE er verdens tredje vanligste hjerte- og karsykdom, og forårsaker betydelig morbiditet og mortalitet. Til tross for økt fokus på forebygging, har forekomsten vært stabil eller økende de siste tiårene. Hvert år rammes 1-2 av 1000 personer, og halvparten av de rammede har ingen åpenbar årsak. Nye bio- og risikomarkører kan bidra til bedre forståelse og forebygging av sykdommen, og kan dermed redusere de helsemessige konsekvensene. Graden av variasjon i størrelse på de røde blodcellene kan kvantifiseres som «Red Cell distribution Width» (RDW). RDW kan øke som en følge av ulike tilstander som påvirker beinmargens produksjon av røde blodceller, som for eksempel jernmangel. De siste årene har man funnet en sammenheng mellom RDW og risiko for flere sykdommer. Formålet med denne avhandlingen var å undersøke om RDW kan forutsi fremtidig risiko for VTE, samt dødelighet blant VTE-pasienter. Videre ønsket vi å undersøke om bakenforliggende jernmangel eller utvikling av andre sykdommer kunne forklare sammenhengen.

Alle artiklene i avhandlingen er basert på Tromsøundersøkelsen, som er en stor kohortestudie. Vi har benyttet data fra den fjerde undersøkelsen (Tromsø 4), som ble gjennomført i 1994-95. I den første artikkelen ble studiedeltakerne fulgt fra studiestart til og med 1. januar 2012. I den tredje og fjerde artikkelen ble deltakerne fulgt til og med 31. desember 2010. Den andre artikkelen er en nestet kasus-kontroll-studie med deltakere rekruttert fra Tromsø 4. Alle tilfeller av VTE, hjerteinfarkt, slag og kreft ble registret og validert gjennom studieperiodene.

Vi fant en sammenheng mellom høy RDW og økt risiko for fremtidig VTE, samt risiko for dødelighet blant VTE-pasienter. Sammenhengen kom ikke som en følge av jernmangel. Vi fant også en sammenheng mellom høy RDW og risiko for kreft, men hverken utvikling av kreft, hjerneslag eller hjerteinfarkt kunne forklare sammenhengen mellom RDW og VTE.



## Abbreviations

ACT	anticoagulant treatment
AUC	area under receiver-operating curve
BMI	body mass index
CI	confidence interval
CRP	C-reactive protein
CTEPH	chronic thromboembolic pulmonary hypertension
CTPA	computed tomography pulmonary angiography
CTPH	chronic thromboembolic pulmonary hypertension
DVT	deep vein thrombosis
FtH	ferritin heavy-chain
FtL	ferritin light-chain
FVIII	coagulation factor VIII (8)
FVL	factor V Leiden
GWAS	genome-wide association studies
HR	Hazard ratio
HUNT	The Nord-Trøndelag Health Study
LITE	the Longitudinal Investigation of Thromboembolism Etiology study
MCV	mean corpuscular volume
MI	myocardial infarction
MV	microvesicles
NETs	neutrophil extracellular traps
OR	odds ratio
PE	pulmonary embolism
PTS	post-thrombotic syndrome
RDW	red cell distribution width
SD	standard deviation
SNPs	single-nucleotide polymorphisms
TF	tissue factor
VTE	venous thromboembolism

# 1. Introduction

## 1.1 Epidemiology of venous thrombosis

Deep vein thrombosis (DVT) is the formation of a blood clot in a deep vein, usually in the lower limbs, obstructing the venous flow. Pulmonary embolism (PE) might occur secondary to a DVT if a part of the thrombus, an emboli, breaks loose and is carried by the bloodstream. emboli). As the pulmonary arteries decrease in size, the emboli is eventually wedged, obstructing the blood flow. However, no detectable DVT is found in up to half of all PE patients investigated with ultrasonography or magnetic resonance imaging (MRI) (1-3). In these cases, the entire thrombus might have embolized or dissolved. Alternatively, the emboli might have formed in the right atrium of the heart, or the thrombus could arise de novo in the lungs. Atrial fibrillation was associated with higher risk of PE than DVT, and explained 20% of PE events in the Tromsø study (4). Venous thromboembolism (VTE) is a collective term for DVT and PE. VTE is the third most common cardiovascular disease, causing significant morbidity and mortality (5). The incidence rate of VTE increases markedly with age, from one per 10 000 person-years in young adults to one per 100 person-years in the elderly (6, 7). Despite preventive strategies, the incidence has been stable or slightly increasing during the last decades (7, 8). Young women are more often affected than young men, while older men are more affected than older women (9).

VTE is a disease with serious acute and long-term complications. PE is the most common avoidable cause of death among hospitalized patients (10). Approximately one in every four PE-event is lethal (11), and the mortality rate remains increased up to 8 years following a VTE-event (12). In observational studies, 5-10% of VTE-patients died within one month, and 20-30% died within 5 years (6, 13-15). The mortality is highest for patients with underlying malignancy. The one-year survival rate in patients with cancer-related VTE was

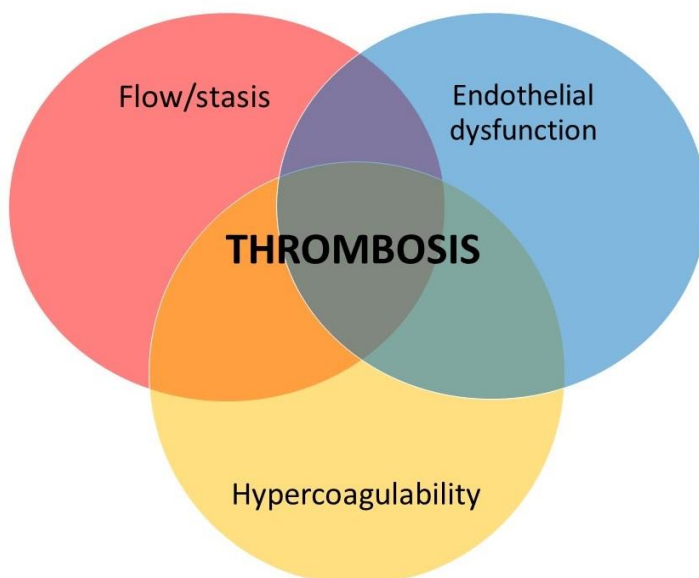
12% in a Danish population-based study (16). In addition to increased mortality, a high proportion of VTE-patients suffer from long-term complications, such as recurrent events, post-thrombotic syndrome (PTS) and/or chronic thromboembolic pulmonary hypertension (CTEPH). The cumulative proportion of patients with VTE recurrence is approximately 30% after 10 years (17, 18), with highest risk the first 6-12 months (13, 17). Transient risk factors, such as surgery, is associated with a low risk of recurrence, while the risk of recurrence is high for patients with persisting risk factors, such as underlying malignancy or inherited thrombophilias, as well as for patients with unprovoked VTE (19-21). Moreover, the recurrent event tend to occur at the same location as the previous event (i.e. PE-patients develop a new PE rather than a DVT) (22). In the California Patient Discharge Data Set, 6.4% of DVT patients developed recurrent VTE within 6 months of hospital discharge, of which 85% developed a second DVT (23). In the same study, 70% of PE patients admitted with recurrent VTE were diagnosed with a PE. CTEPH is defined as chronic (more than 6 months) hypertension in arteria pulmonalis following a PE-event (11). The condition affects 2-4% of PE-patients, often causing dyspnea and progressive heart failure (11, 24). PTS is a condition characterized by chronic pain, edema and dermatitis, and severely affected patients may develop venous leg ulcers (25, 26). PTS affects 20 to 50% of DVT patients, and risk factors include obesity, female sex, proximal DVT location and recurrent DVTs (18, 27). PTS may considerably impair mobility and quality of life (25), which may explain the association between DVT and subsequent work-related disability (28).

The VTE diagnosis is set using a combination of clinical prediction scores, laboratory tests and radiological imaging. For DVT, assessment of the clinical Wells score is combined with levels of D-dimer to exclude the diagnosis, while ultrasound is considered to be a confirmatory diagnostic test (29). Similarly, clinical prediction tools and D-dimer might rule out PE, whereas computed tomography pulmonary angiography (CTPA) confirms the diagnosis

(30). The increasing use of CTPA has led to an increased incidence of PE (31). In a study from the Nationwide Inpatient Sample and Multiple Cause-of-Death databases, the implementation of CTPA led to a 81% increase in the incidence rate of PE, a minimal change in mortality, and a 71% increase in presumed complications of anticoagulation (32). The high sensitivity of CTPA has led to increased ability to detect less severe PEs with unknown clinical significance (33). Thus, several PE-patients receive potential harmful treatment without clear benefit (34). New biomarkers might aid clinicians in VTE risk assessment, and might improve the prognostic prediction. The identification of novel risk factors and prognostic markers can potentially improve the benefit-to-harm ratio for both prevention and treatment of VTE.

## 1.2 Pathophysiology of venous thrombosis

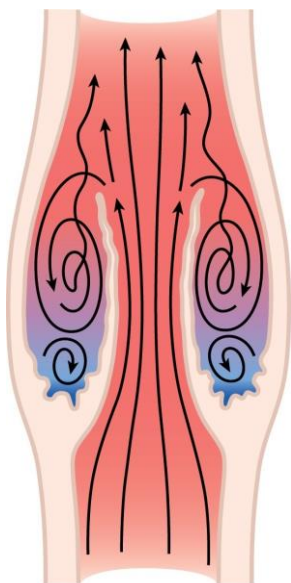
The human body is equipped with a sophisticated system for maintaining a physiological balance between bleeding and thrombosis. Arterial clots are formed under high shear stress, typically after rupture of an atherosclerotic plaque or other damage to the vessel wall (35). If the vessel wall is injured, prothrombotic elements (such as collagen, ADP and tissue factor (TF)) are exposed on the luminal side, leading to accumulation and activation of platelets, activation of the coagulation cascade, and ultimately the formation of a thrombus (36). Venous blood clots form under low shear stress on the surface of a largely intact endothelium (37), and the exact mechanisms are not as well established as for arterial blood clots. The German scientist Rudolph Virchow proposed as early as 1856 that there is a triad of elements central in the pathogenesis of venous thrombosis (Figure 1). The triad, known as Virchow's triad, consists of stasis of the blood flow, changes in the blood composition (i.e. hypercoagulability) and endothelial damage or dysfunction. Most of the recognized risk factors for VTE connect with one or more of these key elements.



**Figure 1.** Virchow's triad illustrating the three central elements in the pathogenesis of venous thrombosis



DVTs most often develop in the sinuses of the venous valves (38). The endothelium contains natural anticoagulants like endothelial protein C-receptor, tissue factor pathway inhibitor, thrombomodulin and heparin-like proteoglycans. In the large veins, the ratio between the endothelial surface area and blood volume is low, and the anticoagulant properties of the endothelium is less effective than in smaller veins (39-41). In addition, the turbulent blood flow and the effect of gravity may lead to the formation of secondary vortexes with localized hypoxia in the venous valve pocket (Figure 2) (42). The innermost layer of the endothelium receives its oxygen supply from the luminal erythrocytes. Hypoxia might cause endothelial dysfunction with downregulation of the endothelial anticoagulant properties, promoting prothrombotic processes (43, 44). Moreover, hypoxia promotes the activation of leukocytes and platelets, which in turn might release TF-containing microvesicles (37). The activated leukocytes might express TF on the cell surface, recruit platelets and release neutrophil extracellular traps (NETs), and is suggested to play a key role in inflammatory mediated thrombosis (45, 46). However, the underlying mechanisms between microvesicles, NETs and VTE remain to be established.

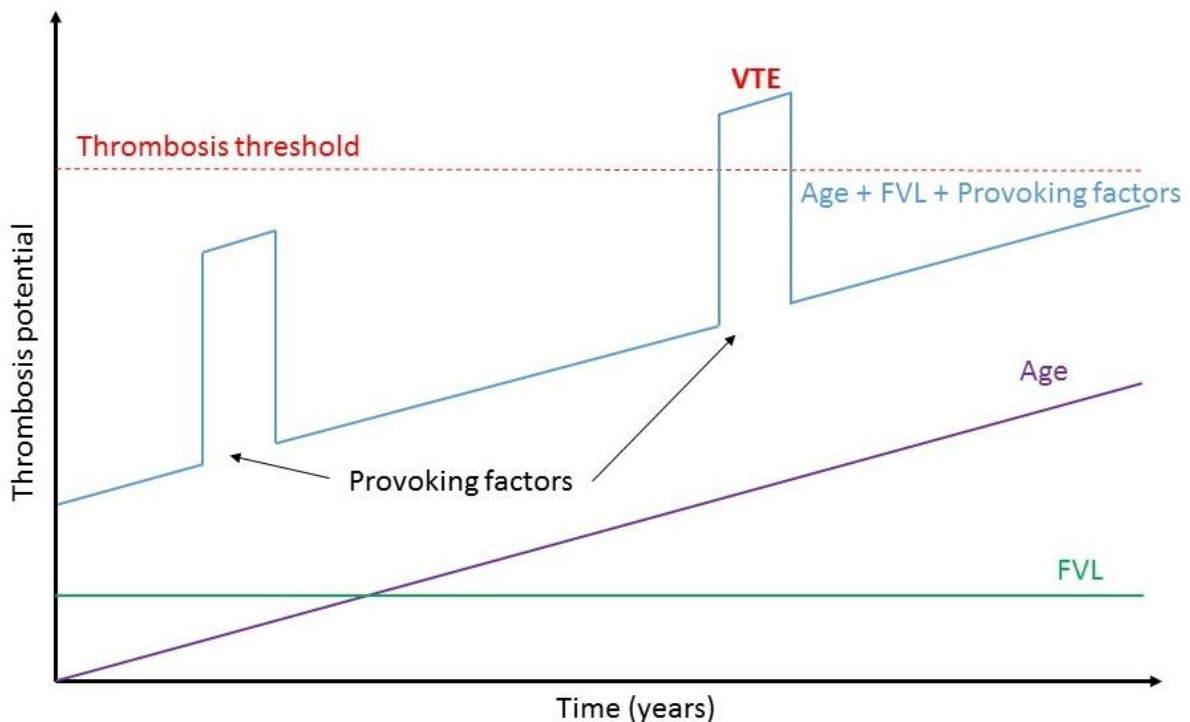


**Figure 2.** Schematic representation of the blood flow observed around the venous valves of the deep veins. Oxygen tension is color coded with a gradient from red to blue: the darker the blue, the greater the hypoxia. The solid arrows illustrate blood flow direction with the development of vortexes in the venous valve pocket (42).

## 2. Risk factors

### 2.1 Risk factors for VTE

VTE is a complex disease with several known risk factors, but the underlying cause is still unknown in up to 50% of the cases (47). VTE is described as a multifactorial disease requiring the presence of more than one risk factor at the same time for an event to occur (48), as illustrated in the thrombosis potential model (Figure 3). A single risk factor is usually not sufficient for a thrombus to form, but the effect of multiple risk factors might overpower the physiological antithrombotic mechanisms.



**Figure 3.** The thrombosis potential model modified from Frits Rosendaal (48). The green line represents an inherited risk factor such as factor V Leiden (FVL) that is stable over time. The purple line represents the effect of a risk factor that increases over time, such as age. The blue line represents the joint effect of an inherited risk factor, age and provoking events. The figure illustrates how several risk factors for VTE might be required to reach the thrombosis threshold.

### 2.1.1 Inherited risk factors for VTE

Studies have indicated that genetic factors account for up to 60% of all VTE events (49, 50), and a family history of VTE is associated with a two- to threefold increase in risk of VTE (51-54). Inherited thrombophilias are generally caused by mutations that either decrease endogenous antithrombotic factors (*loss-of-function*), or potentiate prothrombotic factors (*gain-of-function*) (55). Loss-of-function mutations are quite rare (approximately 1% of the population), but are associated with high risk of VTE (56-58). Mutations causing deficiencies in **antithrombin, protein C and protein S** are associated with a 4 to 30-fold increased risk of VTE, but as they are rare in the total population, these mutations account for only a small proportion of VTE-events (55, 59).

Gain-of-function mutations are generally associated with a weak to moderate increase in risk of VTE, and are rather common in the population. **Factor V Leiden** is a mutation causing resistance to the anticoagulant effect of protein C (60), and is seen in approximately 5% of people of Caucasian heritage (59). Heterozygous carriers have a two to fivefold increased risk of VTE, while homozygous carriers have up to 12-fold increased risk (59, 61, 62). **Prothrombin G20210A** is a mutation causing increased levels of prothrombin and is associated with up to threefold increased risk of VTE for heterozygous carriers (59, 62). The mutation is present in 1-2% of the population, with some geographical variation (63). The most prevalent inherited risk factor for VTE, **non-O blood type**, is present in as much as 60% of the population (64), and is associated with up to twofold higher risk of VTE than people with O blood type (59). As several of the inherited risk factors for VTE are rather common, multiple risk factors might occur in the same person, conducting a synergistic effect on the risk of VTE. For instance, those heterozygous for both FVL and prothrombin G20210A have a 20-fold higher risk of VTE than the general population (65).

As mentioned, family studies indicate that genetic predisposition accounts for 60% of the VTE events. However, the known genetic variants only account for 10-20% of all VTEs (49), and there are likely several unknown inherited risk factors. During the last decades, genome-wide association studies (GWAS) have identified several novel single-nucleotide polymorphisms (SNPs) associated with VTE. GWAS test the association of a huge number of SNPs in participants classified by clinical phenotype (e.g. VTE), and is able to detect common alleles with modest phenotypic effects (66). Indeed, most of the newly discovered SNPs have only modest effect on the VTE risk and the clinical relevance is limited. De Haan and colleagues proposed the inclusion of selected SNPs in a VTE risk prediction model (67). Inclusion of the five most strongly associated SNPs to a non-genetic risk prediction model significantly improved the performance of the model (areas under receiver-operating curve (AUC) increased from 0.77 to 0.82). The risk prediction model included recent leg injury, surgery, pregnancy/postpartum, immobilization, oral contraceptive use, hormone replacement therapy, obesity and cancer. However, a validation study of the five SNP weighted risk score reported AUC of 0.59 in whites and 0.56 in African Americans (68). Nevertheless, these type of genetic risk scores might be clinically useful and cost-effective in high-risk persons (such as cancer or trauma patients), and novel biomarkers can prove helpful in identifying persons who will benefit from genetic profiling (67).

