The risk of venous thromboembolism attributed to established prothrombotic genotypes

Short title: Prothrombotic genotypes and venous thromboembolism

Line H. Evensen[1], Carl Arne Løchen Arnesen[1], Frits R. Rosendaal[1,3], Maiken E. Gabrielsen[4], Ben M. Brumpton[4,5], Kristian Hveem[4,6], John-Bjarne Hansen,[1,2] Sigrid K. Brækkan[1, 2]

[1]Thrombosis Research Center (TREC), Department of Clinical Medicine, UiT-The Arctic University of Norway, Tromsø, Norway

[2]Division of Internal Medicine, University Hospital of North Norway, Tromsø, Norway

[3]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands

[4]K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway

[5]Clinic of Thoracic and Occupational Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway

[6]HUNT Research Center, Department of Public Health and Nursing, Norwegian University of Science and Technology, Levanger, Norway

Correspondence to:

Sigrid K. Brækkan, Thrombosis Research Center (TREC), Department of Clinical Medicine, UiT The Arctic University of Norway, Postboks 6050 Langnes, 9037 Tromsø, Norway

E-mail: sigrid.brakkan@uit.no; Phone: +47 77 62 51 06

Author contributions:

Conception and design: LHE, JBH, SKB, FRR

Data collection: JBH, SKB, MEG, KH

Draft of manuscript: LHE, JBH, SKB

Interpretation of results: LHE, CALA, FRR, MEG, BMB, KH, JBH, SKB

Critical revision: LHE, CALA, FRR, MEG, BMB, KH, JBH, SKB

Abstract

Background: The proportion of venous thromboembolism (VTE) events that can be attributed to established prothrombotic genotypes has been scarcely investigated in the general population. We aimed to estimate the proportion of VTEs in the population that could be attributed to established prothrombotic genotypes using a population-based case-cohort.

Methods: Cases with incident VTE (n=1,493) and a randomly sampled sub-cohort (n=13,069) were derived from the Tromsø Study (1994-2012) and the Nord-Trøndelag Health (HUNT) Study (1995-2008). DNA-samples were genotyped for 17 single nucleotide polymorphism (SNPs) associated with VTE. Hazard ratios with 95% confidence intervals (CIs) were estimated in Cox regression models. Population attributable fraction (PAF) with 95% bias-corrected CIs (based on 10,000 bootstrap samples) were estimated using a cumulative model where SNPs significantly associated with VTE were added one-by-one in ranked order of the individual PAFs.

Results: Six SNPs were significantly associated with VTE (rs1799963 [Prothrombin], rs2066865 [FGG], rs6025 [FV Leiden], rs2289252 [F11], rs2036914 [F11] and rs8176719 [ABO]. The cumulative PAF for the six-SNP model was 45.3% (95% CI 19.7-71.6) for total VTE and 61.7% (95% CI 19.6-89.3) for unprovoked VTE. The PAF for prothrombotic genotypes was higher for DVT (52.9%) than for PE (33.8%), and higher for those aged <70 years (66.1%) than for those aged ≥70 years (24.9%).

Conclusions: Our findings suggest that 45-62% of all VTE events in the population can be attributed to known prothrombotic genotypes. The PAF of established prothrombotic genotypes was higher in DVT than in PE, and higher in the young than in the elderly.

What is known on this topic	What does this paper add?
• Venous thromboembolism (VTE) has a strong heritable component • Several prothrombotic genotypes have been identified • The proportion of VTEs in the population that can be attributed to known prothrombotic genotypes is unclear	• We have estimated the population attributable fraction (PAF) of prothrombotic genotypes in VTE • We show that 45-62% of all VTE events in the population can be attributed to five known prothrombotic genotypes • A higher proportion of VTEs can be attributed to prothrombotic genotypes in the young than in the elderly

Introduction

Venous thromboembolism (VTE) manifests clinically as deep vein thrombosis (DVT) or pulmonary embolism (PE), and is recognized as a multicausal disease that develops in a complex interplay between environmental, acquired and inherited risk factors (1). The inheritance of VTE follows a multifactorial non-Mendelian pattern, which indicates polygenic contribution (2). Since the discovery of antithrombin deficiency in 1965 (3), improved molecular insights and technological advances have led to the identification of more than 50 single-nucleotide polymorphisms (SNPs) associated with VTE risk (4-9). In a large case-control study, deHaan and colleagues investigated the performance of 31 SNPs for prediction of incident VTE (10). They found that a parsimonious model comprising the five SNPs most strongly associated with VTE (i.e., rs6025 [FV Leiden], rs1799963 [Prothrombin], rs8176719 [ABO], rs2066865 [FGG] and rs2036914 [F11] performed as well as the full 31-SNP model, suggesting that a substantial amount of the genetic predisposition of VTE could be attributed to relatively few prothrombotic variants.

Several measures can be used to evaluate the genetic contribution to the development and burden of disease, such as heritability, sibling recurrence risk, impact on the overall genetic variance or population attributable fraction (PAF) (11). In a public health perspective, the PAF is of interest, as it reflects the proportion of cases in a population that is attributable to the risk factor, and indicates by what proportion the incidence of disease would decrease if the risk factor could hypothetically be removed (12). The proportion of VTEs in the population that can be attributed to the already established prothrombotic genotypes has so far been fragmentarily investigated. In a case-control study from 2011, Heit et al. reported a joint PAF of 40% for the SNPs F5 (rs6025), F2 (rs1799963), ABO (rs8176719), and ABO (rs2519093) (13). Since then, other prothrombotic genotypes have been identified, and the joint PAF of these genotypes has, to our knowledge, not been assessed in a cohort of unselected VTE patients within a wide age range. The risk of VTE increases exponentially with age (14), and previous studies have indicated a higher PAF in the young than in the elderly for some

individual genotypes (15). However, the joint PAF of prothrombotic genotypes has not been well quantified and compared between different age groups in the same population.

Previous studies have shown that the FV Leiden (FVL) mutation is associated with a higher risk of DVT than of PE, also referred to as the FVL paradox (16). To what extent this applies to the other genotypes, and whether the joint PAF of prothrombotic genotypes differs in DVT and PE is not well addressed.

