



Universiteit
Leiden
The Netherlands

Genetic determinants of venous thrombosis

Haan, H.G. de

Citation

Haan, H. G. de. (2020, January 8). *Genetic determinants of venous thrombosis*. Retrieved from <https://hdl.handle.net/1887/82479>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/82479>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/82479> holds various files of this Leiden University dissertation.

Author: Haan, H.G. de

Title: Genetic determinants of venous thrombosis

Issue Date: 2020-01-08

CHAPTER 6

Multiple SNP testing improves risk prediction of first venous thrombosis

de Haan HG, Bezemer ID, Doggen CJ, Le Cessie S, Reitsma PH, Arellano AR, Tong CH, Devlin JJ, Bare LA, Rosendaal FR, Vossen CY.

ABSTRACT

There are no risk models available yet that accurately predict an individual's risk for developing venous thrombosis. Our aim was therefore to explore whether inclusion of established thrombosis-associated SNPs in a venous thrombosis risk model improves the risk prediction. We calculated genetic risk scores by counting risk-increasing alleles from 31 venous thrombosis-associated SNPs for subjects of a large case-control study including 2712 patients and 4634 controls (MEGA). Genetic risk scores based on all 31 SNPs or on the 5 most strongly associated SNPs performed similarly (areas under receiver-operating characteristic curves (AUCs) of 0.70 and 0.69 respectively). For the 5-SNP risk score, the odds ratios for venous thrombosis ranged from 0.37 (95% CI 0.25-0.53) for individuals with 0 risk alleles to 7.48 (95% CI 4.49-12.46) for individuals with ≥ 6 risk alleles. The AUC of a risk model based on known non-genetic risk factors was 0.77 (95% CI 0.76-0.78). Combining the non-genetic and genetic risk models improved the AUC to 0.82 (95% CI 0.81-0.83), indicating good diagnostic accuracy. In order to become clinically useful, subgroups of high-risk individuals must be identified in whom genetic profiling will also be cost-effective.

INTRODUCTION

Venous thrombosis is the result of innate thrombotic tendency and non-genetic triggers. Many common genetic variants, mainly single-nucleotide polymorphisms (SNPs), with modest effects on risk of venous thrombosis have been reported.¹ Individual SNPs have little predictive value due to their modest effect on risk, but combinations of gene variants may improve the predictive ability and could be used to model susceptibility to venous thrombosis.

Simulation studies have shown that so-called genetic profiling may be useful to discriminate between individuals with high risk of disease and those with low risk. The discriminative accuracy of genetic profiling depends on the heritability and incidence of the disease and on the frequencies of risk alleles.^{2,3}

Genetic profiling has become a popular aim in epidemiologic studies of many common diseases since a large amount of data from genome-wide association studies (GWAS) has become available.²⁻⁸ For recurrent venous thrombosis, we previously investigated the potential clinical utility of multiple SNP testing for recurrent events.⁹ In that study, individual SNPs were not significantly associated with recurrent venous thrombosis. However, when the risk alleles of the individual SNPs were combined, the risk estimates as well as the significance of the association increased. The predictive ability of multiple SNP analysis has not been studied for first events of venous thrombosis. Genetic profiling may guide decisions on prophylactic measures in high-risk groups such as cancer patients, individuals undergoing surgery, requiring a plaster cast or those subject to prolonged immobilization.

In order to explore to what extent venous-thrombosis associated SNPs can be used as predictors for a first venous thrombosis in the general population and in high-risk groups, we investigated 31 SNPs in two large population-based case-control studies, of which one was used as a validation set. We created genetic risk scores based on these SNPs and a risk score based on non-genetic risk factors. We also compared and combined our genetic risk score with the non-genetic risk score to determine whether genetic profiling with the currently known SNPs will improve the assessment of venous thrombosis risk.

METHODS

Study populations

The Multiple Environmental and Genetic Assessment of risk factors for venous thrombosis (MEGA study) is a population-based case-control study of venous thrombosis. Collection and ascertainment of events have been described in detail previously.^{10,11} The MEGA analysis included 2712 consecutive patients with a diagnosis of a first deep vein thrombosis of the leg or arm (with or without pulmonary embolism) and 4634 control subjects (partners of patients and random population controls).

The Leiden Thrombophilia Study (LETS), another population-based case-control study of venous thrombosis, was used to validate the risk scores and included 443 consecutive patients with a diagnosis of a first deep vein thrombosis of the leg (with or without pulmonary embolism) and 453 control subjects (acquaintances or partners of patients), all without a known malignancy. Collection and ascertainment of events have been described in detail previously.¹² Both studies were approved by the Medical Ethics Committee of the Leiden University Medical Center, Leiden, the Netherlands.