### *2.1.2 Acquired risk factors for VTE*

Age and gender: The incidence of VTE increases exponentially by increasing age (6, 7, 69). In The Tromsø-study, persons above 70 years have 11-fold higher risk of VTE than those below 50 years (70). Similar results were found in The Nord-Trøndelag Health Study (HUNT) and the Longitudinal Investigation of Thromboembolism Etiology (LITE) study (69, 71). The incidence of hospitalization doubles with age, and hospitalization might account for 40-50% of VTE-events in the elderly (72). Moreover, ageing may be associated with increased prevalence of conventional risk factors (e.g. immobility, malignancy, co-morbidity and haemostatic factors) (72-74). However, further studies are warranted to identify age-specific risk factors for VTE, as well as possible interactions between established risk factors and high age.

In the LITE study, which consists of participants aged 45 years and older, men had higher incidence of VTE than women (69). In contrast, the incidence rate of VTE was slightly higher for women than men in the HUNT study (71). The HUNT study consist of subjects 20 years or older. No association between sex and risk of VTE was seen in the Tromsø Study (75). Women have increased risk of VTE during pregnancy (76), and use of estrogen in combined oral contraceptives or as hormone replacement therapy is associated with a two to threefold increased risk of VTE (77-79). This may explain the different risk of VTE in men and women of different age groups.

Obesity and body height: Worldwide, the obesity (BMI>30) prevalence has nearly tripled since 1975, and 13% of adults above 18 years were obese in 2016 (80). Obesity is associated with two to threefold increased risk of VTE (75, 81, 82), and there is a dose-response relationship between increasing BMI and risk of VTE (81). In the Tromsø study, the VTE risk increased by 30% per 1 SD increase in BMI (83), and further weight gain was associated with further increased VTE risk (84). Results from two Mendelian randomization studies imply a



causal relationship between obesity and VTE (85, 86). Obese people have higher concentrations of plasminogen activator inhibitor, TF, fibrinogen, Factor VIII and von Willebrand factor (87). Adipose tissue secrete leptin, a hormone found to promote thrombosis by increased platelet aggregation and TF expression (88). In addition, obesity might promote venous thrombosis by impaired venous return due to increased intra-abdominal pressure (89). Moreover, C-reactive protein (CRP) was associated with both obesity, myocardial infarction (MI) and VTE among female participants in the Tromsø-study, suggesting that inflammation might be a shared pathway for MI and VTE in obese patients (90). Tall men have higher risk of VTE than those of short stature, and there is a synergistic effect of body height and weight on the risk of VTE (5, 91, 92). Tall stature might cause venous stasis, and it increases the number of venous valves. However, the exact mechanisms for the association between height and VTE is unknown.

Cancer: The association between cancer and VTE was first described by Jean Baptiste Bouillaud in 1823 (93). The French physician Armand Trousseau described the phenomena of migratory thrombophlebitis (inflammation in the vessel wall at various locations) as an early sign of cancer (94), a sign later known as Trousseau's syndrome (95). Approximately 20% of incident VTEs are associated with underlying malignancy, and cancer patients have a four to sevenfold increased risk of VTE (96, 97). Malignant diseases might increase the activation of platelets, reduce the synthesis of antithrombotic factors and increase the half-life of coagulation factors (98). Cancer cells might express TF, or produce TF-containing microvesicles (99). Moreover, tumor growth can damage or compress veins (100), and cancer patients are exposed to numerous risk factors for VTE such as surgery, infectious diseases and immobilization. Khorana et al developed a predictive model for VTE in cancer patients in 2008 (101), and several risk assessment models have been developed during the following years (102). However, in an external validation study, none of the included models had sufficient discriminating properties to justify introduction in clinical practice (103).

Trauma, surgery and acute medical conditions: The incidence of in-hospital VTE is more than 100 times greater than the incidence in the general population (104), and 25-60% of VTE-events occur in either hospitalized, institutionalized or recently hospitalized patients (105-107). The risk of VTE remains elevated for a substantial period of time after discharge, and might even be higher during the first three months after discharge than during the admission period (108). Between 70 and 80% of all fatal PE events occur in hospitalized medical patients (109). Surgery caused more than 40% of all provoked VTE events in the LITE study (6). Patients admitted to a trauma unit had 50% risk of developing a DVT without prophylactic treatment (110), and the risk of VTE was increased also in trauma patients receiving prophylactic anticoagulant treatment (111). VTE prevention in hospitalized patients is challenging, as these patients are at high risk of both bleeding and thrombosis. Several risk assessment models have been developed to improve VTE prevention in hospitalized medical patients, but the majority have limited generalizability and validation (112, 113). The Padua prediction score is perhaps the most recognized. The Padua score consists of 11 clinical features and was developed to discriminate between medical patients with high and low risk of VTE (114). In the original paper, patients with a high Padua score ( $\geq 4$ ) had a 32-fold higher risk of VTE than patients with a Padua score  $< 4$ . In a large, external validation study, patients with a Padua score  $\geq 4$  had only threefold higher risk of VTE than patients with a Padua score  $< 4$  (112), and the Padua score demonstrated a moderate degree of discrimination (Harrell's C-index 0.60). Similar results were found in another large, multicenter validation study (115)

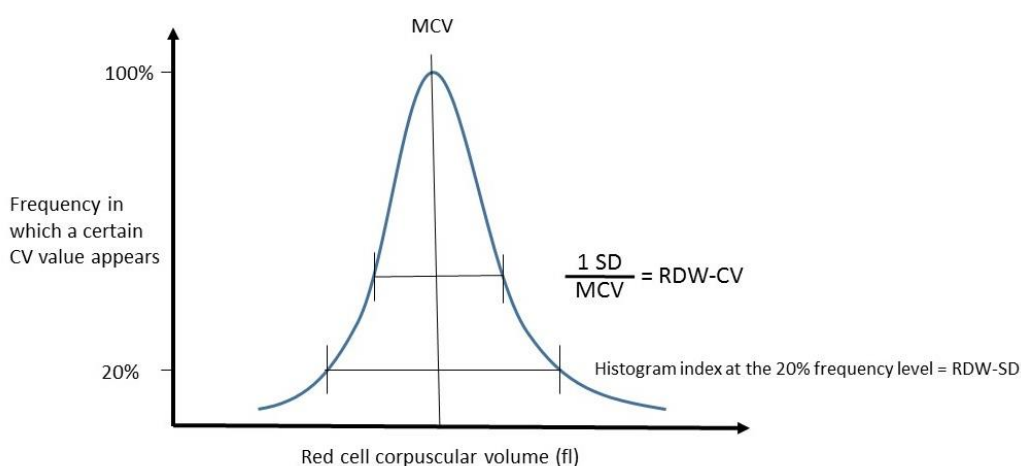
Immobilization: Hospitalized patients tend to be immobilized, and immobilization is a known risk factor for VTE (116). However, the risk of VTE by hospitalization far exceeds the risk of VTE by immobilization alone, and underlying disease is a major and independent cause of VTE. Long haul traveling (airplane, bus, train, car) is associated with an almost threefold increased risk of VTE, and the risk increases with increasing travel time (117).

Gene-environmental interactions: Several of the inherited and acquired risk factors for VTE display a synergistic relationship, i.e. the effect of the combined risk factors exceeds the sum of the effect from the individual components. For instance, patients with factor V Leiden have approximately a 4-fold increased risk of VTE, and women taking oral contraceptives have approximately a 3-fold increased risk of VTE, while patients with factor V Leiden taking oral contraceptives have a 15-fold increased risk of VTE (118). Women with inherited thrombophilias are at high risk of thrombosis during pregnancy and puerperium (48), and a synergistic effect has been described for obesity, inherited thrombophilia and use of oral contraceptives (82, 119, 120).

Despite the discovery of several inherited and acquired risk factors, up to 50% of all VTE events are unprovoked, i.e. the patients have no obvious risk factors triggering the event (47). Moreover, the incident rate of VTE has not decreased during the last decades (8, 11), and identification of novel risk factors and biomarkers could improve our understanding, prevention and treatment of the disease.

## 2.2 Red cell distribution width and risk of diseases

The British hematologist Cecil Price-Jones published a paper in 1922 in which he described a method for calculating the coefficient of the variation in the diameter of red blood cells (121). A Price-Jones curve refers to a distribution curve of the measured diameter of red blood cells, whereas the red cell distribution width express the width of the distribution curve of the measured volume of the red blood cells (i.e. the variation of corpuscular volume) (122). RDW can be calculated by a few different approaches. Most automated blood cell counters calculate the RDW-coefficient of variation (RDW-CV). (Figure 4)(122). RDW-CV is calculated by dividing the standard deviation of the mean corpuscular volume (MCV) by MCV and multiplying by 100 to yield a percentage value (122). As MCV is in the equation, a homogenous cell population with low MCV might have increased RDW, whereas a heterogeneous cell population with high MCV might have a normal RDW. The RDW-standard deviation (RDW-SD) is not dependent on MCV, and thus a more accurate reflection of red cell size variance. Nevertheless, as the vast majority of published studies and clinicians use the RDW-CV, I will refer to RDW-CV when I discuss RDW in this thesis.



**Figure 4:** Obtainment of RDW-CV and RDW-SD from a tentative erythrocyte volume distribution histogram. RDW-CV: Red cell distribution width – coefficient of variation; RDW-SD: Red cell distribution width – standard deviation; CV: corpuscular volume; MCV: Mean corpuscular volume. Modified from Caporal et al (122).

RDW increases with increasing age, but no consistent association is found between RDW and sex (123-125). People of African origin have higher RDW than Caucasians (125-127), and RDW is increased during pregnancy (128, 129). In a study of 26 individuals having monthly blood samples during one calendar year, the intra-individual biological variability in RDW was 3.4%, while the variability due to monthly differences was 1.6% (130). To date, no condition has been associated with an abnormally low RDW, but several diseases are associated with increased RDW (discussed in detail later). RDW has traditionally been used in a classification system for anemias developed by Bessman and colleagues in 1983 (131). High RDW may be caused by deficiencies in iron, B12 or folate (Table 1), but can also be caused by hemoglobinopathies, hemolysis and hemolytic anemia (132-134).

	Low MCV	Normal MCV	High MCV
Normal RDW	Thalassemia minor Chronic disease	Chronic disease Hereditary spherocytosis Hemorrhage	Aplastic anemia Myelodysplastic syndrome
High RDW	Established iron deficiency. Red cell fragmentation	Early iron, B12 or folate deficiency Hemoglobinopathies	Established B 12 or folate deficiency Hemolytic anemia

**Table 1.** Classification system for anemias by RDW and MCV. Modified from Bessman et al (131)

Recent studies have found several associations between RDW and human pathology. In 2007, Felker et al. published the first study suggesting that RDW could be used in clinics beyond the differential diagnosis of anemia (135). A broad range of laboratory measures, including RDW, was measured in 2679 symptomatic heart failure patients from the North American Candesartan in Heart Failure: Assessment of Reduction in Mortality and Morbidity (CHARM) program and 2140 patients with symptomatic heart failure from the Duke Databank. During a median follow-up of 34 months, 952 of the patients in the CHARM program suffered



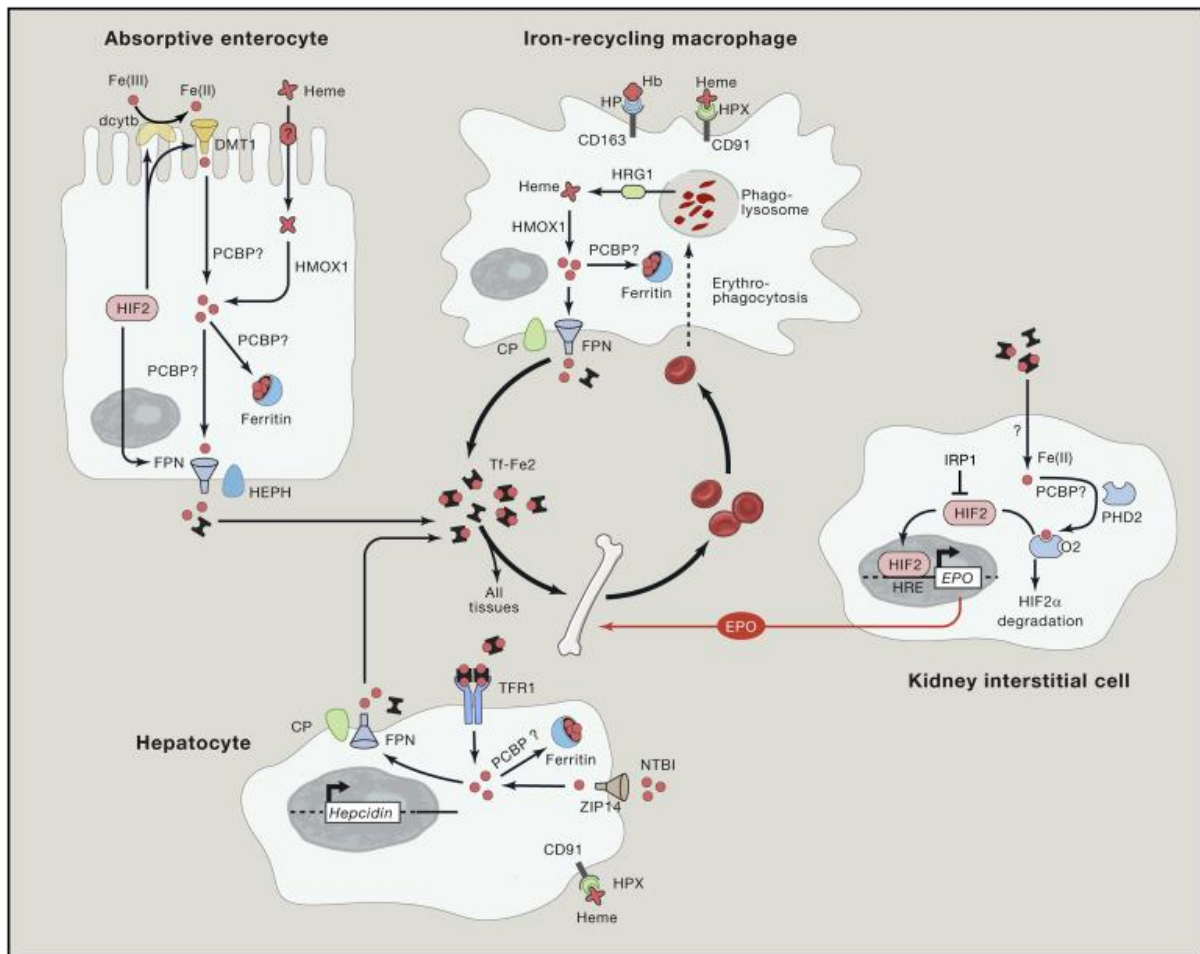
a primary outcome (cardiovascular death or heart failure hospitalization). One standard deviation (SD) increase in RDW was associated with a 17% increased risk of adverse outcome (HR 1.17, 95% CI 1.10-1.25). The results were replicated in patients from the Duke Databank: RDW was associated with increased risk of all-cause mortality (HR 1.29 per 1-SD increase, 95% CI 1.16-1.43) (135). The study by Felker and colleagues was followed by several studies demonstrating that RDW was associated with risk of mortality and/or cardiovascular events in patients with cardiovascular disease (136-139), as well as in the general population (126, 140, 141).

Only a few studies have investigated the association between RDW and VTE. In 2013, a small case-control study by Cay et al. found that increased RDW was associated with presence of DVT (142). These findings were supported by findings from the Multiple Environmental and Genetic Assessment of risk factors for venous thrombosis (MEGA) study (143). In the latter study, individuals with RDW above the 95<sup>th</sup> percentile (RDW>14.1%) had a threefold higher risk of VTE compared to those with RDW values between the 5<sup>th</sup> and 95<sup>th</sup> percentiles (RDW range 11.7-14.1%). In both studies, the thrombotic event preceded the blood sampling. It was therefore impossible to determine whether high RDW was an actual risk factor or a consequence of VTE. An acute VTE event is accompanied by a prolonged inflammatory response, and inflammatory processes might influence RDW (139, 144). The effect of RDW on risk of incident VTE, VTE recurrence and mortality among VTE patients has not been studied in a prospective study.

### *2.2.1 Iron metabolism and risk of VTE*

Iron has the capacity to accept and donate electrons, interconverting between ferric (Fe<sup>2+</sup>) and ferrous (Fe<sup>3+</sup>) forms (145), and iron is useful and required for adequate erythrocyte function, oxidative metabolism and many enzymes (146). However, iron can also damage

tissues by catalyzing the formation of free-radical ions that attack cellular membranes, proteins and DNA (145). Therefore, the iron metabolism is carefully regulated (Figure 5) (147).



**Figure 5:** Systemic iron homeostasis. Enterocytes take up iron from the intestinal lumen. Iron is exported into the bloodstream via ferroportin (FPN), a transporter regulated by hepcidin. Iron is transported in the bloodstream bound to transferrin (Tf-Fe<sub>2</sub>), which binds to the transferrin-receptor in iron demanding tissue, and is mostly used in the hemoglobinization of new red blood cells (RBC). Hepatocytes store excess iron as ferritin and express FPN for iron export. Iron-recycling macrophages engulf and degrade senescent RBCs to release heme into the phagolysosome. Heme is then processed to release iron which is either stored in ferritin or exported into the circulation via FPN with the assistance of ceruloplasmin (CP) ferroxidase. Fibroblasts of the kidney sense iron and oxygen deficiency and release erythropoietin (EPO) to enhance erythropoiesis. Iron utilization by the erythroid marrow and its recycling by macrophages account for the major iron fluxes in the body. Courtesy of Elsevier, license number 4223710421685 (147).