The aim of the present study was to estimate the proportion of VTEs in the population that could be attributed to the established VTE-related SNPs, individually and in a cumulative model. Moreover, we aimed to assess the PAF in different age groups and in clinical phenotypes (DVT and PE).

Methods

Study population

Participants were recruited from the fourth survey of the Tromsø Study (Tromsø 4, 1994-95) and the second survey of the Nord-Trøndelag Health Study (HUNT 2, 1995-97). These are two Norwegian population-based cohort studies of the inhabitants in Tromsø municipality and Nord-Trøndelag County, respectively. In Tromsø 4, the entire population aged ≥25 years was invited, and 77% (27,158) participated. In HUNT 2, all inhabitants aged ≥20 years were invited and 71% (66,140) participated. Detailed methodology of the Tromsø Study (17) and the HUNT Study (18) have been published elsewhere.

The participants were followed from the date of enrolment in the respective studies until the date of incident VTE, migration, death or to the end of follow-up, whichever occurred first. Follow-up ended on December 31, 2012 in the Tromsø Study and on December 31, 2008 in the HUNT Study. The process of VTE identification and adjudication in the Tromsø Study (19) and the HUNT Study (20) have previously been described in detail. In the Tromsø Study, VTE events were identified by searching the hospital discharge diagnosis registry, the radiology procedure registry and the autopsy registry at the

University Hospital of North Norway (UNN). The medical records were reviewed by trained personnel and the adjudication criteria were signs and symptoms of DVT or PE, combined with objective confirmation by a radiological procedure that resulted in treatment unless contraindications were specified. In the HUNT Study, VTE events were identified by searching the discharge diagnosis registry and the radiology procedure registry at two local hospitals (Levanger and Namsos) and the discharge diagnosis registry at the tertiary-care center of the region, St. Olavs Hospital in Trondheim. Two physicians reviewed the medical records and the adjudication criterion for VTE was objective confirmation by a radiological procedure.

The composition of the case-cohort is summarized in Supplementary figure 1. In total, there were 1,493 incident VTE events during follow-up, and these were included as cases in the present study. From the Tromsø and HUNT cohorts, 13,072 individuals without previous VTE were randomly selected for the sub-cohort. As all participants in the original cohort had an equal chance of being selected for the sub-cohort, 217 VTE cases were included in the sub-cohort. Participants who were not officially registered as inhabitants of Tromsø or Nord-Trøndelag at baseline (n=3) were excluded from the study. Both studies were approved by the Regional Committees for Medical and Health Research Ethics, and all participants signed an informed consent form prior to inclusion.

Classification of VTE

All events were classified as either PE (with or without DVT) or isolated DVT, and as provoked or unprovoked based on the presence of provoking factors at the time of diagnosis. In the Tromsø Study, provoking factors were: surgery or trauma (within the previous 8 weeks), acute medical conditions (acute myocardial infarction, ischemic stroke or major infectious disease), active cancer, marked immobilization (bedrest ≥ 3 days, confined to wheelchair, or long-distance travel for ≥ 4 days within the previous 14 days), or another provoking factor described by a physician in the medical record (e.g. intravascular catheters). In the HUNT Study, provoking factors were: trauma or surgery, cancer (active malignancy at the time of the event or within 6 months after the event), marked immobilization

(paresis, paralysis, prolonged bedrest due to acute medical illness, or > 8h travel) within the previous 3 months, pregnancy or puerperium at the time of the event or use of oral contraceptives at the time of the event or up to one month prior to the event.

Baseline measurements

Baseline information in both studies was obtained from physical examinations, blood samples and self-administered questionnaires. Body height and weight were measured with participants wearing light clothes and no shoes. Body mass index (BMI) was calculated as weight in kilograms divided by the square of height in meters (kg m^{-2}). Information on diabetes and arterial cardiovascular disease (CVD; myocardial infarction, stroke and angina) was collected via self-report.

Selection of SNPs and genotyping

Based on current knowledge on the genetics of VTE, 17 SNPs with established associations with VTE risk were selected for genotyping in the present study (5, 6). The Tromsø sample was genotyped using the Sequenom and the TaqMan platforms, as previously described (21). The HUNT sample was genotyped using the Illumina HumanCore Exome array.

Individuals were classified as carriers (≥1 risk allele) or non-carriers (0 risk alleles), and no differentiation was made between hetero- and homozygous carriers in the analyses. For the SNPs rs4524 (F5), rs2036914 (F11), rs1801020 (F12) rs1039084 (STXBP5) and rs1613662 (GP6) the major allele was defined as the risk allele (5, 6, 22, 23) . For the rs8176719 SNP in the ABO gene, the G allele tags non-O blood type (risk group), while homozygous carriers for deletion at this site are phenotyped as blood type O (reference group) (24).

Statistical analysis

Statistical analyses were performed with STATA version 15.1 (Stata Corp, College Station, TX, USA). For each participant, person-years of follow-up were accrued from the date of enrolment in the respective

study (The Tromsø Study: 1994-95 and the HUNT Study: 1995-97) to the date of incident VTE, migration, death or to the end of the study period (The Tromsø Study: December 31, 2012 and the HUNT Study: December 31, 2008).

For each SNP, hazard ratios (HRs) of VTE with 95% CIs were estimated in Cox proportional hazards regression models with non-carriers (0 risk alleles) as the reference group. Age was used as time scale with the age at enrolment defined as entry time, and the age at incident VTE or censoring defined as exit time. The analyses were adjusted for age (as time scale), sex and BMI. The proportional hazards assumption was evaluated on the basis of Schoenfeld residuals.