SNP selection

Initially we selected 40 SNPs for the genetic risk score, based on the literature and our previous work. Eighteen SNPs had been reported and repeatedly confirmed to be associated with venous thrombosis.^{1,13} Twelve SNPs were added from the Group Health study,^{13,14} these SNPs were associated with venous thrombosis in the original study and replicated in the MEGA study. Nine SNPs were added from a large SNP association analysis including subsequent fine mapping that we performed recently in LETS and MEGA.^{15,16} Another added SNP was recently identified in a follow-up study of a GWAS and replicated in the FARIVE study and the MEGA study.¹⁷ Among the 40 SNPs in the initial selection, we studied linkage disequilibrium and mutually adjusted SNPs within genes. Four SNPs in *PROC* (rs1799808, rs1799810, rs2069915 and rs5937) were explained by rs1799809 in *PROC*; 4 SNPs in the fibrinogen genes (rs6050 and rs2070006 in *FGA*, rs1800788 in *FGB* and rs2066854 in *FGG*) were explained by rs2066865 in *FGG*; and rs3753305 in *F5* was explained by rs6025 (factor V Leiden). Consequently, we excluded 9 SNP associations that were explained by other SNPs. The remaining 31 SNPs (Table 1) were included in the genetic risk score.

Genetic risk score

We defined a genetic risk score that counts the total number of risk-increasing alleles in individuals. To take into account the stronger association of some SNPs with venous thrombosis, we also constructed a weighted risk score assigning weights to the risk alleles of each SNP corresponding to the logarithm of the average risk estimates found in literature. In addition to the full genetic model including 31 SNPs, we constructed a parsimonious model with fewer SNPs. To determine which SNPs should be included in this model, we added SNPs one-by-one to create the genetic risk score. We started with the SNP with the highest odds ratio in literature and assessed whether adding SNPs to the risk score improved the AUC after each SNP addition. The addition of SNPs was stopped when the AUC of the risk score including the newly added SNP did not differ from the AUC of the full genetic model.

Non-genetic risk factors

We constructed a non-genetic risk score, which included the following risk factors: recent (within three months prior to the index date) leg injury, surgery, pregnancy or postpartum, immobilization (i.e. plaster cast, bedridden at home, hospitalization), travel for more than four hours in two months prior to the index date, oral contraceptives (OC) use or hormone replacement therapy (HRT) at the index date, obesity (body mass index $>30\text{kg/m}^2$) and a cancer diagnosis between five years before and six months after the index date. The index date was defined as date of diagnosis for patients and their partner controls, and the date of completing the questionnaire for random controls. We also included family history in the non-genetic risk score. Family history was defined as positive when a parent or sibling had experienced venous thrombosis and negative when none of these relatives had experienced venous thrombosis, or when the participant was not aware of venous thrombosis in the family. We assigned weights to each non-genetic risk factor corresponding to the logarithm of the risk estimates in MEGA (Supplemental Table 1) and constructed a simple risk scoring system counting the weighted risk factors. We also constructed a combined risk score including both the genetic risk score and the non-genetic risk score using a logistic regression model.

Application of genetic profiling may be most useful in high-risk groups, i.e. individuals exposed to known non-genetic risk factors. We therefore studied the discriminative accuracy of our genetic risk score as well as the combined scores in high-risk situations of surgery, plaster cast, hospitalization, young women (under 50 years) using oral

contraceptives, women using HRT, pregnancy or postpartum, middle-aged individuals (above 50 years) and travel. We also studied individuals with a positive family history and individuals with malignant disorders.

Statistical analyses

Crude and sex-adjusted (in case SNPs were located on the X chromosome) odds ratios and 95% confidence intervals were calculated by logistic regression for individual SNPs and the genetic, non-genetic and combined risk scores. When assessing the magnitude of risk associated with number of risk alleles, we used the median number of risk alleles among control subjects as the reference group.

To assess how well a score classifies venous thrombosis patients and control subjects, we calculated the area under the receiver-operating characteristic (ROC) curve (AUC). The AUC ranges from 0.5 (no discrimination between patients and control subjects) to 1.0 (perfect discrimination). We compared the AUCs of the different genetic and non-genetic risk models according to the method of Hanley *et al.*¹⁸ Nagelkerke's pseudo- r^2 statistic was used to approximate the proportion of variability explained by the different risk models. All analyses, including ROC curves and AUC calculation were performed in SPSS for Windows, 17.0.2 (SPSS Inc, Chicago, Ill).