The body has no mechanism for active excretion of iron and the regulation of dietary iron absorption from the duodenum plays a critical role in human iron homeostasis (148). Ferroportin 1 (Fpn1, also known as Ireg1 and MTP 1) is the only known iron exporter identified to date (149). Ferroportin exports iron from enterocytes, macrophages and hepatocytes (148). The expression of ferroportin is regulated by the hormone hepcidin (150). Hepcidin binds to ferroportin causing its internalization and degradation, thereby inhibiting iron influx into plasma from duodenal enterocytes, macrophages and hepatocytes (151). The synthesis of hepcidin is stimulated by iron loading and inflammation, while iron deficiency decreases the hepcidin levels (152). Once transported across the basal membrane, iron is bound to plasma transferrin, and iron uptake from plasma is highly dependent on receptor-mediated endocytosis of transferrin bound to transferrin receptors (145). When increased iron uptake is needed (as in iron deficiency anemia), the expression of transferrin receptors are upregulated (146). Most of the body iron content is incorporated into hemoglobin in erythroid precursors or mature erythrocytes, and in myoglobin in muscles. The majority of remaining body iron is stored in the liver and in macrophages, mostly bound to ferritin proteins (153). Human ferritin is composed of two protein subunits, the H and L chain (154). Ferritin heavy-chain (FtH) takes up and releases iron rapidly, while the ferritin light-chain (FtL) is mainly an iron storage with less catalytic property (154). Moreover, inflammation upregulates FtH rather than FtL, as the catalytic properties of FtH are important for reducing iron availability (which is an important defense mechanism against pathogens) (153). The iron metabolism is continuously regulated in order to meet cellular iron demands, avoid oxidative damage and assist the immune system in its defense against pathogens. Hepcidin, ferritin and transferrin receptors seem to be key players in maintaining iron homeostasis (145, 147, 152, 153).

The association between body iron levels and VTE is scarcely investigated. Iron deficiency is the most common cause of anemia (155), and anemia has been associated with

cerebral venous thrombosis in several case reports and a case-control study (156, 157). Moreover, patients with PE had lower hemoglobin levels than age and sex-matched controls in a case-control study (158). Iron deficiency might cause thrombocytosis, and secondary thrombocytosis might explain an association between anemia and VTE (159). The abovementioned studies were conducted on small or highly selected study populations, and might not be relevant for the general population. Moreover, anemia might be caused by several conditions other than iron deficiency (160, 161), and iron deficiency does not always lead to anemia (162). In the third national health and nutrition examination survey, 9-11% of adolescent girls and women of childbearing age were iron deficient, whereas the prevalence of iron deficiency anemia was 2-5% in the same group (163). Thus, the association between iron metabolism and VTE remains to be investigated.

### 3. Aims of the thesis

- To investigate the association between RDW and risk of incident and recurrent VTE, and the association between RDW and risk of all-cause mortality among VTE-patients.
- To assess the impact of iron metabolism on risk of VTE and whether altered iron metabolism could explain the association between RDW and VTE
- To investigate the association between RDW and risk of incident cancer, cancer stage and mortality among cancer patients
- To assess whether the effect of RDW on VTE could be explained by intermediate development of MI, stroke or cancer.

## 4. Study Population and Methods

### 4.1 The Tromsø Study

The Tromsø Study is a population-based cohort study founded by the University of Tromsø in 1974. Its primary aim was to determine factors related to the high cardiovascular mortality observed in the northern part of Norway (164). Seven surveys have been conducted, the first in 1974 and the most recent in 2015-16, and several diseases and conditions have been included.

All four papers included in this thesis are based on the fourth survey of the Tromsø study (Tromsø 4) conducted in 1994-95. All inhabitants aged  $\geq 25$  years were invited to the first screening visit, and 27 158 participated (77% of the eligible population). All participants aged 55 to 74 years and 5-8% samples in the other 5-years birth cohorts were invited to a more extensive second screening visit. Of the 10 542 eligible subjects, 78% (7965) attended the second visit (164).

Papers I-III in this thesis are based on prospective follow-up studies on Tromsø 4 participants. For paper I, subjects were followed from the date of inclusion in 1994/95 through January 1, 2012. In paper II and III, study participants were followed from the date of inclusion through December 31, 2010. Paper IV is a nested case-control study with cases and controls selected among participants in Tromsø 4. The cases were drawn from participants diagnosed with a VTE event between 1994/95 and 2007. For each VTE-patient, two age-, sex- and index-date-matched controls were selected. The index-date was the time of the VTE event, meaning the controls had to be alive and without a VTE-diagnosis at the time of the VTE event in the corresponding case.

## 4.2 Baseline measurements (Tromsø 4)

Baseline information was collected by self-administered questionnaires, blood samples and a physical examination. Non-fasting blood samples were collected from an antecubital vein and analyzed at the department of Clinical Chemistry, University Hospital of North Norway. Serum and citrated plasma were prepared by centrifugation after one hour respite at room temperature, and frozen at  $-70^{\circ}\text{C}$ . For measurement of blood cell count, including RDW, 5 ml of blood was drawn into a vacutainer tube containing EDTA (K3- EDTA 40  $\mu\text{L}$ , 0.37mol/L per tube) as an anticoagulant, and analyzed within 12 hours in an automated blood cell counter (Coulter Counter®, Coulter Electronics, Luton, UK). RDW was calculated by dividing the standard deviation of the mean corpuscular volume (MCV) by MCV and multiplying by 100 to express the result as a percentage (133). For measurements of CRP, hepcidin, FtL and sTfR, plasma samples were thawed and analyzed by enzyme-linked immunosorbent assays. Height and weight were measured with participants wearing light clothes and no shoes, and BMI was calculated as weight in kilograms divided by the square of height in meters ( $\text{kg}/\text{m}^2$ ). Information on smoking habits, family history of cardiovascular diseases, hormone therapy and concurrent diseases was obtained from a self-administered questionnaire.

### 4.3 Outcome measures

The University Hospital of North Norway is the only hospital in the region, and all diagnostic radiology and hospital care is provided exclusively by this hospital. The hospital discharge diagnosis registry covers both hospitalizations and outpatient clinic visits. Moreover, the unique Norwegian national 11-digit identification number allows linkage to national and local diagnosis registries. The National Registry covers all subjects registered as inhabitants of Norway at the time of their death, without regard to whether the death took place in Norway or abroad.

#### *4.3.1 Identification and validation of venous thromboembolic events*

All VTE events during follow-up were identified by searching the hospital discharge diagnosis registry, the autopsy registry and the radiology procedure registry at the University Hospital of North Norway. The relevant discharge codes were International Classification of Diseases (ICD), revision 9 codes 325, 415.1, 451, 452, 453, 671.3, 671.4, 671.9 for the period 1994-98, and ICD-10 codes I26, I80, I81, I82, I67.6, O22.3, O22.5, O87.1, O87.3 for the period 1999-2012. Trained personell reviewed and validated each potential VTE case by assessment of the patient's medical record. For the potential cases derived from either the hospital discharge registry or the radiological procedure registry, all following four criteria were required for a VTE event to be recorded; 1) The presence of signs and symptoms accordant with either a DVT, PE, or both; 2) Objective confirmation by a diagnostic procedure (i.e. compression ultrasonography, venography, spiral computed tomography, perfusion-ventilation scan, pulmonary angiography or autopsy); 3) A diagnosis of a DVT or PE noted by a physician in the patient's medical records, and 4) Initiation of VTE treatment (i.e. anticoagulant medication, thrombolysis, vascular surgery) unless contraindications were specified. For the potential cases



derived from the autopsy registry, a VTE was only recorded when the autopsy report indicated VTE as the cause of death or as a significant condition associated with the cause of death.

VTE events were classified as DVT or PE, and if DVT and PE occurred simultaneously, it was recorded as a PE. Moreover, the VTE events were classified as provoked or unprovoked depending on the presence of provoking factors at the time of diagnosis. Provoking factors were recent surgery or trauma (within the previous 8 weeks), acute medical conditions (MI, ischemic stroke or major infectious disease), active cancer, immobilization (bed rest >3 days, wheelchair use or long-distance travel) or any other specific factors described by a physician in the medical record (for instance intravascular catheter).

Recurrent VTE was defined as symptomatic, objectively confirmed DVT or PE at (i) another location than the first VTE, or at (ii) the same location as the first VTE in cases where the recurrence occurred more than 7 days after the initial event.

#### *4.3.2 Identification and validation of cancer*

All cancer diagnoses in the Norwegian population are registered in the Cancer Registry of Norway and information about cancer in the cohort was obtained by linkage to the cancer registry using the unique 11-digit personal identification number. In a recent evaluation of the data quality, the Cancer Registry of Norway had a completeness of 98.8% with 94% of the cases being histologically verified (165). The cancer registry provided information on the date of diagnosis, location of disease (ICD-7 codes 140-205, except for non-melanoma skin cancer (ICD-7 code 191)) and cancer stage. Information on mortality was obtained by linkage to the National Causes of Death registry kept by Statistics Norway.

#### *4.3.3 Identification and validation of myocardial infarction*

All first-time events of MI were identified by linkage to the diagnosis registries at University Hospital of North Norway (outpatient diagnoses included) and the National Causes of Death Registry at Statistics Norway. Cases of possible incident nonfatal and fatal MI were identified by a broad search for the ICD-9 codes 410 to 414, 430 to 438, and 798 to 799 in the period 1994–1998 and thereafter for the ICD-10 codes I20 to I25, I60 to I69, and R96, R98, and R99. An independent endpoint committee validated all possible events of MI. The hospital medical records were retrieved for case validation. Information from the National Causes of Death Registry and from death certificates was used to collect relevant information of the event from additional sources such as autopsy reports and records from nursing homes, ambulance services, and general practitioners. Manual and/or electronic text searches in paper versions (used until 2001) and digital versions of hospital records were conducted for notes on MI in all participants with one or more of the abovementioned diagnosis (ICD-codes).

All incident events were classified as definite, probable, or possible MI, based on a classification algorithm that included clinical symptoms and signs, findings in electrocardiograms (ECG), values of cardiac biomarkers, and autopsy reports, when applicable. Definite MI was defined by one of the following sets of conditions; (1) Typical, atypical, or inadequately described symptoms plus a definite new infarction in ECG recordings; (2) Typical symptoms plus significantly higher myocardial enzymes and/or troponin levels; (3) Atypical or inadequately described symptoms plus significantly higher myocardial enzymes and/or troponin levels plus a probable new infarction in ECG recordings or (4) Postmortem evidence of recent MI or thrombosis.

#### *4.3.4 Identification and validation of stroke*

Stroke was defined according to the WHO definition as rapidly developing clinical signs of focal or global disturbance of cerebral function, with symptoms lasting 24 hours or longer or leading to death, and with no apparent cause other than vascular origin (166). First-ever non-fatal and fatal strokes were identified by a search for the ICD-9 codes 430-438 and ICD-10 codes I60-I69 in the diagnosis registries at the University Hospital of North Norway (diagnoses from outpatient clinics included), and the National Causes of Death Registry at Statistics Norway. An independent endpoint committee reviewed hospital medical records for validation of all possible hospitalized and out of hospital stroke events. Information from the National Causes of Death Registry and from death certificates was used to collect relevant information of the event from additional sources such as autopsy reports and records from nursing homes, ambulance services, and general practitioners. Event ascertainment followed a detailed protocol, according to established diagnostic criteria. Each case was reviewed separately. Moreover, manual and/or electronic text searches were performed in paper versions (used until 2001) and digital versions of hospital records for notes on stroke in all participants with a diagnosis of ICD-9 410-414, 798-799, and ICD-10 I20-I25, R96, R98, and R99.

## 5. Main results

### 5.1 Paper I

#### RED CELL DISTRIBUTION WIDTH IS ASSOCIATED WITH INCIDENT VENOUS THROMBOEMBOLISM (VTE) AND CASE-FATALITY AFTER VTE IN A GENERAL POPULATION

RDW is associated with cardiovascular event rate, cardiovascular mortality and all-cause mortality. Previously, two case-control studies and one cohort study suggested an association between RDW and incident VTE. Whether RDW is associated with recurrence or mortality in VTE patients had not been investigated. This prospective, population-based cohort study was conducted to investigate the impact of RDW on risk of incident and recurrent VTE, and case-fatality in a general population. RDW was measured in 26 223 subjects enrolled in the Tromsø Study in 1994-95. All verified incident and recurrent VTE events among the study participants were registered until January 1, 2012. Overall, study participants with  $RDW \geq 13.3\%$  (the upper population-based quartile) had a 50% higher risk of incident VTE than those with  $RDW < 12.5\%$  (lower quartile). The risk estimates were especially high for unprovoked deep vein thrombosis (HR highest versus lowest quartile of RDW: 1.9, 95% CI 1.2-3.2), supporting the idea that RDW may influence rheological properties of circulating erythrocytes. RDW was not associated with risk of VTE recurrence, but subjects with  $RDW \geq 13.3\%$  had 30% higher risk of all-cause mortality after the initial VTE-event compared to those with  $RDW < 13.3\%$ . In conclusion, high RDW is a risk factor of incident VTE, and a predictor of all-cause mortality in VTE patients.

## 5.2 Paper II

### PLASMA HEPcidIN IS ASSOCIATED WITH FUTURE RISK OF VENOUS THROMBOEMBOLISM

RDW is associated with numerous diseases, including VTE, but the underlying mechanism(s) are unsettled. Iron deficiency anemia is associated with high RDW, and studies suggest an association between iron deficiency and VTE. Thus, underlying iron deficiency may explain the association between RDW and VTE. In this nested case-control study, 390 VTE-patients and 802 age- and sex-matched controls were selected from the fourth survey of the Tromsø study. Iron status was assessed by measurement of hepcidin and ferritin-light-chain (FtL). RDW was inversely correlated with plasma levels of hepcidin, FtL and hemoglobin (Pearson correlation coefficient: -0.17, -0.11 and -0.17, respectively). The risk of VTE increased linearly across categories of higher plasma hepcidin levels. Participants with hepcidin in the highest quartile had an OR for VTE of 1.32 (95% CI 1.00-2.42) and those above the 90% percentile had an OR for VTE of 1.66 (95% CI 1.14-2.42) compared to the reference group (quartile 2 and 3). Similar results were found for FtL, albeit statistically not significant. The risk of VTE increased by categories of higher RDW, and was strengthened after inclusion of hepcidin and FtL in the multivariable model. In conclusion, our findings rejected the hypothesis that iron deficiency explained the association between RDW and VTE, and suggested, in contrast, that high body iron levels might increase the risk of VTE.

### 5.3 Paper III

#### IMPACT OF RED CELL DISTRIBUTION WIDTH ON FUTURE RISK OF CANCER AND ALL-CAUSE MORTALITY AMONG CANCER PATIENTS – THE TROMSØ STUDY

RDW has been associated with several malignancies, poor cancer prognosis and all-cause mortality in cancer patients and the general population. To the best of our knowledge, this prospective, population-based study was the first to investigate the association between RDW and future risk of cancer, cancer stage and mortality among cancer patients. RDW was measured in 25 383 participants in the Tromsø study, and all incident cancer diagnoses and deaths among the study participants were recorded until December 31, 2010. During a median follow-up time of 15.7 years, 1191 men and 1114 women were diagnosed with cancer. Men with  $RDW \geq 13.2\%$  (the upper quartile) had a 30% increased risk of cancer compared to men with  $RDW < 12.3\%$  (lower quartile) (HR 1.30, 95% CI 1.07-1.59). Furthermore, RDW was associated with increased risk of cancer in women of postmenopausal age ( $\geq 55$  years), whereas no association was found in women of premenopausal age ( $< 55$  years). In men, a 1% increase in RDW was associated with a 21% increased risk of regional cancer spread (HR 1.21, 95% CI 1.11–1.33) and a 19% increased risk of distal metastasis (HR 1.19, 95% CI 1.06–1.33). High RDW was associated with increased all-cause mortality in both male and female cancer patients. In conclusion, RDW was associated with future risk of incident cancer in men and women of postmenopausal age, as well as increased risk of all-cause mortality among cancer patients.

## 5.4 Paper IV

### THE ASSOCIATION BETWEEN RED CELL DISTRIBUTION WIDTH AND VENOUS THROMBOEMBOLISM IS NOT EXPLAINED BY MYOCARDIAL INFARCTION, STROKE OR CANCER

RDW is a risk marker of VTE, myocardial infarction (MI), stroke and cancer. As RDW is a numerical concept, it is likely to assume that the observed associations between RDW and various diseases are explained by underlying pathological mechanisms. However, the underlying explanation for the association between RDW and VTE remains unsettled. A relation between arterial thromboembolic diseases and risk of VTE has been reported in several studies. Moreover, cancer is associated with increased risk of VTE. Due to the established interrelations between these diseases, the apparent association between RDW and VTE could be explained by intermediate development of MI, stroke or cancer. In this prospective, population-based study, 24 363 participants from the Tromsø Study were followed for a median of 16 years. RDW was measured at baseline, and all incident events of VTE, MI, stroke and cancer among the study participants were registered. Conventional and cause-specific Cox-regression models were used to estimate hazard ratios for VTE according to RDW. In cause specific analysis, where each individual contributed with person-time until the first occurring event (i.e. first occurring event out of MI, stroke, cancer or VTE), the risk estimates were essentially similar to the results from the conventional Cox-regression analysis. These findings suggest that the association between RDW and future risk of VTE could not be explained by intermediate development of MI, stroke or cancer.