Several measures can be used to evaluate the impact of a risk factor on the development and burden of disease. In a public health perspective, the population attributable fraction (PAF) is of large interest, as it expresses the proportion of cases in a population that can be attributed to a risk factor. PAF is also known as the "preventable fraction" as it indicates the proportion by which the incidence of disease would decrease if the risk factor could (hypothetically) be removed and the distribution of all other risk factors remained unchanged (25, 26). For SNPs significantly associated with VTE risk in the Cox regression models, PAFs were calculated using the formula $\frac{p(HR-1)}{p(HR-1)+1}$, where p is the prevalence in the population (i.e. the sub-cohort) and HR is the risk of VTE in carriers compared with non-carriers of the respective risk allele (25). A cumulative PAF model was constructed by adding SNPs one-by-one into a combined dichotomous exposure variable (i.e., carriers of ≥1 risk allele of any of the added SNPs were categorized as exposed and those with zero risk alleles were categorized as unexposed). The corresponding PAF for each SNP-combination was calculated based on prevalence and estimated HR of VTE for the combined exposure variable. The SNPs were added in order of the size of the individual PAFs, starting with the two SNPs with the highest PAF values. For all PAF estimates, 95% bias-corrected CIs were calculated based on 10 000 bootstrap samples. Separate PAF estimates were calculated for total and unprovoked VTE, for PE and DVT, and by age group (<70 or ≥70 years). Participants with missing information on SNPs included in the PAF model (n=185; 13 cases and 172 in the sub-cohort) and BMI (n=91; 11 cases and 80 in the sub-cohort) were excluded from these

analyses in order to have an identical sample at each step. Due to the high risk allele frequency for the F5 SNP rs4524 (92.4% with ≥1 risk allele), PAF was not estimated for this variant.

Results

The mean age of the study population was 51 ± 17 years (range: 19-97 years). Baseline characteristics of the VTE cases and the sub-cohort are shown in Table 1. Mean age, BMI, systolic blood pressure and prevalence of arterial CVD and diabetes were higher in VTE cases compared with the sub-cohort.

The allele frequency and risk allele distribution in cases and the subcohort are shown in Table 2. With the exception of rs3813948 (C4BPB) and rs1063857 (vWF), the risk allele frequencies for all SNPs were higher among cases. Likewise, the proportions of individuals genotyped with ≥1 risk alleles were higher among the cases for the majority of SNPs. The allele frequencies in the subcohort resembled those previously reported in Caucasian populations (5, 6, 27).

The HRs of VTE for the individual SNPs are shown in Table 3. Of the initially 17 included SNPs, seven were significantly associated with VTE risk. These were (multivariable adjusted HRs): rs6025 (FVL; HR: 2.32, 95% CI 2.01-2.68), rs4524 (F5; HR: 1.52, 95% CI 1.20-1.92), rs1799963 (F2; HR: 1.51, 95% CI 1.05-2.17), rs8176719 (ABO; HR: 1.38, 95% CI 1.24-1.54), rs2289252 (F11; HR: 1.24, 95% CI 1.11-1.38), rs2036914 (F11; HR: 1.22, 95% CI 1.07-1.40) and rs2066865 (FGG; HR: 1.18, 95% CI 1.06-1.30). For DVT and PE the risk estimates differed for FVL, which was more strongly associated with DVT (HR: 2.92, 95% CI 2.46-3.47) than with PE (HR: 1.52 95% CI 1.17-1.98).

Table 4 shows the individual PAFs for the six SNPs associated with VTE risk, as well as the HRs and prevalences used in the PAF calculations. Ranked in descending order, the overall proportion of cases in the population attributable to the individual SNPs (i.e. PAF) was 18.9% (95% CI 10.8-26.4) for rs8176719 (ABO), 14.3% (95% CI 1.4-26.7) for rs2036914 (F11), 13.1% (95% CI 3.8-20.3) for rs2289252 (F11), 8.0% (95% CI 5.3-10.8) for rs6025 (FVL), 6.8% (95% CI 1.2-10.5) for rs2066865 (FGG) and 0.7% (95% CI -0.4-1.7) for rs1799963 (F2). The PAF was higher for DVT than for PE for rs8176719 (ABO, 21.2% vs. 15.6%) and for rs6025 (FVL, 11.2% vs. 3.3%). Analyses stratified by age revealed that the PAFs were

higher among those aged <70 years for rs8176719 (ABO, 28.7% vs. 10.6%), rs2036914 (F11, 20.4% vs. 9.5%) and rs2066865 (FGG, 11.3% vs. 2.9%), while the estimates for the remaining SNPs were comparable between the age groups (Table 5).

The cumulative PAF estimates are shown in Figures 1-3. The PAF increased stepwise with the addition of each SNP until five SNPs were added. These SNPs were rs8176719 (ABO), rs2036914 (F11), rs2289252 (F11), rs6025 (FVL) and rs2066865 (FGG). For total VTE, the cumulative PAF for this model was 45.3% (95% CI 19.7-71.2) (Figure 1, Panel A), while for unprovoked VTE the corresponding PAF was 61.7% (95% CI 19.6-89.3) (Figure 1, Panel B). The cumulative PAF was higher for DVT (52.9%) than for PE (33.8%). For total VTE, the cumulative PAF was higher in those aged <70 years (66.1%, 95% CI 25.5-90.3) compared with those aged ≥70 years (24.9%, 95% CI -13.3-54.6). The HRs that formed the basis for the cumulative PAF models are shown in Supplementary figures 2-4.

Discussion

In the present study, we estimated the proportion of VTEs in the population that could be attributed to already established prothrombotic genotypes. We found that a combination of six SNPs in a cumulative PAF model accounted for 45% of all VTEs and 62% of all unprovoked VTEs in the general population. A larger proportion of the VTEs could be attributed to genotypes in the young (66%) than in the elderly (25%), and the cumulative PAF was also higher in DVT (53%) than in PE (34%).

When the SNPs associated with VTE risk were added in the cumulative model, the PAF estimate increased until five SNPs were added. The SNPs that were included were rs8176719 (ABO), rs2036914 (F11), rs2289252 (F11), rs6025 (FVL) and rs2066865 (FGG), and the summary PAF for this model suggested that 45% of the VTE events in the population could be attributed to these variants. This is in accordance with a previous study reporting a joint PAF of 35% for three of these SNPs (i.e., rs1799963, rs6025 and rs8176719) (13). Interestingly, four of the SNPs were also included in the 5-SNP genetic risk score developed by deHaan and colleagues, which was found to have similar discriminatory accuracy as a risk score comprised of 31 SNPs (10).