RESULTS

SNPs associated with venous thrombosis

Table 1 lists all associations between SNPs and venous thrombosis in the MEGA population and the average estimated effect-size in literature.^{13-17,19-26} Not all SNPs were associated with venous thrombosis in our study populations; nevertheless, we included all 31 SNPs in the genetic risk score because these SNPs had been associated with venous thrombosis in other studies.

Table 1. 31 SNP associations with venous thrombosis in MEGA and literature.^{13-17; 19-26}

Gene	SNP	Chr	Position	MEGA				Literature
				Cases	Controls	OR	95% CI	Average OR
<i>F5</i>	rs6025	1	167.785.673	10	3	4.30	(3.70 -4.99)	3.79
<i>F2</i>	rs1799963	11	46.717.631	6	2	3.01	(2.36 -3.85)	2.78
<i>ABO</i>	rs8176719	9	136.132.908	47	34	1.74	(1.63 -1.87)	1.85
<i>FGG</i>	rs2066865	4	155.744.726	34	27	1.41	(1.32 -1.51)	1.56
<i>F11</i>	rs2036914	4	187.429.475	59	52	1.35	(1.26 -1.44)	1.32
<i>PROCR</i>	rs2069951	20	33.227.425	7	5	1.32	(1.16 -1.51)	1.30
<i>F11</i>	rs2289252	4	187.444.375	48	41	1.36	(1.28 -1.45)	1.26
<i>F9</i>	rs4149755	X	138.451.778	7	6	1.11	(0.99 -1.24)	1.24
<i>PROCR</i>	rs2069952	20	33.227.612	64	60	1.21	(1.13 -1.29)	1.21
<i>SERPINC1</i>	rs2227589	1	172.152.839	11	9	1.27	(1.15 -1.41)	1.20
<i>HIVEP1</i>	rs169713	6	11.920.517	22	20	1.10	(1.01-1.19)	1.20
<i>F2</i>	rs3136516	11	46.717.332	52	49	1.12	(1.06 -1.20)	1.19
<i>F5</i>	rs1800595	1	167.776.972	6	5	1.18	(1.03 -1.36)	1.18
<i>PROC</i>	rs1799809	2	127.892.345	47	43	1.17	(1.10 -1.25)	1.17
<i>PROCR</i>	rs867186	20	33.228.215	14	12	1.18	(1.07 -1.29)	1.17
<i>VWF</i>	rs1063856	12	6.153.534	37	33	1.18	(1.10-1.26)	1.16
<i>GP6</i>	rs1613662	19	60.228.407	84	82	1.18	(1.09 -1.29)	1.15
<i>F2</i>	rs3136520	11	46.699.808	3	2	1.09	(0.89 -1.32)	1.13
<i>F8</i>	rs1800291	X	153.811.479	85	83	1.12	(1.05 -1.20)	1.13
<i>STXBP5</i>	rs1039084	6	147.635.413	42	45	0.90	(0.84-0.96)	0.90
<i>NAT8B</i>	rs2001490	2	73.781.606	40	37	1.13	(1.06 -1.20)	1.10
<i>F13B</i>	rs6003	1	195.297.644	9	10	1.11	(1.00 -1.24)	1.09
<i>RGS7</i>	rs670659	1	239.228.398	67	64	1.14	(1.06 -1.22)	1.09
<i>F9</i>	rs6048	X	138.460.946	72	70	1.09	(1.03 -1.16)	1.08
<i>F5</i>	rs4524	1	167.778.379	79	74	1.31	(1.22 -1.42)	0.92
<i>F13A1</i>	rs5985	6	6.263.794	76	76	1.03	(0.95 -1.10)	0.93
<i>F3</i>	1208 indel	1	94.780.000	46	46	1.02	(0.96 -1.09)	1.06
<i>TFPI</i>	rs8176592	2	188.040.937	69	68	1.04	(0.97 -1.11)	1.06
<i>F11</i>	rs3822057	4	187.425.146	55	49	1.31	(1.23 -1.39)	1.06
<i>NR1I2</i>	rs1523127	3	120.983.729	41	38	1.15	(1.08 -1.23)	1.05
<i>CPB2</i>	rs3742264	13	45.546.095	69	68	1.04	(0.97 -1.11)	1.01