## 6. General discussion

### 6.1 Methodological considerations

#### 6.1.1 Study design

All papers in the thesis utilize a large population-based cohort study, the Tromsø Study. Paper I, III and IV have a traditional, prospective cohort study design, whereas paper II is a nested case-control study derived from the Tromsø Study. Cohort studies have several advantages compared to other study designs. Entire populations might participate, increasing the external generalizability and the chance of finding a valid association. In a cohort study, participants are followed from attendance date until the occurrence of an outcome (e.g. disease), until the participant cannot be followed (i.e. diseased or moved from the study area) or until the end of the study period. Thus, it is possible to assess and compare rates of outcome in participants with various traits (e.g. exposures) with a clear temporal sequence between exposure and outcome. However, cohort studies are not well suited for rare diseases unless the cohort is very large (167). VTE is a relatively common disease and suitable for cohort studies.

Cohort studies are non-experimental, but not without ethical concerns. Cohort studies are costly and resource demanding, and will, at least if publically funded, draw resources from other areas of the health care system. More importantly, cohort studies require healthy individuals to undergo various medical tests. Even the best tests generate false positive results, and the positive predictive value depends on the prevalence of disease in the test population. As most diseases are relatively rare in the general population, the positive predictive value of most tests are low in cohort studies. Thus, cohort studies generate a significant amount of false positive test results requiring resource demanding and potentially harmful confirmation tests, and both the economic and human cost/benefit-ratio require attention when designing cohort studies. All participants in the Tromsø study with deviating results (outside normal ranges) are



referred to their GPs or a relevant outpatient clinic at the University hospital of North Norway for further examinations. Thus, the Tromsø study generates substantial economic and human costs in addition to the actual survey, and invited participants must receive valid and thorough information on both benefits and risks before deciding to participate or not. However, the society considers the information and knowledge obtained from the study to outweigh the cost.

Large cohort studies increase the likelihood of statistically significant results (e.g. p-values below 0.05). However, statistical significance does not always equal clinical significance, and the researcher should formulate a clinical or biological meaningful hypothesis before performing the analysis. Case-control studies are more efficient and less resource demanding than cohorts, and better suited to investigate rare disorders. However, most case-control studies are retrospective and prone to reverse causation and recall bias, and cannot assess cause-and-effect relationships. For instance, a meta-analysis based on four case-control studies and one prospective study concluded that HDL-cholesterol protected against VTE (81). However, inflammation might decrease HDL-cholesterol levels (168), and low HDL could be an effect rather than a cause of VTE. In accordance with this, three independent cohort studies found no significant association between low HDL and risk of VTE (169-171). Furthermore, the selection of controls can be problematic in case-control studies. Ideally, controls should be sampled from the source population of cases, rather than from the entire non-diseased population (172). However, a nested case-control study design can draw advantages from “both worlds” (173). In paper II of this thesis, cases and controls were sampled from the background cohort and blood samples taken several years previously were thawed and analyzed. This enabled us to limit the use of expensive analyses without sacrificing the temporal advantage of cohort studies. Moreover, the cases and controls were sampled from the same population and matched on age, sex and index-date, as previously described.

Even though cohort studies can establish temporal associations, they cannot establish causality. Randomized controlled trials (RCT) randomly allocate individuals to an exposure (or intervention) before assessing the effect on outcome. In cohort studies, allocation of individuals is not by chance, and the observed difference in outcome might be attributed to factors other than the exposure (confounding). Moreover, cohort studies are not able to obtain experimental evidence for the detected association, one of the Bradford-Hill criteria of causality (174). Nevertheless, as RCTs require large resources and substantial ethical considerations, cohort studies are useful to identify associations for further investigation in RCTs.

#### *6.1.2 Information bias, misclassification and regression dilution bias*

All measurements and calculations will have some degree of error, and could possibly lead to information bias. Measurement errors occur when continuous variables are measured incorrectly, while misclassification errors might be present in categorical variables. Both errors might be random or systematic. If the error is systematic, we get information bias. For instance, if only one cuff size is used to measure blood pressure in a study and an inappropriate cuff size yield erroneous results, obese subjects would have more erroneous measurements than normal-weight subjects. Self-administered questionnaires may cause errors and possible bias. The questions might be difficult to answer correctly, either because the participants do not understand the question, do not know what to answer (e.g. do not remember how many units of alcohol they consume during a period), or study participants might answer incorrectly (e.g. underestimate their body weight). Alternatively, the questions might be interpreted differently than intended. Many of these errors are likely to be of a random nature, while some information is more likely to be systematically wrong. It is more likely to get an overestimation of body height and an underestimation of body weight than the opposite. In the Tromsø study, height

and weight were measured at the time of the physical examination. Paper I includes self-reported information on previous cardiovascular disease, smoking habits and hormone therapy. The validity of self-reported stroke has been assessed by Engstad et.al (175). They estimated the sensitivity of self-reported stroke in Tromsø4 to be 80% and the specificity to be 99%. They concluded that the self-administered questionnaire is suitable to assess the prevalence of stroke. The questions regarding smoking and use of hormone therapy are not validated.

Another type of measurement error arises from errors in the measurement process, either human error or error within measurement instruments. Measurement errors due to analytical variation are typically of a random nature as long as the instruments are calibrated regularly. The department of Clinical Chemistry at the University Hospital of North Norway report the analytical variation of RDW measurement to be less than 3%. Finally, punching the collected data into the dataset could be a source of errors.

Taken together, it is likely to assume some misclassification errors in our data. The effect of misclassification on the risk estimates depends on the type of misclassification error. If the misclassification is evenly distributed among study participants and independent of other variables, it is non-differential. In general, non-differential misclassification is thought to lead to bias towards the null, i.e. the observed effect is smaller than the true effect (176).. Even though the classification process yields errors non-differentially, the misclassification in the data might in fact be differential due to chance (177). Thus, even random measurement errors might cause biased estimates greater than the true estimates, and one cannot always claim that the observed effect must be an underestimation (178). Differential misclassification errors would inevitably lead to bias as the error is not evenly distributed among the study participants (179). One example of the latter is recall bias. Study participants with a disease are more likely to report the presence of a risk factor than healthy study participants are. The magnitude of the bias depends on the sensitivity and specificity of the measurement. For instance, if the sum of

the sensitivity and specificity for the self-administered questionnaire on smoking habits in the Tromsø Study were less than one, the direction of the observed association (for an outcome with smoking as risk factor) would be reversed from the true association. However, the abovementioned study by Engstad et.al indicate much better test properties for the self-administered questionnaire. Moreover, information from self-administered questionnaires were only used to control for possible confounders in paper I of this thesis and the effect of misclassification errors on the risk estimates should not be substantial. The endpoint-committees were blinded for the baseline variables, and any misclassification in outcome assessment should not be influenced by exposure status.

In all papers of this thesis, RDW was measured at baseline only, and could potentially change during the relatively long follow-up. Mean time from baseline to outcome was 9 years in paper I, III and IV and 7 years in paper II. For cohort studies with long time-lapse from exposure and outcome, modifiable risk factors might cause regression dilution bias (180). For instance, if a person has a low RDW-value at baseline, she would be placed in quartile 1 (e.g. the reference group). If the RDW-value subsequently increases, she would still be in quartile 1 in the study, but perhaps in quartile 3 or 4 in reality. Another person might have measured a high RDW at baseline that subsequently decreases. Thus, person 1 has a higher risk during follow-up than the baseline value indicates, whereas the opposite is the case for person 2. As a result, the risk estimates for VTE is attenuated (180). The regression dilution bias increases the risk of missing true positive associations (type II errors), but can be corrected using repeated measurements from a representative sample of the study population. Unfortunately, no repeated measurements were available for RDW in our study population.

### *6.1.3 Confounding and effect modification*

Confounding is a central issue for all epidemiologic study designs. A simple definition of confounding is “the confusion of effects”, implying that the effect of the exposure is mixed with the effect of another variable, leading to a bias (167). A confounder is a variable related to both the exposure and the outcome that is not directly in the causal pathway (181). In a cohort study, this means that the confounding variable must differ between the comparison groups and it must influence the outcome. Naturally, confounding is a potential problem for causality assessment, and it might, if not noticed, lead to wrong conclusions on associations. Confounding can lead to over- or underestimation of the true effect. It can also reverse the apparent direction of an effect. It differs from selection and information bias in the way that a confounded association is real (but not causal). Cohort studies are prone to confounding because the study design does not contain randomization (as RCTs do), but there are approaches for controlling the effect of confounding (181, 182). First, one must identify possible confounders, ideally by review of the available literature. The next step is to assess whether the distribution of potential confounders varies between study participants. After selecting available confounders, there are in general two possible actions to control confounding: stratification and multivariable models. Stratification of continuous variables is possible after categorization. The range in these categories must not be too wide in order to avoid residual confounding. However, narrow ranges would create small study populations for each analysis, which is a major limitation to stratification (183). Small sample sizes reduce the statistical power and yield uncertain risk estimates with wide confidence intervals. Moreover, as each subgroup is analyzed individually, the risk of false positive results (type 1 error) in one or more of the subgroups increases with the number of subgroups analyzed (184). If we test a null hypothesis that is in fact true, using 0.05 as the significance level, the probability of coming to a non-significant (correct) conclusion is 0.95. If the hypothesis is tested in two subgroups, the probability that

neither test will be significant is  $0.95^2 = 0.90$ . The probability of getting at least one significant result (i.e. a type 1 error) would be  $1-0.90=0.10$ . If the hypothesis is tested in many subgroups, the probability of conducting at least one type 1 error would in fact be greater than 0.5. The Bonferroni correction is a method used to control for the increased risk of type 1 errors (185). However, the Bonferroni correction has been criticized for yielding a high rate of false negative results (type II errors) due to its conservative design. Traditionally, scientific journals are reluctant to publish papers without statistical significant findings (publication bias), and the Bonferroni correction might enhance this problem. Thus, alternative methods have been developed (186, 187). However, the major concern regarding subgroup analysis and multiplicity is the lack of concern. Wang et al assessed the quantity and quality of subgroup analysis reported in a high-ranked journal during a one year period (188). Results from subgroup analysis were reported for 59 out of 97 trials. For 43 (72%) of these trials, it was unclear whether the subgroup analysis were predefined or post hoc, and interaction tests were completely reported in only 16 (27%) trials. Multiple comparison issues were addressed in only two of the 15 trials claiming heterogeneity of treatment effects. Stratification should be predefined to avoid post hoc data exploration, and statistical tests for interaction between subgroups should be evaluated (184, 189). Moreover, findings in subgroup analysis should be interpreted with caution and confirmed in independent studies before conclusions can be drawn (190).

In paper III, there was a significant interaction between RDW and sex on the risk of cancer. Therefore, we stratified the cohort on gender. Moreover, we stratified the female cohort into two age groups based on presumed menopausal status (55 years). The stratification by age was done with a biological rationale, as premenopausal women are prone to iron deficiency which might cause elevated RDW. Thus, the rationale for stratification is valid, but the risk

estimates must nevertheless be interpreted with caution. In paper IV, we performed the same stratifications as in paper III, for the same reasons.

#### *6.1.4 Multivariable regression models*

Multivariable regression modelling solves the problem of abundant subgroups and allows a more efficient way to control for several variables simultaneously (167). In multivariable regression, the regression coefficient for each variable is estimated by fitting the model to the data and adjusting for all other variables in the model (183). However, results from the regression model are susceptible to bias if the model is not a good fit to the data. For the Cox proportional hazard model, the underlying assumption is that the exposure is associated with a constant increase in relative risk of the outcome, i.e. an exponential dose-response curve between exposure and outcome (181). This assumption is also referred to as the proportionality assumption (183), and can be tested by adding Schoenfeld residuals in the model and applying the tests of the non-zero slope developed by Therneau and Grambsch (191). This method was applied in paper I, III and IV, and no violation was found.

The selection of covariates in multivariable models requires consideration. Adjusting for any given variable implies an adjustment, at least partially, for other variables related to it (181). Over-adjustment occurs if the adjusted variable is in the causal pathway between the exposure and outcome, or if the variable is so strongly associated with the exposure or outcome that their true relationship is altered (181). In order to avoid over-adjustment, it is important to know the biological mechanisms behind the observed association, and to assess the relationship between the confounding variable and the exposure and outcome. In addition, careless selection of covariates may give rise to a statistical phenomenon described as the vibration of effects. Patel et al assessed the variance of the effect size obtained by different combinations of

adjustment variables (192). Using data from the National Health and Nutrition Examination Survey (NHANES), they found that vitamin E was associated with both increased and decreased risk of all-cause mortality, depending on which covariates included in the model. One exposure cannot both increase and decrease the risk of the same outcome in the same study population simultaneously, so some of the obtained effect sizes are clearly misleading. This study emphasizes the importance of predetermination of covariates with valid biological or clinical significance when designing statistical models. As the underlying biological mechanism(s) for the association between RDW and VTE is largely unknown, it is likely that some of the covariates included in the papers I-IV are not actual confounders, and over-adjustment might have occurred to some extent. However, our findings are in accordance with several findings from other study populations.

Although several known confounders have been adjusted for, there is still a possibility of residual confounding. Residual confounding occurs when either the categories of the confounder controlled for are too broad or when some confounding variables are not measured or for other reasons remain unaccounted for (181). The latter is most likely for the papers in the present thesis, as the underlying mechanisms for the association between RDW and VTE are unknown. Underlying conditions such as inflammatory diseases and the use of various medication may influence RDW, and some of them are also likely to influence the risk of VTE. Unfortunately, we have limited baseline information on these conditions, and residual confounding cannot be ruled out.



### *6.1.5 Detection and validation of outcomes*

The outcome measurements are also prone to error. All outcomes in this thesis were identified by searching for specific diagnoses in different registers. It is likely to assume that some of the diagnoses were erroneously registered or coded. For instance, some DVT patients may have been registered with a superficial vein thrombosis or vice versa. In order to minimize the risk of missing outcomes due to coding errors, the initial search for diagnosis included several ICD-codes in addition to the “VTE-codes”, such as thrombophlebitis, unspecified venous complication and more. Secondly, trained personnel reviewed all possible cases. They applied, as previously described, strict and objective criteria to each possible case. This approach minimizes the risk of false positive cases in the study population. The University Hospital of North Norway is the only hospital in the region, providing all diagnostic radiology and hospital care. However, some cases may have been missed if they were diagnosed and treated elsewhere, or if the participant decided not to seek medical assistance. The personnel registering the outcomes were blinded to the baseline information, and any misclassification errors are likely to be non-differential.

### *6.1.6 Selection bias, missing variables and external validity*

Even though the participation rate is quite high in the Tromsø Study, there is still a significant part of the population not attending and non-response bias cannot be ruled out. In the fourth survey, 77% of the eligible population participated. It is a common perception that those not attending health surveys are from a lower socioeconomic class and thus likely to have a poorer health in general than those attending (193). Selection bias occurs when a systematic error in the recruitment of study participants results in a tendency toward distorting the measure expressing the association between exposure and outcome (181). More than 85% of those aged 45-74 participated in the study, but the remaining 15% might differ in disease burden.

Moreover, the attendance rate among the oldest and youngest was considerably lower. Among those younger than 30 years, the low attendance rate can reflect a proportion living outside Tromsø even though officially registered as inhabitants of Tromsø (i.e. for studying, working offshore etc.). There are, to the best of my knowledge, no obvious reasons why young people living outside of Tromsø should differ much from those remaining in the city. Langhammer et al. conducted a study on a similar cohort (the HUNT study) where they stated that the most important reason for non-participation in the 30-year-old subjects was lack of time or an inconvenient time for appointment (194). However, the low attendance rate for those above 80 years of age is more likely to cause some selection bias. Langhammer et al. found that the non-participation reason of feeling too ill to participate was of increasing importance with increasing age (194). This effect of the most ill persons missing out is called survivor bias (182). On the other hand, those who volunteer for health surveys might have considerations for their own health. These biases counteract each other, but the net bias is unknown as we do not have information regarding those not attending. However, we assessed whether the effect of RDW on risk of VTE varied among different age groups (Paper I). The risk estimates were similar for all three age groups, suggesting that selection bias is not a major problem for the results presented in this thesis. In paper II, which is a nested case-control study, cases and controls were selected from the same population, and they were matched on age, sex and index-date. These measures reduce the risk of selection bias.

Missing data is a problem occurring in almost all studies, due to several reasons. Study participants might not complete the questionnaire, data values are missing because of equipment failure, laboratory samples are lost in transit or technically unsatisfactory, study participants are lost from follow-up, or data is missing because of unknown reasons (195). The standard approach to the problem is to exclude individuals with missing data from the analysis. This practice is appropriate if the data is missing completely at random and the magnitude is

limited. However, it might yield problems with statistical power if many subjects are excluded, or it can yield biased results if those excluded differ from the remaining study population (196). Alternative approaches is to omit the variables with many missing values or to replace the missing values with a predicted value (imputation). However, no really satisfactory solution exists for missing data (195). The main concern is whether the missing data could introduce bias. In paper I, III and IV, around 2% of the eligible population was omitted due to missing data, and in paper II, 5 out of 1266 subjects were omitted due to missing hepcidin values. Thus, missing data is not likely to introduce bias in the papers of this thesis.