In accordance with previous studies, we found that the non-O blood type (rs8176719) had the largest impact on VTE in the population, accounting for 19% of all events (13, 24, 28). The relatively large contribution from this variant is explained by a weak-to-moderate effect size combined with a relatively high prevalence. Our PAF estimate is in accordance with reports from a population-based cohort which reported a PAF of 20% for non-O blood type (24), while the estimates derived from case-control studies tended to be higher (30-40%) (13, 28). Further, we found that the two F11 SNPs contributed significantly to VTE in the population, with PAFs of 13-14%. However, our PAF estimates were markedly lower than those previously reported in a case-control study (24-28%) (22). For FVL and the prothrombin mutation G2021A, our PAF estimates were in accordance with previous studies (13, 24, 28-30), although considerable variations exist between geographical locations due to differences in prevalence (31). For instance, in the Netherlands and in southern Sweden, the high prevalence of the FVL mutation yielded PAF estimates of 20-50% (31). In addition, the observed differences in reported PAF estimates could to some extent be explained by differences in study design, as higher estimates are typically reported in case-control studies. If the distribution of exposure in the control population differs from that of the source population, the true effect may be overestimated, and consequently, the PAF will also be overestimated.

Though we are not aware of previous studies investigating the joint PAF for unprovoked VTE, our findings of a stronger genetic influence is in accordance with studies reporting higher prevalence of thrombophilia among patients with unprovoked VTE (32). Higher effect sizes in unprovoked compared with total or provoked VTE have been suggested for several prothrombotic genotypes, including the prothrombin G2021A mutation and FVL (30). The observation that 60% of the unprovoked events can be attributed to these genotypes, highlights the need for unraveling novel genetic or environmental risk factors for unprovoked VTE. The characteristics of yet unknown genotypes may be either common variants with very low effect sizes, or extremely rare or private mutations with large effect sizes. In a recent genome-wide/transcriptome-wide association study (GWAS/TWAS) including >30000 VTE cases and >170000 controls from 18 studies in the INVENT

consortium, 16 novel susceptibility loci for VTE were identified (27). However, the effect sizes (odds ratios) for these variants were small, ranging from 0.80 to 1.14. Furthermore, the INVENT consortium also performed an exome-wide array analysis based on data from 11 studies to identify novel associations with VTE for low-frequency variants (33). While associations with previously known loci were confirmed, no new variants were identified, and larger studies with sequencing data were requested (33). In the other end of the continuum, rare mutations with large effect sizes may be identified in family studies. Importantly, variants with these characteristics will have a small impact on the burden of VTE at the population level.

Using the cumulative model, we found that the PAF of prothrombotic genotypes was higher for DVT than for PE. This difference appeared to be mainly driven by a higher PAF for DVT by the individual SNPs in ABO and FVL. For ABO, previous studies have not reported major differences in the risk of DVT and PE (16). Accordingly, the hazard ratios for DVT (HR 1.46) and PE (HR 1.29) according to ABO status were not statistically different in our study, but the resulting individual PAF estimates were 21.2% for DVT and 15.6% for PE. For FVL, the risk estimates for DVT were significantly higher than for PE, and the individual PAFs were 11.2% in DVT and 3.3% in PE. The higher risk of DVT than of PE in subjects with FVL, is well recognized and often referred to as the “FVL paradox” (16). This phenomenon could be explained by the formation of more stable clots less susceptible for embolization in patients with FVL. Indeed, a recent experimental study in mice reported that thrombi in FVL carriers were larger and embolized less compared with wild type (34).

In the present study, we found that the prothrombotic genotypes exerted a stronger influence on the occurrence of VTE among individuals <70 years. The higher PAF for prothrombotic genotypes in the younger age group was particularly explained by higher HRs of VTE for the SNPs rs8176719 (ABO), rs2303914 (F11) and rs2066865 (FGG). An age effect has previously been reported for the association between family history and the risk of VTE, with the highest effect sizes observed in the younger age groups (35, 36). Further, the proportion of cases with a prothrombotic genotype has been reported to be significantly higher among those aged 20 years or younger (49.3%) compared with those

aged 70 years and older (21.9%), suggesting that individuals with a genetic susceptibility experience VTE at a younger age (32). As aging is associated with an accumulation of acquired and environmental risk factors, the relative contribution of the genetic factors may be diluted. The age effect may also be explained by the phenomenon 'attrition of susceptibles' where those highly vulnerable are likely to develop thrombosis early in life, and an apparently resilient elderly population (15).

Main strengths of our study include the recruitment of participants from a general population cohort with a wide age range, unselected and objectively validated VTEs, and comprehensive information on a large number of prothrombotic genotypes. Our findings are derived from a Caucasian population, which lowers the risk of population stratification, but limits the generalizability of the results to other ethnicities. The PAF is specific for the population it is derived from, and differences in allele frequency between populations have implications for the relevance of the mutation and the occurrence of thrombosis. PAF depends on both the prevalence of the risk factor and its effect size. Importantly, these estimates are only valid under the assumption that the risk factor is causal and are also specific for the population from which they were derived (26, 31). Although PAF is a useful tool to study the impact of genetic factors, its direct dependence on prevalence might lead to overestimation when the prevalence is very high (11, 37). Due to a high prevalence in the cumulative PAF model, the percentages may be overestimated. Finally, the rs8176719 used in this study is not optimal for evaluating VTE risk mediated by the ABO locus as it does not take into account the A2 and O2 blood groups which are associated with a lower risk of VTE compared to A1 and B (38). However, as the prevalence of A2 and O2 in the population is low (haplotype frequencies <5% and <1%, respectively), this is likely to have a negligible influence on our estimated PAF.

In conclusion, we found that 45% of all VTEs and 62% of the unprovoked VTEs in the population could be attributed to established prothrombotic genotypes, and that this was mainly explained by five SNPs. The proportion of events that could be attributed to genes was higher in the young than in the elderly, and higher in DVT than in PE.

Sources of Funding

K.G. Jebsen Thrombosis Research and Expertise Center (TREC) received an independent grant from Stiftelsen Kristian Gerhard Jebsen in the period 2014-2020.

Disclosures

The authors report no conflict of interest. No funding was received to assist with the preparation of the manuscript.

References

1. Lijfering WM, Rosendaal FR, Cannegieter SC. Risk factors for venous thrombosis - current understanding from an epidemiological point of view. British journal of haematology 2010; 149: 824-33.

2. Crous-Bou M, Harrington LB, Kabrhel C. Environmental and Genetic Risk Factors Associated with Venous Thromboembolism. Seminars in thrombosis and hemostasis 2016.