Abbreviations: Chr=chromosome; OR=odds ratio; CI=confidence interval

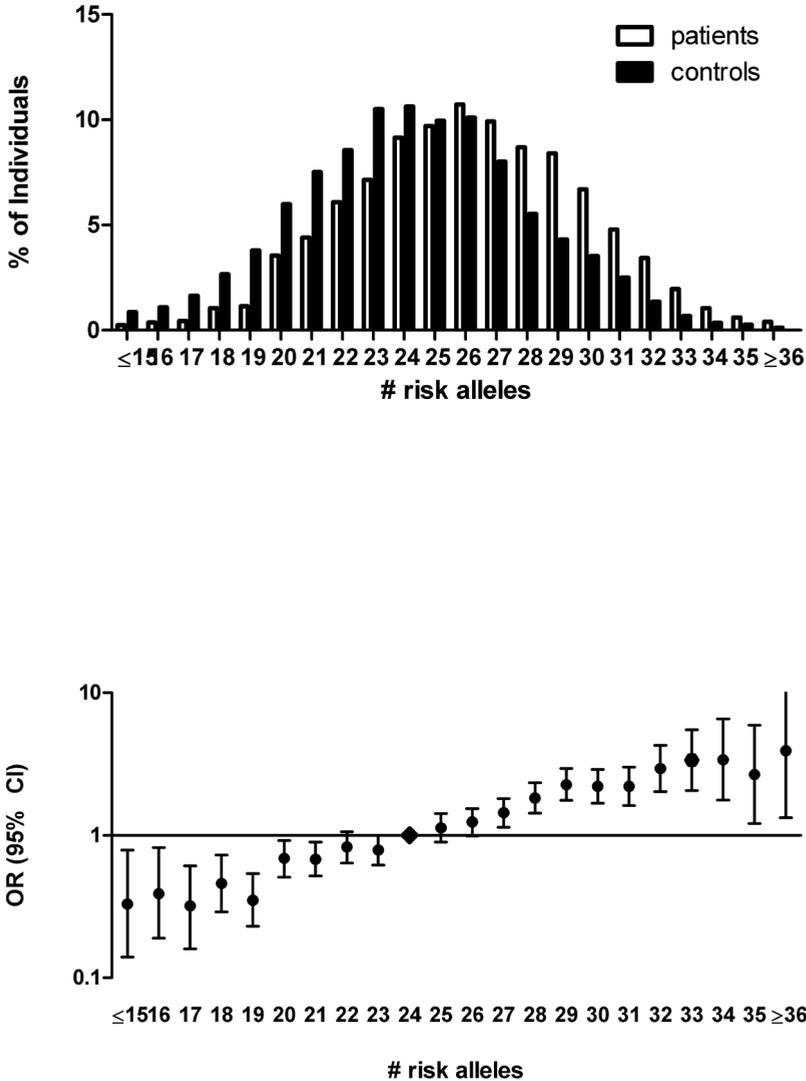


Figure 1. 31-SNP risk allele distribution in patients with venous thrombosis and control subjects (upper panel of figure) and corresponding odds ratios (lower panel).

Odds ratios (95% confidence interval) for venous thrombosis were calculated relative to the median number of risk alleles among control subjects (24 risk alleles). Individuals with 15 or less and 36 or more risk alleles were combined for the calculation of the odds ratio because of the low numbers of individuals with that few or many risk alleles.

Genetic risk score

We first included all 31 SNPs in the genetic risk score. For each individual we counted the number of risk-increasing alleles. The number of risk alleles ranged from 13 to 38 with a median of 24 among control subjects and 26 among cases (Figure 1). The risk for venous thrombosis was estimated for each number of risk alleles, relative to the median number of risk alleles of 24, and ranged from an odds ratio of 0.27 (95% CI 0.13-0.56) for 16 risk alleles to an odds ratio of 3.23 (95% CI 1.96-5.30) for 33 risk alleles. At the more extreme ends of the risk distribution, confidence intervals around risk estimates became very wide due to small numbers. The average relative risk increase per risk allele, when treated as an ordinal variable, however, could be estimated with a high level of precision, and was 1.14 (95% CI 1.12-1.16). This corresponds to an about 100-fold difference in risk between the lowest and the highest number of risk alleles in our population.

We also constructed a weighted risk score thereby assigning weight to the risk alleles according to their risk estimates found in literature (Table 1). A few SNPs have only been studied in the MEGA population; in that case we used the risk estimate in MEGA as weight. The ROC curve for the weighted 31-SNP risk score had an AUC of 0.71 (Table 2: 95% CI 0.69-0.72); i.e., there is a 71% probability that a randomly chosen patient will have a higher score than a randomly chosen control subject. The weighted 31-SNP risk score was a better predictor than the non-weighted 31-SNP risk score (AUC 0.64, 95% CI 0.63-0.65). The average relative risk increase per unit in the risk score, when treated as an ordinal variable, was 7.89 (95% CI 6.76-9.21). The proportion of variability explained by the 31-SNP risk score was 16.1% (Nagelkerke's pseudo- r^2 ; Table 2).