The external validity of a study is determined by several factors, most importantly the attendance rate and quality of the data. A study has high external validity if the results are applicable to the whole source population and similar populations, i.e. the study has high generalizability. The age and sex distribution of the Tromsø population is comparable with other Western populations (6, 197). The incidence and prevalence of cardiovascular diseases, traditional cardiovascular risk factors as well as the incidence of VTE and cancer is comparable with the incidence found in other studies of similar populations (6, 11, 197). The high attendance rate and quality of the data suggest that the study has high external validity, but the possible selection bias might, at least to a limited degree, reduce the generalizability for the older (above 80) age groups. Moreover, the results of our studies might not be representative to other ethnicities, as the vast majority of participants in the Tromsø study are of Caucasian ethnicity and RDW differs somewhat between ethnic groups (125-127),

The measurement and calculation of RDW is not standardized, which is a considerable limitation to the external validity of our results. As mentioned, RDW can be calculated via more than one method. Some studies, for instance studies from the Malmö Cancer and Diet cohort, calculate the RDW-SD in contrast to the more commonly used RDW-CV (198). In a study of 806 whole blood samples, Caporal and Comar calculated the ability of various measurements

of RDW to identify anisocytosis, using manual microscopic evaluation as the gold standard (122). For microcytic samples, RDW-CV outperformed RDW-SD in regards to sensitivity and negative predictive value, while RDW-SD had higher specificity and positive predictive value. For normocytic and macrocytic samples, RDW-SD had better overall performance than RDW-CV. Lippi and colleagues assessed and compared four hematological analyzers and found that RDW values varied among different analyzers, which limits the comparability between laboratories (199). The imprecision within the different hemocytometers was between 0.3-1.2 percent, indicating that the internal validity for RDW measurements is high as long as the same analyzer is used for the whole study population. Thus, the results obtained in our studies are valid, but no optimal cut-off values for RDW can be made until the measurement is standardized.

## 7. Discussion of main results

### 7.1 RDW as a biomarker for risk of VTE, cancer and mortality

Growing evidence suggest that RDW may have clinical applications for many disorders, including cardiovascular disease, heart failure and all-cause mortality (126, 135-140, 200, 201). Previously, two case-control studies (142, 143) and one cohort study (198) have reported an association between RDW and VTE. However, the effect of RDW on risk of all-cause mortality among VTE patients is not previously studied. RDW has been associated with presence of colon cancer (202), malign causes of biliary obstruction and weight loss (203, 204), as well as poor prognosis in patients with lung cancer and multiple myeloma (205, 206). Previously, the association between RDW and cancer has not been assessed in a prospective study.

In a large population-based case-control study of 2473 VTE patients and 2935 controls, Rezende et al found a strong and consistent association between RDW and risk of VTE (143). Individuals with RDW above the 95<sup>th</sup> percentile (RDW>14.1%) had a threefold higher risk of VTE compared to those with RDW values between the 5<sup>th</sup> and 95<sup>th</sup> percentiles (RDW range 11.7-14.1%). Moreover, patients with DVT (n=216) had higher RDW than controls (n=215) referred to duplex ultrasonography with suspicion of DVT (142). In a Swedish population-based cohort study, individuals with RDW in the upper quartile had a 1.7-fold higher risk of VTE than those in the lower quartile (198). Similarly, we found that individuals with RDW in the upper quartile had a 1.5-fold higher risk of incident VTE than those in the lowest quartile of RDW (paper I). The risk estimate was particularly high for unprovoked DVT, and RDW was associated with increased risk of mortality among VTE patients. RDW predicted long-term mortality independent of hemoglobin-levels in patients undergoing percutaneous coronary intervention (137), and in patients with myocardial infarction (207). In a study of participants 45 years or older in the NHANES, the risk of all-cause mortality increased by 22% for every

1% increase in RDW (126). Moreover, a meta-analysis of seven community-based studies showed a dose-response relationship between RDW and risk of death among elderly with and without age-associated disease (201). Our findings in paper I suggest that RDW also predicts mortality among VTE patients.

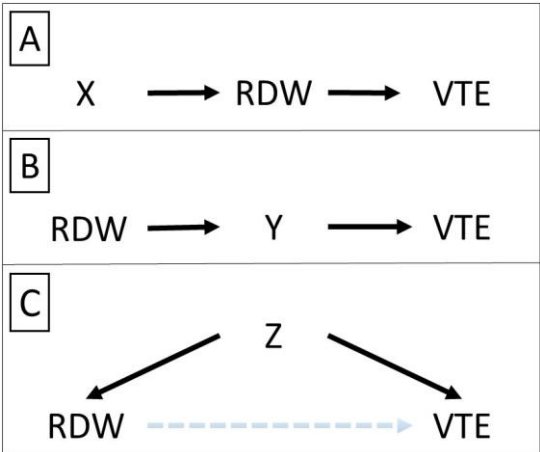
In a case-control study, RDW was higher in 225 patients with colon cancer compared to 494 cancer-free controls (202), and RDW was a useful tool to differentiate between benign and malign causes of biliary obstruction in a retrospective analysis (203). As RDW was measured at the time of diagnosis in these studies, it is possible that the observed elevation of RDW is a consequence of rather than a cause of cancer. As occult cancer may affect RDW through low-grade inflammation, we excluded 131 participants who were diagnosed with cancer within one year after the inclusion date, resulting in a mean time from inclusion to diagnosis of 9 years (median: 9 years) (paper III). Further, sensitivity analysis was performed by extending the exclusion interval from one to two years from study inclusion to cancer diagnosis (excluding additional 127 subjects) without affecting the risk estimates for cancer by RDW. Moreover, adding the time from baseline to cancer diagnosis as an extra adjustment variable did not alter the results. Thus, the findings in paper III demonstrate a clear temporal sequence between exposure (RDW) and outcome (incident cancer).

RDW was associated with advanced cancer stage and worsened prognosis in our study. Accordingly, elevated RDW was associated with more advanced cancers and worse prognosis in a study of 146 patients with multiple myeloma (206), and an association was found between elevated RDW, cancer stage and prognosis among lung cancer patients (205). Two previous cohort studies from NHANES found an association between RDW and cancer mortality (126, 140). The risk of cancer-related death was two-fold higher among participants in the highest compared to the lowest quintile (126), and the risk increased by 28% per 1-SD increase in RDW (140). In our study, the apparent association between RDW and mortality in cancer patients

was substantially weakened after adjustment for regional and distant metastasis at cancer diagnosis. These findings suggest that the relation between high RDW and cancer-related mortality is explained by advanced cancer stage at diagnosis.

### 7.2 Possible underlying mechanisms for the association between RDW and VTE

In principle, three theoretical models might explain the underlying mechanisms for the association between RDW and VTE (Figure 6). In model A, RDW is in the causal pathway similar to a mediator. In this case, the increased RDW is caused by factors not directly related to VTE, and the increased risk of VTE is a direct consequence of elevated RDW. Alternatively, the apparent association between RDW and VTE might be a consequence of intermediate development of another condition or disease (model B). Model C is essentially a confounding model as the association is explained by other underlying factors associated with both RDW and VTE. For instance, iron deficiency might cause elevated RDW. If iron deficiency also increases the risk of VTE, this would fit model C. Possible factors for each of the models are discussed in the following sections.



**Figure 6.** Theoretical models explaining the underlying mechanisms for the association between RDW and VTE. Panel A: RDW is increased due to factors (X) not related to VTE, and the increased risk of VTE is explained by direct effects from increased RDW. Panel B: High RDW increases the risk of developing a condition or disease (Y), and this condition increases the risk of VTE. Panel C: A factor (Z) causes both increased RDW and increased risk of VTE. RDW: Red cell distribution width, VTE: Venous thromboembolism.

### *7.2.1 Prothrombotic effects of RDW*

Some line of evidence support a direct effect of RDW on risk of VTE (Figure 6A). In paper I in this thesis, we hypothesized that hypercoagulability or erythrocyte aggregation may play a role in promoting thrombosis formation in individuals with high RDW. Increased RDW has been associated with decreased red blood cell deformability (208), and red cell deformability has been related to erythrocyte aggregation and altered blood viscosity (209). Hematocrit, one of the major determinants of blood viscosity, has been shown to predict future risk of venous thrombosis (210), and increased erythrocyte aggregation might promote thrombosis (211). Similar to RDW, inherited hypercoagulability caused by FV-Leiden favor development of DVT rather than PE (65, 212). However, there are no studies on the direct effect of RDW on blood viscosity. In paper IV, we found that RDW was not associated with VTE in women younger than 55 years, whereas older women had similar risk of VTE as men. If high RDW directly increases coagulability, one would expect the effect to be independent of sex. Moreover, RDW is associated with several diseases in which coagulation is not likely to play a role (213), and RDW is even shown to predict bleeding after percutaneous coronary intervention (214). Taken together, RDW is not likely to be a causal mediator of hypercoagulability.

### *7.2.2 Intermediate development of other diseases*

In addition to VTE and cancer (paper I and III), RDW has been associated with risk of MI, stroke and carotid atherosclerosis progression in the Tromsø Study (215-217). A relation between arterial thromboembolic diseases and risk of VTE was reported in several studies (218-221). Moreover, cancer is associated with a four- to sevenfold increased risk of VTE (97, 222-224). In paper IV, we hypothesized that the observed association between RDW and VTE could



be explained by intermediate development of cancer or arterial thromboembolic disorders (figure 6B). To test this hypothesis, we conducted conventional and cause-specific analyses of the relationship between RDW and VTE. For conventional analysis, person-time of follow-up was calculated from the date of enrolment to the date of an incident VTE diagnosis, to the date when the participant died or moved from the municipality of Tromsø, or to the end of the study period, whichever came first. For the cause-specific analysis, person time was calculated from the date of enrolment to the date of the first occurring diagnosis of VTE, MI, stroke or cancer, to the date when the participant died or moved from Tromsø, or to the end of the study period, whichever came first. The risk estimates for VTE were essentially similar in the conventional and cause-specific Cox-analyses, suggesting that the association between RDW and VTE is not explained by intermediate development of arterial thromboembolic conditions or cancer. To the best of our knowledge, no studies have investigated whether intermediate development of other diseases might contribute to the explanation.

### *7.2.3 Iron metabolism*

In 1981, Jerome Sullivan proposed the hypothesis that the greater incidence of heart disease in men and postmenopausal women compared with the incidence in premenopausal women results from higher levels of stored iron in these two groups (225). This “iron hypothesis” led to debate and controversy. Removal of stored iron has been shown to decrease the amount of iron deposition within atherosclerotic lesions in animals studies, yielding reduced lesion size and increased plaque stability (226, 227). In a study of high- and low-frequency voluntary blood donors, the high-frequency donors had evidence of decreased body iron stores assessed by serum ferritin and iron-binding capacity, decreased oxidative stress assessed by serum levels of 3-nitrotyrosine, and enhanced vascular function assessed by flow-mediated

dilution of the brachial artery (228). Moreover, iron chelation improved endothelial function, assessed by blood flow response to methacholine, in patients with coronary artery disease (229). RDW correlated with markers of iron deficiency and risk of mortality in two prospective studies of patients with heart failure (139, 230). However, high serum ferritin levels were associated with increased risk of acute MI among Finnish men (231), and high serum iron was associated with risk of fatal MI in a Canadian cohort study (232). In summary, both iron overload and iron deficiency seem to increase the risk of cardiovascular disease (233). High levels of iron might promote the formation of reactive oxygen species and the peroxidation of lipids, causing macrophages to develop into foam cells and promoting atherosclerosis (234). Iron deficiency is found to cause reactive thrombocytosis (159), and might lead to decreased antioxidant defense and increased platelet aggregation (235). Moreover, iron deficiency might be caused by underlying inflammation, as inflammatory cytokines such as IL-6 are known to cause increased hepcidin levels (147).

The role of iron in venous thromboembolic disease is scarcely studied. A history of iron deficiency anemia was associated with VTE in a Taiwanese case-control study on 2522 VTE patients and 12610 randomly selected controls (236). The odds ratio of previous iron deficiency anemia for VTE patients was 1.43 (95% CI: 1.10-1.87) compared to controls. In a cohort study from a similar database, patients (n=4001) with aplastic anemia had a 2.56-fold higher risk (95% CI: 1.81-3.63) of VTE than subjects (n=15998) without aplastic anemia (237). Potaczek et al conducted a cohort study of 229 patients with incident, unprovoked VTE to assess the effect of iron deficiency on risk of VTE recurrence (238). Patients with iron deficiency, defined as serum ferritin levels < 30 µg/L, had a three-fold higher risk of VTE recurrence than patients without iron deficiency. Finally, anemia was associated with cerebral venous thrombosis in the MEGA study (239). However, no correlation was found between anemia and PE in a case-

control study of 921 patients undergoing cross-sectional imaging (either CT or MR) to evaluate for PE (240).

RDW is traditionally used in the differential diagnosis of anemia (131), and is strongly associated with iron deficiency anemia (133, 134). Thus, we hypothesized that the association between RDW and VTE might be explained by underlying iron deficiency (paper II). We assessed whether altered iron metabolism, assessed by hepcidin and ferritin light-chain, was associated with risk of VTE. Our findings suggest that there is a dose-dependent association between increasing levels of body iron and risk of VTE. As RDW correlates with decreasing iron stores, it is highly unlikely that iron deficiency explains the association between RDW and VTE.

#### *7.2.4 Inflammation and oxidative stress*

RDW correlates with several inflammatory markers such as the erythrocyte sedimentation rate, high sensitivity CRP, tumor necrosis factor I and II, interleukin 6 and fibrinogen levels (139, 144, 241, 242). Oxidative stress and inflammation might increase RDW by impairing iron metabolism, reducing the lifespan of red blood cells and altering the response of erythropoietin on the bone marrow (243, 244). Low levels of selenium, an important antioxidant, was associated with high RDW and predicted elevated RDW over a two-year period in the Women's Health and Aging Study (245). However, the association between RDW and mortality was independent of CRP in NHANES (140), and the association between inflammatory markers and VTE has been inconsistent in previous observational studies (246, 247). To investigate whether inflammation can explain the association between RDW and VTE, we conducted a nested case-control study on 202 cases with incident VTE and 496 controls sampled from the Tromsø study (248). A combination of high CRP and high RDW had an

additive effect on VTE risk, but the combined risk estimate did not exceed the sum of the individual components. Moreover, adjusting for CRP did not alter the risk estimates for VTE by RDW. Our findings suggested that chronic inflammation could not completely explain the association between RDW and VTE.

#### *7.2.5 Other possible mechanisms*

We assessed whether the association between RDW and VTE was inflicted by joint effect of genetic thrombotic risk factors and RDW in 639 VTE patients and 1730 controls sampled from Tromsø IV (249). The risk of VTE by RDW did not change after adjustment for FV-Leiden (rs6025), F5 (rs4524), FGG (rs2066865), ABO blood type (rs8176719) and F11 (rs2036914 and rs2289252), and the combination of high RDW and risk alleles yielded no synergism. Our findings suggest that the risk of VTE by RDW is not caused by a joint effect of RDW and genetic thrombotic risk factors. Similarly, no interaction between high RDW and thrombophilia abnormality on the risk of VTE was observed in a case-control study of 730 VTE-patients and 352 controls (250).

During the last decade, RDW has been associated with a growing number of diseases and pathological processes (213). In addition to conditions previously discussed, RDW has been associated with metabolic disturbances, vitamin deficiencies, impaired liver and renal function, autoimmune diseases and malnutrition (251-253). RDW was associated with shorter telomere lengths in the Dallas Heart Study (254), and shortening of telomere length is associated with several age associated disorders and all-cause mortality (255, 256). Whether or not the abovementioned diseases and conditions can explain the association between RDW and VTE remain to be investigated.

## 8. Conclusions

- High RDW was associated with future risk of incident VTE and all-cause mortality among VTE-patients. The risk estimates were highest for unprovoked DVT. RDW was not associated with risk of VTE recurrence.
- High iron stores, assessed by hepcidin and ferritin light-chain, was associated with increased risk of VTE. As high RDW correlates with low iron stores, the association between RDW and VTE could not be explained by iron deficiency.
- RDW was associated with increased risk of cancer in men and women of postmenopausal age, whereas no association was found for younger women. Elevated RDW increased the risk of all-cause mortality for all cancer patients, most likely due to more advanced cancer stage among patients with high RDW.
- The effect of RDW on risk of VTE could not be explained by intermediate development of myocardial infarction, stroke or cancer.

## 9. Implications of results and future perspectives

Our findings of RDW as a risk factor and prognostic marker for VTE is consistent with findings from several other studies. The underlying mechanism(s) between these associations have proven difficult to explain, and the suggested models in this thesis do not offer an independent explanation. As VTE is a multicausal disease and numerous conditions influence RDW, it is likely that several factors contribute to the observed association. Even though the association seems robust and consistent, the effect size is not very large (HRs for VTE by upper versus lower quartile of RDW ranging from 1.3-2.0). Thus, detecting the individual effect from each potential factor (i.e. to extract a single effect from the background noise) is very difficult.

Regardless of the underlying mechanisms, RDW has proven to be a prognostic marker for several diseases, and it can predict risk of future disease. Moreover, RDW might identify patients with good prognosis or low risk of disease, i.e. patients without need of further investigations, prevention or prolonged treatment. Overdiagnosis and overtreatment is a major concern in modern medicine, with considerable economical and human costs. For VTE and particularly PE, the increasing use of CTPA has led to the diagnosis of small pulmonary emboli with questionable clinical significance (257). Wiener et al showed that while the PE incidence nearly doubled, the mortality rate remained stable after the advent of CTPA (32). These results suggest that nearly half of the patients diagnosed with PE receive potentially harmful treatment without a clear benefit, and the complication rate from anticoagulant treatment (ACT) substantially increased during the study period (32). Identifying patients without benefit from ACT might prevent unnecessary and harmful complications. The European Society of Cardiology guidelines recommend the use of prognostic scores such as the pulmonary embolism severity index to guide the therapeutic strategy for patients diagnosed with PE (30). However, the present prognostic scores are not accurate enough to identify patients without

benefit from ACT. Future studies should investigate whether RDW can improve assessment of prognosis, ultimately limiting the use of ACT to patients with a favorable benefit-to-risk ratio. Moreover, RDW might improve risk assessment models for selected or unselected patient groups, aiding clinicians in identifying patients in need of primary prophylaxis.