3. Egeberg O. Inherited Antithrombin Deficiency Causing Thrombophilia. Thromb Diath Haemorrh 1965; 13: 516-30.

4. Germain M, Chasman DI, de Haan H, et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. American journal of human genetics 2015; 96: 532-542.

5. Tregouet DA, Morange PE. What is currently known about the genetics of venous thromboembolism at the dawn of next generation sequencing technologies. British journal of haematology 2018; 180: 335-345.

6. Morange PE, Suchon P, Tregouet DA. Genetics of Venous Thrombosis: update in 2015. Thromb Haemost 2015; 114: 910-9.

7. Klarin D, Emdin CA, Natarajan P, Conrad MF, Kathiresan S. Genetic Analysis of Venous Thromboembolism in UK Biobank Identifies the ZFPM2 Locus and Implicates Obesity as a Causal Risk Factor. Circulation Cardiovascular genetics 2017; 10.

8. Klarin D, Busenkell E, Judy R, et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. Nat Genet 2019; 51: 1574-1579.

9. Lindstrom S, Wang L, Smith EN, et al. Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism. Blood 2019; 134: 1645-1657.

10. de Haan HG, Bezemer ID, Doggen CJ, et al. Multiple SNP testing improves risk prediction of first venous thrombosis. Blood 2012; 120: 656-63.

11. Witte JS, Visscher PM, Wray NR. The contribution of genetic variants to disease depends on the ruler. Nature reviews Genetics 2014; 15: 765-776.

12. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. Am J Public Health 1998; 88: 15-9.

13. Heit JA, Cunningham JM, Petterson TM, Armasu SM, Rider DN, M DEA. Genetic variation within the anticoagulant, procoagulant, fibrinolytic and innate immunity pathways as risk factors for venous thromboembolism. Journal of thrombosis and haemostasis : JTH 2011; 9: 1133-42.

14. 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. European journal of epidemiology 2017; 32: 299-305.

15. Engbers MJ, van Hylckama Vlieg A, Rosendaal FR. Venous thrombosis in the elderly: incidence, risk factors and risk groups. J Thromb Haemost 2010; 8: 2105-12.

16. van Langevelde K, Flinterman LE, van Hylckama Vlieg A, Rosendaal FR, Cannegieter SC. Broadening the factor V Leiden paradox: pulmonary embolism and deep-vein thrombosis as 2 sides of the spectrum. Blood 2012; 120: 933-46.

17. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njolstad I. Cohort profile: the Tromso Study. Int J Epidemiol 2012; 41: 961-7.

18. Holmen J, Midthjell K, Krüger Ø, et al. The Nord-Trøndelag Health Study 1995-1997 (HUNT 2): Objectives, contents, methods and participation. Nor Epidemiol 2003; 13: 19-32.

19. 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. J Thromb Haemost 2008; 6: 1851-7.

20. Naess IA, Christiansen SC, Romundstad P, Cannegieter SC, Rosendaal FR, Hammerstrom J. Incidence and mortality of venous thrombosis: a population-based study. J Thromb Haemost 2007; 5: 692-9.

21. Horvei LD, Braekkan SK, Smith EN, et al. Joint effects of prothrombotic genotypes and body height on the risk of venous thromboembolism: the Tromso study. J Thromb Haemost 2018; 16: 83-89.

22. Li Y, Bezemer ID, Rowland CM, et al. Genetic variants associated with deep vein thrombosis: the F11 locus. Journal of thrombosis and haemostasis : JTH 2009; 7: 1802-8.

23. Johnson CY, Tuite A, Morange PE, Tregouet DA, Gagnon F. The factor XII -4C>T variant and risk of common thrombotic disorders: A HuGE review and meta-analysis of evidence from observational studies. American journal of epidemiology 2011; 173: 136-44.

24. Sode BF, Allin KH, Dahl M, Gyntelberg F, Nordestgaard BG. Risk of venous thromboembolism and myocardial infarction associated with factor V Leiden and prothrombin mutations and blood

type. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne 2013; 185: E229-37.

25. Zapata-Diomedi B, Barendregt JJ, Veerman JL. Population attributable fraction: names, types and issues with incorrect interpretation of relative risks. British Journal of Sports Medicine 2018; 52: 212-213.

26. Bhopal RJ. Concepts of Epidemiology. Integrating the ideas, theories, principles, and methods of epidemiology. Third ed. Oxford, Oxford University Press; 2016.

27. Lindstrom S, Wang L, Smith EN, et al. Genomic and Transcriptomic Association Studies Identify 16 Novel Susceptibility Loci for Venous Thromboembolism. Blood 2019.

28. Roach REJ, Cannegieter SC, Lijfering WM. Differential risks in men and women for first and recurrent venous thrombosis: the role of genes and environment. Journal of Thrombosis and Haemostasis 2014; 12: 1593-1600.

29. Zöller B, Melander O, Svensson PJ, Engström G. Factor V Leiden paradox in a middle-aged Swedish population: A prospective study. Vascular Medicine 2018; 23: 52-59.

30. Folsom AR, Cushman M, Tsai MY, Heckbert SR, Aleksic N. Prospective study of the G20210A polymorphism in the prothrombin gene, plasma prothrombin concentration, and incidence of venous thromboembolism. American journal of hematology 2002; 71: 285-90.

31. Rosendaal FR. Risk factors for venous thrombotic disease. Thrombosis and haemostasis 1999; 82: 610-9.

32. Weingarz L, Schwonberg J, Schindewolf M, et al. Prevalence of thrombophilia according to age at the first manifestation of venous thromboembolism: results from the MAISTHRO registry. British journal of haematology 2013; 163: 655-665.

33. Lindstrom S, Brody JA, Turman C, et al. A large-scale exome array analysis of venous thromboembolism. Genetic epidemiology 2019.

34. Shaya SA, Westrick RJ, Gross PL. Thrombus stability explains the factor V Leiden paradox: a mouse model. Blood advances 2019; 3: 3375-3378.

35. Bezemer ID, van der Meer FM, Eikenboom JJ, Rosendaal FR, Doggen CM. The value of family history as a risk indicator for venous thrombosis. Archives of internal medicine 2009; 169: 610-615.

36. 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: 1012-20.

37. Rockhill B, Weinberg CR, Newman B. Population attributable fraction estimation for established breast cancer risk factors: considering the issues of high prevalence and unmodifiability. American journal of epidemiology 1998; 147: 826-33.

38. Goumidi L, Thibord F, Wiggins KL, et al. Association between ABO haplotypes and the risk of venous thrombosis: impact on disease risk estimation. Blood 2021; 137: 2394-2402.

Figure legends:

Figure 1. Cumulative population attributable fraction (PAF) with 95% confidence intervals of total (Panel A) and unprovoked (Panel B) venous thromboembolism (VTE) based on increasing number of single-nucleotide polymorphisms (SNPs). SNPs were added in ranked order of the individual PAF estimates (Table 4)

Figure 2. Cumulative population attributable fraction (PAF) of deep vein thrombosis (DVT) and pulmonary embolism (PE) based on increasing number of single-nucleotide polymorphisms (SNPs)

Figure 3. Cumulative population attributable fraction (PAF) of venous thromboembolism based on increasing number of single-nucleotide polymorphisms stratified by age (< and ≥ 70 years)

Tables and figures

Table 1. Baseline characteristics of the venous thromboembolism (VTE) cases and the sub-cohort

	VTE cases (n=1493)	Sub-cohort (n=13069)
Age (years), mean ± SD	61 ± 15	51 ± 17
Sex (female), % (n)	52.9 (790)	52.9 (6909)
Body mass index (kg m^2), mean ± SD	27.4 ± 4.4	26.2 ± 4.1
Systolic blood pressure (mm Hg), mean ± SD	145 ± 23	138 ± 22
Arterial cardiovascular disease*, % (n)	14.3 (213)	8.3 (1083)
Diabetes, % (n)	4.3 (64)	3.1 (410)

*Myocardial infarction, stroke or angina

SD standard deviation

Table 2. Allele frequency (AF) and distribution of risk alleles in the venous thromboembolism (VTE) cases and the sub-cohort

Gene	SNP	VTE cases (n=1493)*			Sub-cohort (n=13069)*		
		AF (%)	1 risk allele (%, n)	2 risk alleles (%, n)	AF (%)	1 risk allele (%, n)	2 risk alleles (%, n)
F5	**rs4524**	76.6	36.7 (547)	58.3 (870)	72.9	39.2 (5122)	53.2 (6954)
FVL	**rs6025**	7.7	14.4 (214)	0.5 (8)	3.4	6.5 (854)	0.1 (15)
SERP	**rs2227589**	9.8	17.2 (257)	1.1 (17)	8.7	15.8 (2069)	0.8 (107)
C4BPB	**rs3813948**	7.4	13.7 (205)	0.5 (8)	7.7	14.2 (1848)	0.6 (78)
KNG1	**rs710446**	42.1	46.6 (695)	18.8 (281)	41.2	48.4 (6325)	17.0 (2224)
FGG	**rs2066865**	27.3	37.4 (559)	8.6 (128)	24.0	36.3 (4740)	5.8 (756)
F11	**rs2036914**	57.3	49.6 (739)	32.5 (484)	53.1	50.2 (6533)	28.0 (3646)
F11	**rs2289252**	43.4	48.8 (727)	19.0 (284)	39.2	47.4 (6188)	15.5 (2021)
F12	**rs1801020**	76.3	34.8 (431)	58.9 (728)	74.4	38.0 (4865)	55.4 (7092)
F13	**rs5985**	28.8	40.5 (496)	8.5 (104)	26.7	39.1 (4983)	7.1 (907)
STXBP5	**rs1039084**	53.8	48.1 (718)	29.7 (443)	51.5	50.4 (6582)	26.3 (3431)
F2	**rs3136520**	3.1	5.9 (87)	0.2 (3)	3.0	5.8 (763)	0.1 (16)
F2	**rs1799963**	1.0	2.1 (31)	-	0.7	1.3 (173)	-
vWF	**rs1063857**	37.8	48.3 (695)	13.7 (197)	38.1	46.6 (5489)	14.8 (1744)
TC2N	**rs1884841**	44.5	50.7 (755)	19.1 (285)	43.0	48.2 (6299)	18.9 (2471)
GP6	**rs1613662**	86.0	23.2 (346)	74.4 (1110)	82.6	28.7 (3749)	68.3 (8917)
ABO	**rs8176719**	43.1	51.3 (764)	17.4 (260)	38.3	46.3 (6027)	15.2 (1978)

AF allele frequency; *SNP* single-nucleotide polymorphism

*Not all participants have data on all SNPs

Table 3. Incidence rates (IRs) and hazard ratios (HRs) with 95% confidence intervals (CIs) for venous thromboembolism by individual single-nucleotide polymorphisms (SNPs)