## 10. References

1. Yamaki T, Nozaki M, Sakurai H, Takeuchi M, Soejima K, Kono T. Presence of lower limb deep vein thrombosis and prognosis in patients with symptomatic pulmonary embolism: preliminary report. *European journal of vascular and endovascular surgery*. 2009;37(2):225-31.
2. Girard P, Sanchez O, Leroyer C, Musset D, Meyer G, Stern J-B, et al. Deep venous thrombosis in patients with acute pulmonary embolism: prevalence, risk factors, and clinical significance. *CHEST Journal*. 2005;128(3):1593-600.
3. van Langevelde K, Sramek A, Vincken PW, van Rooden JK, Rosendaal FR, Cannegieter SC. Finding the origin of pulmonary emboli with a total-body magnetic resonance direct thrombus imaging technique. *Haematologica*. 2013;98(2):309-15.
4. Enga KF, Rye-Holmboe I, Hald EM, Løchen ML, Mathiesen EB, Njølstad I, et al. Atrial fibrillation and future risk of venous thromboembolism: the Tromsø study. *Journal of Thrombosis and Haemostasis*. 2015;13(1):10-6.
5. Glynn RJ, Rosner B. Comparison of risk factors for the competing risks of coronary heart disease, stroke, and venous thromboembolism. *Am J Epidemiol*. 2005;162(10):975-82.
6. Cushman M, Tsai AW, White RH, Heckbert SR, Rosamond WD, Enright P, et al. Deep vein thrombosis and pulmonary embolism in two cohorts: the longitudinal investigation of thromboembolism etiology. *Am J Med*. 2004;117(1):19-25.
7. Silverstein MD, Heit JA, Mohr DN, Petterson TM, O'Fallon WM, Melton LJ, 3rd. Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study. *Arch Intern Med*. 1998;158(6):585-93.
8. Arshad N, Isaksen T, Hansen JB, Braekkan SK. Time trends in incidence rates of venous thromboembolism in a large cohort recruited from the general population. *Eur J Epidemiol*. 2017;32(4):299-305.
9. Heit JA, Silverstein MD, Mohr DN, Petterson TM, Lohse CM, O'Fallon WM, et al. The epidemiology of venous thromboembolism in the community. *Thromb Haemost*. 2001;86(1):452-63.
10. Heit JA. The epidemiology of venous thromboembolism in the community: implications for prevention and management. *Journal of thrombosis and thrombolysis*. 2006;21(1):23-9.
11. Heit JA. Venous thromboembolism: disease burden, outcomes and risk factors. *Journal of thrombosis and haemostasis*. 2005;3(8):1611-7.
12. Flinterman LE, van Hylckama Vlieg A, Cannegieter SC, Rosendaal FR. Long-term survival in a large cohort of patients with venous thrombosis: incidence and predictors. *PLoS Med*. 2012;9(1):e1001155.
13. Arshad N, Bjøri E, Hindberg K, Isaksen T, Hansen JB, Brækkan SK. Recurrence and mortality after first venous thromboembolism in a large population-based cohort. *Journal of Thrombosis and Haemostasis*. 2017;15(2):295-303.
14. Tagalakis V, Patenaude V, Kahn SR, Suissa S. Incidence of and mortality from venous thromboembolism in a real-world population: the Q-VTE Study Cohort. *The American journal of medicine*. 2013;126(9):832. e13-. e21.
15. Young L, Ockelford P, Milne D, Rolfe-Vyson V, McKelvie S, Harper P. Post-treatment residual thrombus increases the risk of recurrent deep vein thrombosis and mortality. *Journal of thrombosis and haemostasis*. 2006;4(9):1919-24.
16. Sorensen HT, Mellekjaer L, Olsen JH, Baron JA. Prognosis of cancers associated with venous thromboembolism. *The New England journal of medicine*. 2000;343(25):1846-50.
17. Heit JA, Mohr DN, Silverstein MD, Petterson TM, O'Fallon WM, Melton LJ. Predictors of recurrence after deep vein thrombosis and pulmonary embolism: a population-based cohort study. *Archives of internal medicine*. 2000;160(6):761-8.
18. Prandoni P, Lensing AA, Cogo A, et al. THE long-term clinical course of acute deep venous thrombosis. *Annals of Internal Medicine*. 1996;125(1):1-7.



19. Iorio A, Kearon C, Filippucci E, Marcucci M, Macura A, Pengo V, et al. Risk of recurrence after a first episode of symptomatic venous thromboembolism provoked by a transient risk factor: a systematic review. *Archives of internal medicine*. 2010;170(19):1710-6.
20. Baglin T, Luddington R, Brown K, Baglin C. Incidence of recurrent venous thromboembolism in relation to clinical and thrombophilic risk factors: prospective cohort study. *The Lancet*. 2003;362(9383):523-6.
21. Prandoni P, Lensing AW, Piccioli A, Bernardi E, Simioni P, Girolami B, et al. Recurrent venous thromboembolism and bleeding complications during anticoagulant treatment in patients with cancer and venous thrombosis. *Blood*. 2002;100(10):3484-8.
22. Heit JA. Predicting the Risk of Venous Thromboembolism Recurrence. *American journal of hematology*. 2012;87(Suppl 1):S63-S7.
23. Murin S, Romano PS, White RH. Comparison of outcomes after hospitalization for deep venous thrombosis or pulmonary embolism. *Thrombosis and haemostasis*. 2002;88(3):407-14.
24. Piazza G, Goldhaber SZ. Chronic thromboembolic pulmonary hypertension. *The New England journal of medicine*. 2011;364(4):351-60.
25. Prandoni P, Kahn SR. Post-thrombotic syndrome: prevalence, prognostication and need for progress. *Br J Haematol*. 2009;145(3):286-95.
26. Vazquez SR, Kahn SR. Advances in the diagnosis and management of postthrombotic syndrome. *Best practice & research Clinical haematology*. 2012;25(3):391-402.
27. Tick L, Kramer M, Rosendaal F, Faber W, Doggen CJM. Risk factors for post-thrombotic syndrome in patients with a first deep venous thrombosis. *Journal of Thrombosis and Haemostasis*. 2008;6(12):2075-81.
28. Brækkan SK, Grosse SD, Okoroh EM, Tsai J, Cannegieter SC, Næss IA, et al. Venous thromboembolism and subsequent permanent work-related disability. *Journal of Thrombosis and Haemostasis*. 2016;14(10):1978-87.
29. Mazzolai L, Aboyans V, Ageno W, Agnelli G, Alatri A, Bauersachs R, et al. Diagnosis and management of acute deep vein thrombosis: a joint consensus document from the European society of cardiology working groups of aorta and peripheral vascular diseases and pulmonary circulation and right ventricular function. *European Heart Journal*. 2017:ehx003-ehx.
30. Zamorano JL, Achenbach S, Baumgartner H, Bax JJ, Bueno H, Dean V, et al. 2014 ESC Guidelines on the diagnosis and management of acute pulmonary embolism The Task Force for the Diagnosis and Management of Acute Pulmonary Embolism of the European Society of Cardiology (ESC) Endorsed by the European Respiratory Society (ERS). *European Heart Journal*. 2014;35(43):3033-73.
31. Huang W, Goldberg RJ, Anderson FA, Kiefe CI, Spencer FA. Secular trends in occurrence of acute venous thromboembolism: the Worcester VTE study (1985-2009). *Am J Med*. 2014;127(9):829-39 e5.
32. Wiener RS, Schwartz LM, Woloshin S. Time trends in pulmonary embolism in the United States: evidence of overdiagnosis. *Arch Intern Med*. 2011;171(9):831-7.
33. Schissler AJ, Rozenshtein A, Kulon ME, Pearson GDN, Green RA, Stetson PD, et al. CT Pulmonary Angiography: Increasingly Diagnosing Less Severe Pulmonary Emboli. *PloS one*. 2013;8(6):e65669.
34. Wiener RS, Schwartz LM, Woloshin S. When a test is too good: how CT pulmonary angiograms find pulmonary emboli that do not need to be found. *British Medical Journal*. 2013;347.
35. Lippi G, Franchini M, Targher G. Arterial thrombus formation in cardiovascular disease. *Nat Rev Cardiol*. 2011;8(9):502-12.
36. Furie B, Furie BC. Mechanisms of Thrombus Formation. *New England Journal of Medicine*. 2008;359(9):938-49.
37. Mackman N. New insights into the mechanisms of venous thrombosis. *J Clin Invest*. 2012;122(7):2331-6.
38. Esmon CT. Basic mechanisms and pathogenesis of venous thrombosis. *Blood Rev*. 2009;23(5):225-9.

39. Busch C, Cancilla PA, DeBault LE, Goldsmith JC, Owen WG. Use of endothelium cultured on microcarriers as a model for the microcirculation. *Lab Invest.* 1982;47(5):498-504.
40. Esmon CT, Owen WG. Identification of an endothelial cell cofactor for thrombin-catalyzed activation of protein C. *Proc Natl Acad Sci U S A.* 1981;78(4):2249-52.
41. Esmon CT. The roles of protein C and thrombomodulin in the regulation of blood coagulation. *J Biol Chem.* 1989;264(9):4743-6.
42. Bovill EG, van der Vliet A. Venous valvular stasis-associated hypoxia and thrombosis: what is the link? *Annu Rev Physiol.* 2011;73:527-45.
43. Ogawa S, Gerlach H, Esposito C, Pasagian-Macaulay A, Brett J, Stern D. Hypoxia modulates the barrier and coagulant function of cultured bovine endothelium. Increased monolayer permeability and induction of procoagulant properties. *J Clin Invest.* 1990;85(4):1090-8.
44. Ogawa S, Shreeniwas R, Brett J, Clauss M, Furie M, Stern DM. The effect of hypoxia on capillary endothelial cell function: modulation of barrier and coagulant function. *Br J Haematol.* 1990;75(4):517-24.
45. Geddings JE, Mackman N. New players in haemostasis and thrombosis. *Thromb Haemost.* 2014;111(4):570-4.
46. Fuchs TA, Brill A, Duerschmied D, Schatzberg D, Monestier M, Myers DD, Jr., et al. Extracellular DNA traps promote thrombosis. *Proc Natl Acad Sci U S A.* 2010;107(36):15880-5.
47. Goldhaber SZ, Bounameaux H. Pulmonary embolism and deep vein thrombosis. *The Lancet.* 379(9828):1835-46.
48. Rosendaal FR. Venous thrombosis: a multicausal disease. *The Lancet.* 1999;353(9159):1167-73.
49. Souto JC, Almasy L, Borrell M, Blanco-Vaca F, Mateo J, Soria JM, et al. Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. *Genetic Analysis of Idiopathic Thrombophilia.* *Am J Hum Genet.* 2000;67(6):1452-9.
50. Heit JA, Phelps MA, Ward SA, Slusser JP, Petterson TM, De Andrade M. Familial segregation of venous thromboembolism. *Journal of thrombosis and haemostasis.* 2004;2(5):731-6.
51. Sorensen HT, Riis AH, Diaz LJ, Andersen EW, Baron JA, Andersen PK. Familial risk of venous thromboembolism: a nationwide cohort study. *Journal of thrombosis and haemostasis.* 2011;9(2):320-4.
52. Noboa S, Le Gal G, Lacut K, Mercier B, Leroyer C, Nowak E, et al. Family history as a risk factor for venous thromboembolism. *Thrombosis research.* 2008;122(5):624-9.
53. Bezemer ID, van der Meer FJ, Eikenboom JC, Rosendaal FR, Doggen CJ. The value of family history as a risk indicator for venous thrombosis. *Arch Intern Med.* 2009;169(6):610-5.
54. Zoller B, Li X, Sundquist J, Sundquist K. Age- and gender-specific familial risks for venous thromboembolism: a nationwide epidemiological study based on hospitalizations in Sweden. *Circulation.* 2011;124(9):1012-20.
55. Martinelli I, De Stefano V, Mannucci PM. Inherited risk factors for venous thromboembolism. *Nat Rev Cardiol.* 2014;11(3):140-56.
56. Tait RC, Walker ID, Perry DJ, Islam SI, Daly ME, McCall F, et al. Prevalence of antithrombin deficiency in the healthy population. *Br J Haematol.* 1994;87(1):106-12.
57. Tait RC, Walker ID, Reitsma PH, Islam SI, McCall F, Poort SR, et al. Prevalence of protein C deficiency in the healthy population. *Thromb Haemost.* 1995;73(1):87-93.
58. Rosendaal FR. Risk factors for venous thrombosis: prevalence, risk, and interaction. *Semin Hematol.* 1997;34(3):171-87.
59. Morange PE, Tregouet DA. Current knowledge on the genetics of incident venous thrombosis. *Journal of thrombosis and haemostasis.* 2013;11 Suppl 1:111-21.
60. Koster T, Rosendaal FR, de Ronde H, Briet E, Vandenbroucke JP, Bertina RM. Venous thrombosis due to poor anticoagulant response to activated protein C: Leiden Thrombophilia Study. *Lancet.* 1993;342(8886-8887):1503-6.
61. Simone B, De Stefano V, Leoncini E, Zacho J, Martinelli I, Emmerich J, et al. Risk of venous thromboembolism associated with single and combined effects of Factor V Leiden, Prothrombin

- 20210A and Methylenetetrahydrofolate reductase C677T: a meta-analysis involving over 11,000 cases and 21,000 controls. *Eur J Epidemiol.* 2013;28(8):621-47.
62. Seligsohn U, Lubetsky A. Genetic susceptibility to venous thrombosis. *The New England journal of medicine.* 2001;344(16):1222-31.
63. Rosendaal FR, Doggen CJ, Zivelin A, Arruda VR, Aiach M, Siscovick DS, et al. Geographic distribution of the 20210 G to A prothrombin variant. *Thromb Haemost.* 1998;79(4):706-8.
64. Dentali F, Sironi AP, Ageno W, Turato S, Bonfanti C, Frattini F, et al. Non-O blood type is the commonest genetic risk factor for VTE: results from a meta-analysis of the literature. *Seminars in thrombosis and hemostasis.* 2012;38(5):535-48.
65. Emmerich J, Rosendaal FR, Cattaneo M, Margaglione M, De Stefano V, Cumming T, et al. Combined effect of factor V Leiden and prothrombin 20210A on the risk of venous thromboembolism--pooled analysis of 8 case-control studies including 2310 cases and 3204 controls. Study Group for Pooled-Analysis in Venous Thromboembolism. *Thromb Haemost.* 2001;86(3):809-16.
66. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet.* 2005;6(2):95-108.
67. de Haan HG, Bezemer ID, Doggen CJM, Le Cessie S, Reitsma PH, Arellano AR, et al. Multiple SNP testing improves risk prediction of first venous thrombosis. *Blood.* 2012;120(3):656-63.
68. Folsom AR, Tang W, Weng LC, Roetker NS, Cushman M, Basu S, et al. Replication of a genetic risk score for venous thromboembolism in whites but not in African Americans. *Journal of Thrombosis and Haemostasis.* 2016;14(1):83-8.
69. Tsai AW, Cushman M, Rosamond WD, Heckbert SR, Polak JF, Folsom AR. Cardiovascular risk factors and venous thromboembolism incidence: the longitudinal investigation of thromboembolism etiology. *Arch Intern Med.* 2002;162(10):1182-9.
70. Braekkan SK, Mathiesen EB, Njolstad I, Wilsgaard T, Stormer J, Hansen JB. Family history of myocardial infarction is an independent risk factor for venous thromboembolism: the Tromso study. *Journal of thrombosis and haemostasis.* 2008;6(11):1851-7.
71. Naess IA, Christiansen SC, Romundstad P, Cannegieter SC, Rosendaal FR, Hammerstrom J. Incidence and mortality of venous thrombosis: a population-based study. *Journal of thrombosis and haemostasis.* 2007;5(4):692-9.
72. Engbers M, van Hylckama Vlieg A, Rosendaal F. Venous thrombosis in the elderly: incidence, risk factors and risk groups. *Journal of Thrombosis and Haemostasis.* 2010;8(10):2105-12.
73. Franchini M. Hemostasis and aging. *Critical reviews in oncology/hematology.* 2006;60(2):144-51.
74. Ageno W, Agnelli G, Imberti D, Moia M, Palareti G, Pistelli R, et al. Risk factors for venous thromboembolism in the elderly: results of the master registry. *Blood Coagulation & Fibrinolysis.* 2008;19(7):663-7.
75. Braekkan SK, Hald EM, Mathiesen EB, Njolstad I, Wilsgaard T, Rosendaal FR, et al. Competing risk of atherosclerotic risk factors for arterial and venous thrombosis in a general population: the Tromso study. *Arterioscler Thromb Vasc Biol.* 2012;32(2):487-91.
76. Heit JA, Kobbervig CE, James AH, Petterson TM, Bailey KR, Melton LJ, 3rd. Trends in the incidence of venous thromboembolism during pregnancy or postpartum: a 30-year population-based study. *Ann Intern Med.* 2005;143(10):697-706.
77. Hannaford PC. Epidemiology of the contraceptive pill and venous thromboembolism. *Thrombosis research.* 2011;127 Suppl 3:S30-4.
78. Rott H. Thrombotic risks of oral contraceptives. *Curr Opin Obstet Gynecol.* 2012;24(4):235-40.
79. Nelson HD, Humphrey LL, Nygren P, Teutsch SM, Allan JD. Postmenopausal hormone replacement therapy: scientific review. *JAMA.* 2002;288(7):872-81.
80. Obesity and Overweight factsheet from the WHO. 2017. available at <http://www.who.int/mediacentre/factsheets/fs311/en/>
81. Ageno W, Becattini C, Brighton T, Selby R, Kamphuisen PW. Cardiovascular risk factors and venous thromboembolism: a meta-analysis. *Circulation.* 2008;117(1):93-102.

82. Pomp ER, le Cessie S, Rosendaal FR, Doggen CJ. Risk of venous thrombosis: obesity and its joint effect with oral contraceptive use and prothrombotic mutations. *Br J Haematol*. 2007;139(2):289-96.
83. Horvei LD, Brækkan SK, Mathiesen EB, Njølstad I, Wilsgaard T, Hansen J-B. Obesity measures and risk of venous thromboembolism and myocardial infarction. *European Journal of Epidemiology*. 2014;29(11):821-30.
84. Horvei LD, Brækkan SK, Hansen J-B. Weight Change and Risk of Venous Thromboembolism: The Tromsø Study. *PloS one*. 2016;11(12):e0168878.
85. Klovaite J, Benn M, Nordestgaard B. Obesity as a causal risk factor for deep venous thrombosis: a Mendelian randomization study. *Journal of internal medicine*. 2015;277(5):573-84.
86. Klarin D, Emdin CA, Natarajan P, Conrad MF, Kathiresan S, Consortium I. Genetic Analysis of Venous Thromboembolism in UK Biobank Identifies the ZFPM2 Locus and Implicates Obesity as a Causal Risk Factor. *Circulation: Cardiovascular Genetics*. 2017;10(2):e001643.
87. Blokhin IO, Lentz SR. Mechanisms of thrombosis in obesity. *Curr Opin Hematol*. 2013;20(5):437-44.
88. Schafer K, Konstantinides S. Adipokines and thrombosis. *Clin Exp Pharmacol Physiol*. 2011;38(12):864-71.
89. Willenberg T, Schumacher A, Amann-Vesti B, Jacomella V, Thalhammer C, Diehm N, et al. Impact of obesity on venous hemodynamics of the lower limbs. *J Vasc Surg*. 2010;52(3):664-8.
90. Horvei LD, Grimnes G, Hindberg K, Mathiesen EB, Njølstad I, Wilsgaard T, et al. C-reactive protein, obesity, and the risk of arterial and venous thrombosis. *Journal of Thrombosis and Haemostasis*. 2016;14(8):1561-71.
91. Brækkan SK, Borch KH, Mathiesen EB, Njølstad I, Wilsgaard T, Hansen JB. Body height and risk of venous thromboembolism: The Tromsø Study. *Am J Epidemiol*. 2010;171(10):1109-15.
92. Borch KH, Nyegaard C, Hansen JB, Mathiesen EB, Njølstad I, Wilsgaard T, et al. Joint effects of obesity and body height on the risk of venous thromboembolism: the Tromsø Study. *Arterioscler Thromb Vasc Biol*. 2011;31(6):1439-44.
93. Bouillard JB BS. De l'Obliteration des veines et de son influence sur la formation des hydropisies partielles: consideration sur la hydropisies passive et general. *Arch Gen Med*. 1823;1(1):188-204.
94. Trousseau A, Bazire PV, Cornmack JL. *Lectures on Clinical Medicine* London: Hardwicke 1867.
95. Varki A. Trousseau's syndrome: multiple definitions and multiple mechanisms. *Blood*. 2007;110(6):1723-9.
96. Blom JW, Doggen CJ, Osanto S, Rosendaal FR. Malignancies, prothrombotic mutations, and the risk of venous thrombosis. *JAMA*. 2005;293(6):715-22.
97. Cronin-Fenton DP, Sondergaard F, Pedersen LA, Fryzek JP, Cetin K, Acquavella J, et al. Hospitalisation for venous thromboembolism in cancer patients and the general population: a population-based cohort study in Denmark, 1997-2006. *British journal of cancer*. 2010;103(7):947-53.
98. Piazza G. Venous thromboembolism and cancer. *Circulation*. 2013;128(24):2614-8.
99. Donati MB, Lorenzet R. Thrombosis and cancer: 40 years of research. *Thrombosis research*. 2012;129(3):348-52.
100. Falanga A, Rickles FR. Pathophysiology of the thrombophilic state in the cancer patient. *Seminars in thrombosis and hemostasis*. 1999;25(2):173-82.
101. Khorana AA, Kuderer NM, Culakova E, Lyman GH, Francis CW. Development and validation of a predictive model for chemotherapy-associated thrombosis. *Blood*. 2008;111(10):4902-7.
102. Angelini DE, Greene MT, Wietzke JN, Flanders SA, Sood SL. A Novel Risk Assessment Model to Predict Venous Thromboembolism (VTE) in Cancer Inpatients: The Canclot Score. *Am Soc Hematology*; 2016.
103. van Es N, Di Nisio M, Cesarman G, Kleinjan A, Otten H-M, Mahé I, et al. Comparison of risk prediction scores for venous thromboembolism in cancer patients: a prospective cohort study. *Haematologica*. 2017;haematol. 2017.169060.

104. Heit JA, Melton LJ, 3rd, Lohse CM, Petterson TM, Silverstein MD, Mohr DN, et al. Incidence of venous thromboembolism in hospitalized patients vs community residents. *Mayo Clin Proc.* 2001;76(11):1102-10.
105. Heit JA, O'Fallon WM, Petterson TM, Lohse CM, Silverstein MD, Mohr DN, et al. Relative impact of risk factors for deep vein thrombosis and pulmonary embolism: a population-based study. *Arch Intern Med.* 2002;162(11):1245-8.
106. Noboa S, Mottier D, Oger E. Estimation of a potentially preventable fraction of venous thromboembolism: a community-based prospective study. *Journal of Thrombosis and Haemostasis.* 2006;4(12):2720-2.
107. Anderson FA, Spencer FA. Risk Factors for Venous Thromboembolism. *Circulation.* 2003;107(23 suppl 1):I-9-I-16.
108. Spencer FA, Lessard D, Emery C, Reed G, Goldberg RJ. Venous thromboembolism in the outpatient setting. *Arch Intern Med.* 2007;167(14):1471-5.
109. Francis CW. Clinical practice. Prophylaxis for thromboembolism in hospitalized medical patients. *The New England journal of medicine.* 2007;356(14):1438-44.
110. Geerts WH, Code KI, Jay RM, Chen E, Szalai JP. A prospective study of venous thromboembolism after major trauma. *The New England journal of medicine.* 1994;331(24):1601-6.
111. Meissner MH, Chandler WL, Elliott JS. Venous thromboembolism in trauma: a local manifestation of systemic hypercoagulability? *J Trauma.* 2003;54(2):224-31.
112. Greene MT, Spyropoulos AC, Chopra V, Grant PJ, Kaatz S, Bernstein SJ, et al. Validation of Risk Assessment Models of Venous Thromboembolism in Hospitalized Medical Patients. *The American Journal of Medicine.* 2016;129(9):1001.e9-.e18.
113. Huang W, Anderson FA, Spencer FA, Gallus A, Goldberg RJ. Risk-assessment models for predicting venous thromboembolism among hospitalized non-surgical patients: a systematic review. *Journal of thrombosis and thrombolysis.* 2013;35(1):67-80.
114. Barbar S, Noventa F, Rossetto V, Ferrari A, Brandolin B, Perlati M, et al. A risk assessment model for the identification of hospitalized medical patients at risk for venous thromboembolism: the Padua Prediction Score. *Journal of Thrombosis and Haemostasis.* 2010;8(11):2450-7.
115. Nendaz M, Spirk D, Kucher N, Aujesky D, Hayoz D, Beer JH, et al. Multicentre validation of the Geneva Risk Score for hospitalised medical patients at risk of venous thromboembolism. *Thrombosis and haemostasis.* 2014;112(03):531-8.
116. Pottier P, Hardouin JB, Lejeune S, Jolliet P, Gillet B, Planchon B. Immobilization and the risk of venous thromboembolism. A meta-analysis on epidemiological studies. *Thrombosis research.* 2009;124(4):468-76.
117. Chandra D, Parisini E, Mozaffarian D. Meta-analysis: travel and risk for venous thromboembolism. *Ann Intern Med.* 2009;151(3):180-90.
118. Wu O, Robertson L, Langhorne P, Twaddle S, Lowe GDO, Clark P, et al. Oral contraceptives, hormone replacement therapy, thrombophilias and risk of venous thromboembolism: a systematic review The Thrombosis: Risk and Economic Assessment of Thrombophilia Screening (TREATS) Study. *Thrombosis and Haemostasis.* 2005;94(1):17-25.
119. Abdollahi M, Cushman M, Rosendaal FR. Obesity: risk of venous thrombosis and the interaction with coagulation factor levels and oral contraceptive use. *Thrombosis and Haemostasis.* 2003;89(3):493-8.
120. Severinsen MT, Overvad K, Johnsen SP, Dethlefsen C, Madsen PH, Tjønneland A, et al. Genetic susceptibility, smoking, obesity and risk of venous thromboembolism. *British journal of haematology.* 2010;149(2):273-9.
121. Price-Jones C. The diameters of red cells in pernicious anæmia and in anæmia following hæmorrhage. *The Journal of Pathology and Bacteriology.* 1922;25(4):487-504.
122. Caporal FA, Comar SR. Evaluation of RDW-CV, RDW-SD, and MATH-1SD for the detection of erythrocyte anisocytosis observed by optical microscopy. *Jornal Brasileiro de Patologia e Medicina Laboratorial.* 2013;49:324-31.

123. Chen PC, Sung FC, Chien KL, Hsu HC, Su TC, Lee YT. Red blood cell distribution width and risk of cardiovascular events and mortality in a community cohort in Taiwan. *Am J Epidemiol*. 2010;171(2):214-20.
124. Borne Y, Smith JG, Melander O, Engstrom G. Red cell distribution width in relation to incidence of coronary events and case fatality rates: a population-based cohort study. *Heart*. 2014;100(14):1119-24.
125. Zalawadiya SK, Veeranna V, Panaich SS, Afonso L, Ghali JK. Gender and ethnic differences in red cell distribution width and its association with mortality among low risk healthy United state adults. *The American journal of cardiology*. 2012;109(11):1664-70.
126. Patel KV, Ferrucci L, Ershler WB, Longo DL, Guralnik JM. Red blood cell distribution width and the risk of death in middle-aged and older adults. *Arch Intern Med*. 2009;169(5):515-23.
127. Saxena S, Wong ET. Heterogeneity of common hematologic parameters among racial, ethnic, and gender subgroups. *Archives of pathology & laboratory medicine*. 1990;114(7):715-9.
128. Shehata HA, Ali MM, Evans-Jones JC, Upton GJ, Manyonda IT. Red cell distribution width (RDW) changes in pregnancy. *Int J Gynaecol Obstet*. 1998;62(1):43-6.
129. Lurie S. Changes in age distribution of erythrocytes during pregnancy: a longitudinal study. *Gynecol Obstet Invest*. 1993;36(3):141-4.
130. Maes M, Scharpe S, Cooreman W, Wauters A, Neels H, Verkerd R, et al. Components of biological, including seasonal, variation in hematological measurements and plasma fibrinogen concentrations in normal humans. *Experientia*. 1995;51(2):141-9.
131. Bessman JD, Gilmer PR, Jr., Gardner FH. Improved classification of anemias by MCV and RDW. *American journal of clinical pathology*. 1983;80(3):322-6.
132. Rodak B. *Hematology: Clinical Principles and Applications*: Saunders Elsevier; 2007.
133. Evans TC, Jehle D. The red blood cell distribution width. *J Emerg Med*. 1991;9 Suppl 1:71-4.
134. Roberts GT, El Badawi SB. Red blood cell distribution width index in some hematologic diseases. *American journal of clinical pathology*. 1985;83(2):222-6.
135. Felker GM, Allen LA, Pocock SJ, Shaw LK, McMurray JJ, Pfeffer MA, et al. Red cell distribution width as a novel prognostic marker in heart failure: data from the CHARM Program and the Duke Databank. *J Am Coll Cardiol*. 2007;50(1):40-7.
136. Tonelli M, Sacks F, Arnold M, Moye L, Davis B, Pfeffer M, et al. Relation Between Red Blood Cell Distribution Width and Cardiovascular Event Rate in People With Coronary Disease. *Circulation*. 2008;117(2):163-8.
137. Poludasu S, Marmur JD, Weedon J, Khan W, Cavusoglu E. Red cell distribution width (RDW) as a predictor of long-term mortality in patients undergoing percutaneous coronary intervention. *Thromb Haemost*. 2009;102(3):581-7.
138. Kim J, Kim YD, Song TJ, Park JH, Lee HS, Nam CM, et al. Red blood cell distribution width is associated with poor clinical outcome in acute cerebral infarction. *Thromb Haemost*. 2012;108(2):349-56.
139. Forhecz Z, Gombos T, Borgulya G, Pozsonyi Z, Prohaszka Z, Janoskuti L. Red cell distribution width in heart failure: prediction of clinical events and relationship with markers of ineffective erythropoiesis, inflammation, renal function, and nutritional state. *Am Heart J*. 2009;158(4):659-66.
140. Perlstein TS, Weuve J, Pfeffer MA, Beckman JA. Red blood cell distribution width and mortality risk in a community-based prospective cohort. *Arch Intern Med*. 2009;169(6):588-94.
141. Zalawadiya SK, Veeranna V, Niraj A, Pradhan J, Afonso L. Red cell distribution width and risk of coronary heart disease events. *The American journal of cardiology*. 2010;106(7):988-93.
142. Cay N, Unal O, Kartal MG, Ozdemir M, Tola M. Increased level of red blood cell distribution width is associated with deep venous thrombosis. *Blood Coagul Fibrinolysis*. 2013;24(7):727-31.
143. Rezende SM, Lijfering WM, Rosendaal FR, Cannegieter S. Hematological variables and venous thrombosis: red cell distribution width and blood monocytes are associated with an increased risk. *Haematologica*. 2013.

144. Lippi G, Targher G, Montagnana M, Salvagno GL, Zoppini G, Guidi GC. Relation between red blood cell distribution width and inflammatory biomarkers in a large cohort of unselected outpatients. *Archives of pathology & laboratory medicine*. 2009;133(4):628-32.
145. Andrews NC. Disorders of Iron Metabolism. *New England Journal of Medicine*. 1999;341(26):1986-95.
146. Muñoz M, Villar I, García-Erce JA. An update on iron physiology. *World Journal of Gastroenterology : WJG*. 2009;15(37):4617-26.
147. Muckenthaler MU, Rivella S, Hentze MW, Galy B. A red carpet for iron metabolism. *Cell*. 2017;168(3):344-61.
148. Siah CW, Ombiga J, Adams LA, Trinder D, Olynyk JK. Normal Iron Metabolism and the Pathophysiology of Iron Overload Disorders. *Clinical Biochemist Reviews*. 2006;27(1):5-16.
149. Abboud S, Haile DJ. A Novel Mammalian Iron-regulated Protein Involved in Intracellular Iron Metabolism. *Journal of Biological Chemistry*. 2000;275(26):19906-12.
150. Ganz T. Heparin—a regulator of intestinal iron absorption and iron recycling by macrophages. *Best Practice & Research Clinical Haematology*. 2005;18(2):171-82.
151. Nemeth E, Ganz T. Heparin and iron-loading anemias. *Haematologica*. 2006;91(6):727-32.
152. Ganz T. Heparin, a key regulator of iron metabolism and mediator of anemia of inflammation. *Blood*. 2003;102(3):783-8.
153. Arosio P, Levi S. Ferritin, iron homeostasis, and oxidative damage. *Free Radical Biology and Medicine*. 2002;33(4):457-63.
154. Boyd D, Vecoli C, Belcher DM, Jain SK, Drysdale JW. Structural and functional relationships of human ferritin H and L chains deduced from cDNA clones. *Journal of Biological Chemistry*. 1985;260(21):11755-61.
155. Pasricha S-R. Anemia: a comprehensive global estimate. *Blood*. 2014;123(5):611-2.
156. Keung Y-K, Owen J. Iron Deficiency and Thrombosis: Literature Review. *Clinical and Applied Thrombosis/Hemostasis*. 2004;10(4):387-91.
157. Stolz E, Valdeuza JM, Grebe M, Schlachetzki F, Schmitt E, Madlener K, et al. Anemia as a risk factor for cerebral venous thrombosis? An old hypothesis revisited. *Journal of Neurology*. 2007;254(6):729.
158. Can Ç, Topacoglu H, Uçku R. Investigation of relationship between blood hemoglobin level and acute pulmonary embolism in emergency setting. *International Medical Journal*. 2013; 20(5): 584-6.
159. Dan K. Thrombocytosis in iron deficiency anemia. *Intern Med*. 2005;44(10):1025-6.
160. Bach V, Schruckmayer G, Sam I, Kemmler G, Stauder R. Prevalence and possible causes of anemia in the elderly: a cross-sectional analysis of a large European university hospital cohort. *Clinical interventions in aging*. 2014;9:1187.
161. Hoffbrand AV, Moss PA. *Essential haematology*: John Wiley & Sons; 2011.
162. Lopez A, Cacoub P, Macdougall IC, Peyrin-Biroulet L. Iron deficiency anaemia. *The Lancet*. 2016;387(10021):907-16.
163. Looker AC, Dallman PR, Carroll MD, Gunter EW, Johnson CL. Prevalence of iron deficiency in the United States. *JAMA*. 1997;277(12):973-6.
164. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njolstad I. Cohort profile: the Tromso Study. *Int J Epidemiol*. 2012;41(4):961-7.
165. Larsen IK, Småstuen M, Johannesen TB, Langmark F, Parkin DM, Bray F, et al. Data quality at the Cancer Registry of Norway: An overview of comparability, completeness, validity and timeliness. *European journal of cancer*. 2009;45(7):1218-31.
166. Investigators WMPP. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. *Journal of clinical epidemiology*. 1988;41(2):105-14.
167. Rothman KJ. *Epidemiology : an introduction*. 2nd ed. ed. Oxford: Oxford University Press; 2012.