		Venous thromboembolism			Deep vein thrombosis			Pulmonary embolism		
		0 risk alleles	≥1 risk alleles	HR (95% CI)*	0 risk alleles	≥1 risk alleles	HR (95% CI)*	0 risk alleles	≥1 risk alleles	HR (95% CI)*
Gene	SNP	Events	Events	≥1 vs. 0 risk alleles	Events	Events	≥1 vs. 0 risk alleles	Events	Events	≥1 vs. 0 risk alleles
FVL	**rs6025**	1258	221	2.32 (2.01-2.68)	719	159	2.92 (2.46-3.47)	539	62	1.52 (1.17-1.98)
F5	**rs4524**	75	1406	1.52 (1.20-1.92)	50	829	1.34 (1.01-1.79)	25	577	1.87 (1.25-2.79)
F2	**rs1799963**	1446	30	1.51 (1.05-2.17)	860	15	1.26 (0.76-2.10)	586	15	1.88 (1.13-3.14)
ABO	**rs8176719**	464	1015	1.38 (1.24-1.54)	266	612	1.45 (1.26-1.68)	198	403	1.29 (1.09-1.53)
F11	**rs2289252**	474	1006	1.24 (1.11-1.38)	283	596	1.23 (1.06-1.41)	191	410	1.25 (1.05-1.48)
F11	**rs2036914**	265	1214	1.22 (1.07-1.40)	159	719	1.21 (1.02-1.44)	106	495	1.24 (1.01-1.53)
FGG	**rs2066865**	799	683	1.18 (1.06-1.30)	468	412	1.21 (1.06-1.38)	331	271	1.13 (0.97-1.33)
GP6	**rs1613662**	36	1445	1.16 (0.83-1.62)	18	861	1.39 (0.87-2.22)	18	584	0.93 (0.58-1.49)
F13	**rs5985**	621	593	1.11 (0.99-1.24)	376	355	1.10 (0.95-1.28)	245	238	1.13 (0.94-1.34)
TC2N	**rs1884841**	444	1034	1.11 (0.99-1.24)	265	612	1.10 (0.95-1.27)	179	422	1.12 (0.94-1.34)
SERP	**rs2227589**	1212	269	1.09 (0.96-1.25)	728	151	1.02 (0.85-1.21)	484	118	1.20 (0.98-1.47)
STXBP5	**rs1039084**	327	1154	1.09 (0.97-1.23)	177	702	1.22 (1.04-1.44)	150	452	0.94 (0.78-1.13)
vWF	**rs1063857**	543	885	1.03 (0.92-1.15)	326	524	1.02 (0.88-1.17)	217	361	1.05 (0.89-1.24)
F2	**rs3136520**	1377	90	1.04 (0.84-1.29)	819	53	1.03 (0.78-1.36)	558	37	1.07 (0.77-1.50)
KNG1	**rs710446**	516	966	1.02 (0.92-1.13)	309	571	1.00 (0.87-1.15)	207	395	1.05 (0.88-1.24)
F12	**rs1801020**	78	1148	0.97 (0.77-1.22)	40	692	1.14 (0.83-1.57)	38	456	0.79 (0.57-1.10)
C4BPB	**rs3813948**	1270	211	0.97 (0.84-1.13)	764	115	0.87 (0.72-1.06)	506	96	1.12 (0.90-1.40)

*Adjusted for age (as time scale) and sex and body mass index

Table 4. Population attributable fraction (PAF) of venous thromboembolism (VTE), deep vein thrombosis (DVT) and pulmonary embolism (PE) for individual single-nucleotide polymorphisms (SNPs)

Gene	SNP	HR (95% CI)*	Sub-cohort prevalence†	PAF (95% CI)
ALL				
ABO	**rs8176719**	1.38 (1.23-1.54)	0.62	18.9 (10.8-26.4)
F11	**rs2036914**	1.21 (1.06-1.39)	0.78	14.3 (1.4-26.7)
F11	**rs2289252**	1.24 (1.11-1.38)	0.63	13.1 (3.8-20.3)
FVL	**rs6025**	2.31 (2.00-2.66)	0.07	8.0 (5.3-10.8)
FGG	**rs2066865**	1.17 (1.06-1.30)	0.42	6.8 (1.2-10.5)
F2	**rs1799963**	1.54 (1.07-2.21)	0.01	0.7 (-0.4-1.7)
DVT				
ABO	**rs8176719**	1.44 (1.24-1.66)	0.62	21.2 (10.8-31.5)
F11	**rs2036914**	1.20 (1.01-1.43)	0.78	13.6 (-2.6-29.1)
F11	**rs2289252**	1.22 (1.06-1.41)	0.63	12.3 (1.4-23.3)
FVL	**rs6025**	2.90 (2.44-3.44)	0.07	11.2 (7.9-14.9)
FGG	**rs2066865**	1.20 (1.05-1.37)	0.42	7.6 (0.0-15.4)
F2	**rs1799963**	1.28 (0.77-2.14)	0.01	0.4 (-0.7-1.7)
PE				
ABO	**rs8176719**	1.30 (1.10-1.54)	0.62	15.6 (2.7-28.2)
F11	**rs2036914**	1.23 (1.00-1.52)	0.78	15.5 (-3.7-34.1)
F11	**rs2289252**	1.27 (1.07-1.50)	0.63	14.3 (1.2-27.3)
FVL	**rs6025**	1.52 (1.17-1.97)	0.07	3.3 (0.2-6.8)
FGG	**rs2066865**	1.14 (0.97-1.34)	0.42	5.6 (-3.2-14.8)
F2	**rs1799963**	1.93 (1.16-3.23)	0.01	1.2 (-0.4-3.0)

*Adjusted for age (as time scale), sex and body mass index; †≥1 risk allele

Table 5. Population attributable fraction (PAF) of venous thromboembolism (VTE) for individual single-nucleotide polymorphisms (SNPs) stratified by age group (< and ≥ 70 years).

Gene	SNP	HR (95% CI)*	Sub-cohort prevalence†	PAF (95% CI)
<70 years				
ABO	**rs8176719**	1.65 (1.40-1.96)	0.62	28.7 (16.7-34.6)
F11	**rs2036914**	1.33 (1.09-1.62)	0.78	20.4 (3.9-35.3)
F11	**rs2289252**	1.26 (1.07-1.48)	0.63	13.9 (0.6-24.6)
FVL	**rs6025**	2.31 (1.88-2.85)	0.07	8.1 (4.2-13.1)
FGG	**rs2066865**	1.30 (1.12-1.51)	0.42	11.3 (3.0-20.5)
F2	**rs1799963**	1.51 (0.90-2.52)	0.01	0.7 (-0.7-2.0)
≥70 years				
ABO	**rs8176719**	1.19 (1.03-1.38)	0.61	10.6 (2.6-23.1)
F11	**rs2036914**	1.13 (0.95-1.35)	0.79	9.5 (-9.1-23.5)
F11	**rs2289252**	1.23 (1.06-1.43)	0.63	12.8 (-0.7-23.6)
FVL	**rs6025**	2.30 (1.88-2.81)	0.07	7.9 (4.7-11.2)
FGG	**rs2066865**	1.07 (0.93-1.23)	0.41	2.9 (-2.7-11.3)
F2	**rs1799963**	1.56 (0.94-2.60)	0.01	0.7 (-0.6-2.0)

*Adjusted for age (as time scale), sex and body mass index; †≥1 risk allele

Figure 1. Cumulative population attributable fraction (PAF) with 95% confidence intervals of total (Panel A) and unprovoked (Panel B) venous thromboembolism (VTE) based on increasing number of single-nucleotide polymorphisms (SNPs). SNPs were added in ranked order of the individual PAF estimates (Table 4)