168. Mainous AG, 3rd, Wells BJ, Everett CJ, Gill JM, King DE. Association of ferritin and lipids with C-reactive protein. *The American journal of cardiology*. 2004;93(5):559-62.
169. Braekkan SK, Borch KH, Mathiesen EB, Njolstad I, Hansen JB. HDL-cholesterol and future risk of venous thromboembolism: the Tromso Study. *Journal of thrombosis and haemostasis*. 2009;7(8):1428-30.
170. Chamberlain AM, Folsom AR, Heckbert SR, Rosamond WD, Cushman M. High-density lipoprotein cholesterol and venous thromboembolism in the Longitudinal Investigation of Thromboembolism Etiology (LITE). *Blood*. 2008;112(7):2675-80.
171. Everett BM, Glynn RJ, Buring JE, Ridker PM. Lipid biomarkers, hormone therapy and the risk of venous thromboembolism in women. *Journal of thrombosis and haemostasis*. 2009;7(4):588-96.
172. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. ed: Wolters Kluwer Health; 2015.
173. Ernster VL. Nested Case-Control Studies. *Preventive Medicine*. 1994;23(5):587-90.
174. Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med*. 1965;58:295-300.
175. Engstad T, Bonna KH, Viitanen M. Validity of self-reported stroke : The Tromso Study. *Stroke*. 2000;31(7):1602-7.
176. Gullen WH, Bearman JE, Johnson EA. Effects of misclassification in epidemiologic studies. *Public Health Reports*. 1968;83(11):914-8.
177. Wacholder S, Hartge P, Lubin JH, Dosemeci M. Non-differential misclassification and bias towards the null: a clarification. *Occupational and environmental medicine*. 1995;52(8):557.
178. Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. *International journal of epidemiology*. 2005;34(3):680-7.
179. Hofler M. The effect of misclassification on the estimation of association: a review. *International journal of methods in psychiatric research*. 2005;14(2):92-101.
180. Clarke R, Shipley M, Lewington S, Youngman L, Collins R, Marmot M, et al. Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. *American journal of epidemiology*. 1999;150(4):341-53.
181. Szklo M, Nieto FJ. *Epidemiology: beyond the basics*: Jones & Bartlett Publishers; 2012.
182. Bhopal RS. *Concepts of epidemiology: an integrated introduction to the ideas, theories, principles and methods of epidemiology*: Oxford University Press Oxford; 2002.
183. Katz MH. *Multivariable analysis: a primer for readers of medical research*. *Ann Intern Med*. 2003;138(8):644-50.
184. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in medicine*. 2002;21(19):2917-30.
185. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ*. 1995;310(6973):170.
186. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society Series B (Methodological)*. 1995:289-300.
187. Garamszegi LZ. Comparing effect sizes across variables: generalization without the need for Bonferroni correction. *Behavioral Ecology*. 2006;17(4):682-7.
188. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. *Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials*. *New England Journal of Medicine*. 2007;357(21):2189-94.
189. Lagakos SW. *The Challenge of Subgroup Analyses — Reporting without Distorting*. *New England Journal of Medicine*. 2006;354(16):1667-9.
190. Gyawali B, Prasad V. Pemetrexed in nonsquamous non-small-cell lung cancer: The billion dollar subgroup analysis. *JAMA Oncology*. 2017.
191. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81(3):515-26.



192. Patel CJ, Burford B, Ioannidis JP. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *Journal of clinical epidemiology*. 2015;68(9):1046-58.
193. Sogaard AJ, Selmer R, Bjertness E, Thelle D. The Oslo Health Study: The impact of self-selection in a large, population-based survey. *Int J Equity Health*. 2004;3(1):3.
194. Langhammer A, Krokstad S, Romundstad P, Heggland J, Holmen J. The HUNT study: participation is associated with survival and depends on socioeconomic status, diseases and symptoms. *BMC Med Res Methodol*. 2012;12:143.
195. Altman DG, Bland JM. Missing data. *BMJ*. 2007;334(7590):424-.
196. Clark TG, Altman DG. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. *Journal of clinical epidemiology*. 2003;56(1):28-37.
197. Jensvoll H, Severinsen MT, Hammerstrom J, Braekkan SK, Kristensen SR, Cannegieter SC, et al. Existing data sources in clinical epidemiology: the Scandinavian Thrombosis and Cancer Cohort. *Clinical epidemiology*. 2015;7:401-10.
198. Zoller B, Melander O, Svensson P, Engstrom G. Red cell distribution width and risk for venous thromboembolism: A population-based cohort study. *Thrombosis research*. 2013.
199. Lippi G, Pavesi F, Bardi M, Pipitone S. Lack of harmonization of red blood cell distribution width (RDW). Evaluation of four hematological analyzers. *Clinical Biochemistry*. 2014;47(12):1100-3.
200. Aronson D, Musallam A, Lessick J, Dabbah S, Carasso S, Hammerman H, et al. Impact of diastolic dysfunction on the development of heart failure in diabetic patients after acute myocardial infarction. *Circ Heart Fail*. 2010;3(1):125-31.
201. Patel KV, Semba RD, Ferrucci L, Newman AB, Fried LP, Wallace RB, et al. Red cell distribution width and mortality in older adults: a meta-analysis. *J Gerontol A Biol Sci Med Sci*. 2010;65(3):258-65.
202. Spell DW, Jones DV, Jr., Harper WF, David Bessman J. The value of a complete blood count in predicting cancer of the colon. *Cancer Detect Prev*. 2004;28(1):37-42.
203. Beyazit Y, Kekilli M, Ibis M, Kurt M, Sayilir A, Onal IK, et al. Can red cell distribution width help to discriminate benign from malignant biliary obstruction? A retrospective single center analysis. *Hepatogastroenterology*. 2012;59(117):1469-73.
204. Baicus C, Caraiola S, Rimbas M, Patrascu R, Baicus A, for Grupul de Studiu al Scaderii Ponderale I. Utility of routine hematological and inflammation parameters for the diagnosis of cancer in involuntary weight loss. *J Investig Med*. 2011;59(6):951-5.
205. Koma Y, Onishi A, Matsuoka H, Oda N, Yokota N, Matsumoto Y, et al. Increased red blood cell distribution width associates with cancer stage and prognosis in patients with lung cancer. *PloS one*. 2013;8(11):e80240.
206. Lee H, Kong SY, Sohn JY, Shim H, Youn HS, Lee S, et al. Elevated red blood cell distribution width as a simple prognostic factor in patients with symptomatic multiple myeloma. *Biomed Res Int*. 2014;2014:145619.
207. Dabbah S, Hammerman H, Markiewicz W, Aronson D. Relation between red cell distribution width and clinical outcomes after acute myocardial infarction. *The American journal of cardiology*. 2010;105(3):312-7.
208. Patel KV, Mohanty JG, Kanapuru B, Hesdorffer C, Ershler WB, Rifkind JM. Association of the red cell distribution width with red blood cell deformability. *Adv Exp Med Biol*. 2013;765:211-6.
209. Vaya A, Suescun M. Hemorheological parameters as independent predictors of venous thromboembolism. *Clin Hemorheol Microcirc*. 2013;53(1-2):131-41.
210. Braekkan SK, Mathiesen EB, Njolstad I, Wilsgaard T, Hansen JB. Hematocrit and risk of venous thromboembolism in a general population. The Tromso study. *Haematologica*. 2010;95(2):270-5.
211. Yu FT, Armstrong JK, Tripette J, Meiselman HJ, Cloutier G. A local increase in red blood cell aggregation can trigger deep vein thrombosis: evidence based on quantitative cellular ultrasound imaging. *Journal of thrombosis and haemostasis*. 2011;9(3):481-8.
212. Meyer G, Emmerich J, Helley D, Arnaud E, Nicaud V, Alhenc-Gelas M, et al. Factors V Leiden and II 20210A in patients with symptomatic pulmonary embolism and deep vein thrombosis. *Am J Med*. 2001;110(1):12-5.

213. Salvagno GL, Sanchis-Gomar F, Picanza A, Lippi G. Red blood cell distribution width: A simple parameter with multiple clinical applications. *Critical reviews in clinical laboratory sciences*. 2014;1-20.
214. Fatemi O, Torguson R, Chen F, Ahmad S, Badr S, Satler LF, et al. Red cell distribution width as a bleeding predictor after percutaneous coronary intervention. *American heart journal*. 2013;166(1):104-9.
215. Lappegård J, Ellingsen TS, Skjelbakken T, Mathiesen EB, Njølstad I, Wilsgaard T, et al. Red cell distribution width is associated with future risk of incident stroke. The Tromsø Study. *Thrombosis and Haemostasis*. 2015(2015-08-20 00:00:00).
216. Lappegård J, Ellingsen TS, Vik A, Skjelbakken T, Brox J, Mathiesen EB, et al. Red cell distribution width and carotid atherosclerosis progression. The Tromsø Study. *Thromb Haemost*. 2015;113(3):649-54.
217. Skjelbakken T, Lappegård J, Ellingsen TS, Barrett-Connor E, Brox J, Lochen ML, et al. Red cell distribution width is associated with incident myocardial infarction in a general population: the Tromsø Study. *Journal of the American Heart Association*. 2014;3(4).
218. Sorensen HT, Horvath-Puho E, Pedersen L, Baron JA, Prandoni P. Venous thromboembolism and subsequent hospitalisation due to acute arterial cardiovascular events: a 20-year cohort study. *The Lancet*. 2007;370(9601):1773-9.
219. Lind C, Flinterman LE, Enga KF, Severinsen MT, Kristensen SR, Braekkan SK, et al. Impact of Incident Venous Thromboembolism on Risk of Arterial Thrombotic Diseases. *Circulation*. 2013.
220. Sorensen HT, Horvath-Puho E, Sogaard KK, Christensen S, Johnsen SP, Thomsen RW, et al. Arterial cardiovascular events, statins, low-dose aspirin and subsequent risk of venous thromboembolism: a population-based case-control study. *Journal of thrombosis and haemostasis*. 2009;7(4):521-8.
221. Sørensen HT, Horvath-Puho E, Lash TL, Christiansen CF, Pesavento R, Pedersen L, et al. Heart Disease May Be a Risk Factor for Pulmonary Embolism Without Peripheral Deep Venous Thrombosis. *Circulation*. 2011;124(13):1435-41.
222. Agnelli G, Verso M. Thrombosis and cancer: clinical relevance of a dangerous liaison. *Haematologica*. 2005;90(2):154-6.
223. Heit JA, Silverstein MD, Mohr DN, Petterson TM, O'Fallon W, Melton L, et al. Risk factors for deep vein thrombosis and pulmonary embolism: A population-based case-control study. *Archives of Internal Medicine*. 2000;160(6):809-15.
224. Timp JF, Braekkan SK, Versteeg HH, Cannegieter SC. Epidemiology of cancer-associated venous thrombosis. *Blood*. 2013;122(10):1712-23.
225. Sullivan JL. Iron and the sex difference in heart disease risk. *The Lancet*. 1981;1(8233):1293-4.
226. Lee T-S, Shiao M-S, Pan C-C, Chau L-Y. Iron-deficient diet reduces atherosclerotic lesions in apoE-deficient mice. *Circulation*. 1999;99(9):1222-9.
227. Ponraj D, Makjanic J, Thong PS, Tan BK, Watt F. The onset of atherosclerotic lesion formation in hypercholesterolemic rabbits is delayed by iron depletion. *FEBS letters*. 1999;459(2):218-22.
228. Zheng H, Cable R, Spencer B, Votto N, Katz SD. Iron stores and vascular function in voluntary blood donors. *Arteriosclerosis, thrombosis, and vascular biology*. 2005;25(8):1577-83.
229. Duffy SJ, Biegelsen ES, Holbrook M, Russell JD, Gokce N, Keaney JF, et al. Iron chelation improves endothelial function in patients with coronary artery disease. *Circulation*. 2001;103(23):2799-804.
230. Allen LA, Felker GM, Mehra MR, Chiong JR, Dunlap SH, Ghali JK, et al. Validation and potential mechanisms of red cell distribution width as a prognostic marker in heart failure. *Journal of cardiac failure*. 2010;16(3):230-8.
231. Salonen JT, Nyssönen K, Korpela H, Tuomilehto J, Seppänen R, Salonen R. High stored iron levels are associated with excess risk of myocardial infarction in eastern Finnish men. *Circulation*. 1992;86(3):803-11.
232. Morrison HI, Semenciw RM, Mao Y, Wigle DT. Serum iron and risk of fatal acute myocardial infarction. *Epidemiology*. 1994:243-6.

233. Franchini M, Targher G, Montagnana M, Lippi G. Iron and thrombosis. *Annals of hematology*. 2008;87(3):167-73.
234. Habib A, Finn AV. The role of iron metabolism as a mediator of macrophage inflammation and lipid handling in atherosclerosis. *Frontiers in Pharmacology*. 2014;5:195.
235. Tekin D, Yavuzer S, Tekin M, Akar N, Cin S. Possible effects of antioxidant status on increased platelet aggregation in childhood iron-deficiency anemia. *Pediatr Int*. 2001;43(1):74-7.
236. Hung S-H, Lin H-C, Chung S-D. Association between venous thromboembolism and iron-deficiency anemia: a population-based study. *Blood Coagulation & Fibrinolysis*. 2015;26(4):368-72.
237. Lin C-L, Lin C-L, Tzeng S-L, Chung W-S. Aplastic anemia and risk of deep vein thrombosis and pulmonary embolism: A nationwide cohort study. *Thrombosis research*. 2017;149:70-5.
238. Potaczek DP, Jankowska EA, Wypasek E, Undas A. Iron deficiency: a novel risk factor of recurrence in patients after unprovoked venous thromboembolism. *Pol Arch Med Wewn*. 2016;126(3):159-65.
239. Coutinho JM, Zuurbier SM, Gaartman AE, Dikstaal AA, Stam J, Middeldorp S, et al. Association Between Anemia and Cerebral Venous Thrombosis. Case–Control Study. 2015;46(10):2735-40.
240. Haringa JB, Bracken RL, Nagle SK, Schiebler ML, Patterson BW, Svenson JE, et al. Anemia is not a risk factor for developing pulmonary embolism. *The American Journal of Emergency Medicine*. 2017;35(1):146-9.
241. Vayá A, Sarnago A, Fuster O, Alis R, Romagnoli M. Influence of inflammatory and lipidic parameters on red blood cell distribution width in a healthy population. *Clinical hemorheology and microcirculation*. 2015;59(4):379-85.
242. Emans ME, van der Putten K, van Rooijen KL, Kraaijenhagen RJ, Swinkels D, van Solinge WW, et al. Determinants of red cell distribution width (RDW) in cardiorenal patients: RDW is not related to erythropoietin resistance. *Journal of cardiac failure*. 2011;17(8):626-33.
243. Weiss G, Goodnough LT. Anemia of Chronic Disease. *New England Journal of Medicine*. 2005;352(10):1011-23.
244. Ghaffari S. Oxidative stress in the regulation of normal and neoplastic hematopoiesis. *Antioxid Redox Signal*. 2008;10(11):1923-40.
245. Semba RD, Patel KV, Ferrucci L, Sun K, Roy CN, Guralnik JM, et al. Serum antioxidants and inflammation predict red cell distribution width in older women: the Women's Health and Aging Study I. *Clin Nutr*. 2010;29(5):600-4.
246. Hald EM, Brækkan SK, Mathiesen EB, Njolstad I, Wilsgaard T, Brox J, et al. High-sensitivity C-reactive protein is not a risk factor for venous thromboembolism: the Tromso study. *Haematologica*. 2011;96(8):1189-94.
247. Ridker PM, Cushman M, Stampfer MJ, Tracy RP, Hennekens CH. Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *The New England journal of medicine*. 1997;336(14):973-9.
248. Ellingsen TSL, J.; Brækkan, S.K.; Aukrust, P.; Ueland, T.; Hansen, J.B. ISTH Abstract PB 1392 - Impact of Chronic Inflammation, Assessed by C-reactive Protein, on the Association between Red Cell Distribution Width and Risk of Incident Venous Thromboembolism. *Research and Practice in Thrombosis and Haemostasis*. 2017;1:1-1451.
249. Ellingsen TL, J; Smith, E; Wilsgaard, T; Brækkan, S; Solomon, T; Frazer, K; Hansen, JB. ISTH Abstract PO493TUE - Impact of inherited risk factors of venous thromboembolism (VTE) on the associations between red cell distribution width and risk of VTE. *Journal of Thrombosis and Haemostasis*. 2015;13:1-997.
250. Bucciarelli P, Maino A, Felicetta I, Abbattista M, Passamonti SM, Artoni A, et al. Association between red cell distribution width and risk of venous thromboembolism. *Thrombosis research*. 2015;136(3):590-4.
251. Lippi G, Targher G, Montagnana M, Salvagno GL, Zoppini G, Guidi GC. Relationship between red blood cell distribution width and kidney function tests in a large cohort of unselected outpatients. *Scandinavian journal of clinical and laboratory investigation*. 2008;68(8):745-8.

252. Bujak K, Wasilewski J, Osadnik T, Jonczyk S, Kołodziejska A, Gierlotka M, et al. The prognostic role of red blood cell distribution width in coronary artery disease: a review of the pathophysiology. *Disease markers*. 2015;2015.
253. Montagnana M, Cervellin G, Meschi T, Lippi G. The role of red blood cell distribution width in cardiovascular and thrombotic disorders. *Clinical chemistry and laboratory medicine : CCLM / FESCC*. 2012;50(4):635-41.
254. Kozlitina J, Garcia CK. Red Blood Cell Size Is Inversely Associated with Leukocyte Telomere Length in a Large Multi-Ethnic Population. *PloS one*. 2012;7(12):e51046.
255. Cawthon RM, Smith KR, O'Brien E, Sivatchenko A, Kerber RA. Association between telomere length in blood and mortality in people aged 60 years or older. *The Lancet*. 2003;361(9355):393-5.
256. Blasco MA. Telomeres and human disease: ageing, cancer and beyond. *Nature Reviews Genetics*. 2005;6(8):611-22.
257. Prasad V, Rho J, Cifu A. The diagnosis and treatment of pulmonary embolism: a metaphor for medicine in the evidence-based medicine era. *Archives of internal medicine*. 2012;172(12):955-8.