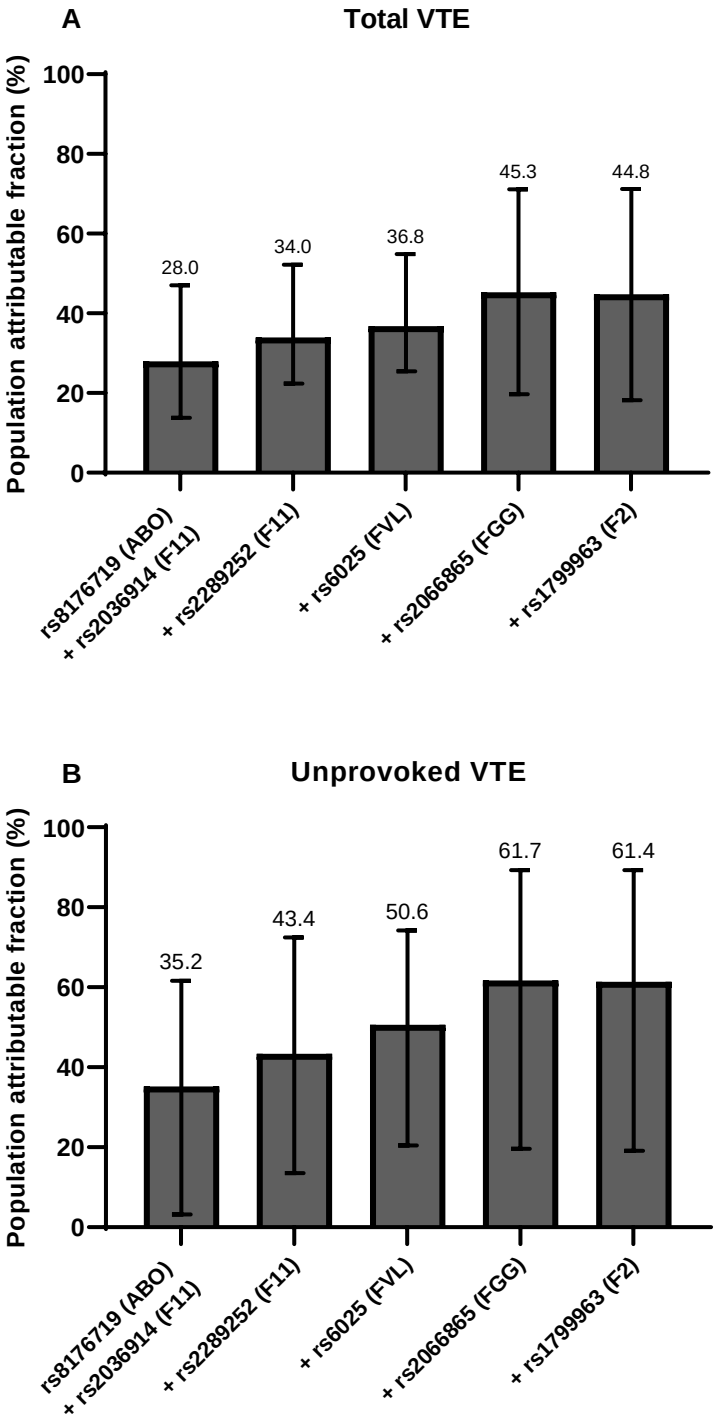

Figure 2. Cumulative population attributable fraction (PAF) of deep vein thrombosis (DVT) and pulmonary embolism (PE) based on increasing number of single-nucleotide polymorphisms (SNPs)

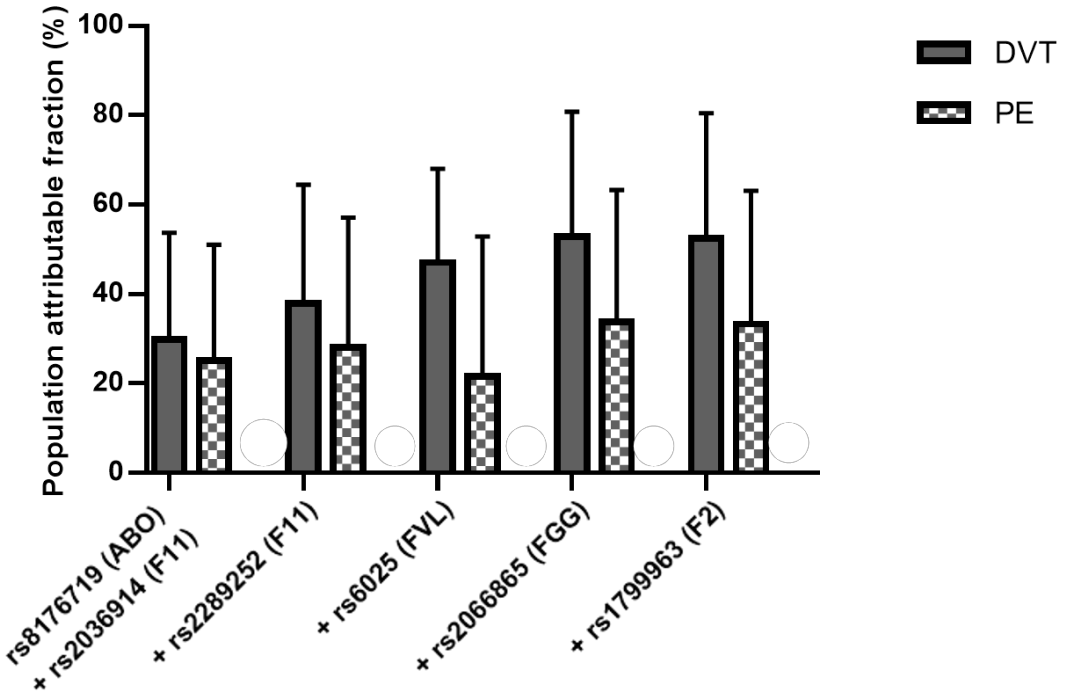

Figure 3. Cumulative population attributable fraction (PAF) of venous thromboembolism based on increasing number of single-nucleotide polymorphisms stratified by age (< and ≥ 70 years)