Table 2. Venous thrombosis prediction using genetic, non-genetic and combined risk scores. The LETS study was used as a validation set.

	MEGA (N=7092)		LETS (N=881)	
	AUC (95% CI)	Nagelkerke pseudo r^2	AUC (95% CI)	Nagelkerke pseudo r^2
31-SNP risk score	0.71 (0.69-0.72)	0.161	0.69 (0.65-0.72)	0.149
5-SNP risk score	0.69 (0.67-0.70)	0.135	0.67 (0.64-0.71)	0.138
Non-genetic risk score	0.77 (0.76-0.78)	0.288	0.71 (0.68-0.74)	0.200
Combined risk score	0.82 (0.81-0.83)	0.378	0.77 (0.74-0.80)	0.292

Abbreviations: AUC=area under the receiver-operating characteristic (ROC) curve; CI=confidence interval

In order to construct a genetic risk score using the most parsimonious model, we added SNPs one-by-one to the genetic risk score, starting with the SNP with the highest OR in literature (Factor V Leiden, rs6025), and calculated the AUC after the addition of each SNP (Figure 2). The AUC for each single SNP ranged from 0.50 (95% CI 0.49-0.52) for rs3136520 in *F2* to 0.60 (95% CI 0.59-0.61) for rs8176719 in *ABO*. The discriminative accuracy of the model improved rapidly with the addition of each SNP, until 5 SNPs were included in the model (Figure 2). These SNPs were rs6025 (*F5*, factor V Leiden), rs1799963 (*F2*, 20210 G>A), rs8176719 (*ABO*), rs2066865 (*FGG* 10034 C>T) and rs2036914 (*F11*). The AUC for this 5-SNP risk score was 0.69 (Table 2, 95% CI 0.67-0.70). Moreover, a model based on the three most well-known prothrombotic polymorphisms (i.e. rs6025, rs1799963 and rs8176719; AUC 0.65, 95% CI 0.64-0.66) performed significantly worse than the 5-SNP risk score. The average relative risk increase per unit in the risk score, when treated as an ordinal, was 9.50 (95% CI 7.92-11.39). The 5-SNP risk score explained 13.5% of the total variability (Nagelkerke's pseudo- r^2 ; Table 2).

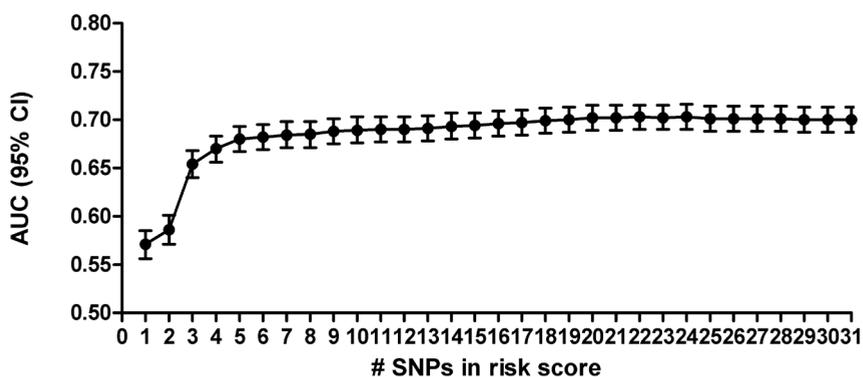


Figure 2. Area under the ROC of genetic risk scores based on increasing numbers of SNPs.

SNPs were added in order of the odds ratio as found in the literature, starting with rs6025 in the score based on 1 SNP, and ending with *CPB2* included in the score of 31 SNPs (Table 1).

The number of risk alleles in the 5-SNP risk score ranged from 0 (OR 0.37, 95% CI 0.26-0.53) to 8 (OR 7.48, 95% CI 4.49-12.46 for ≥ 6 risk alleles), with a median number of risk alleles of 2 among control subjects (Figure 3). The relative increase in risk per increase in number of risk alleles was 1.61 (95% CI 1.54-1.68), again corresponding to an over 100-fold difference in risk between the lowest and the highest number of risk alleles. The weighted 5-SNP risk score was a better predictor than a non-weighted model based on number of risk alleles (AUC 0.66, 95% CI 0.64-0.67).

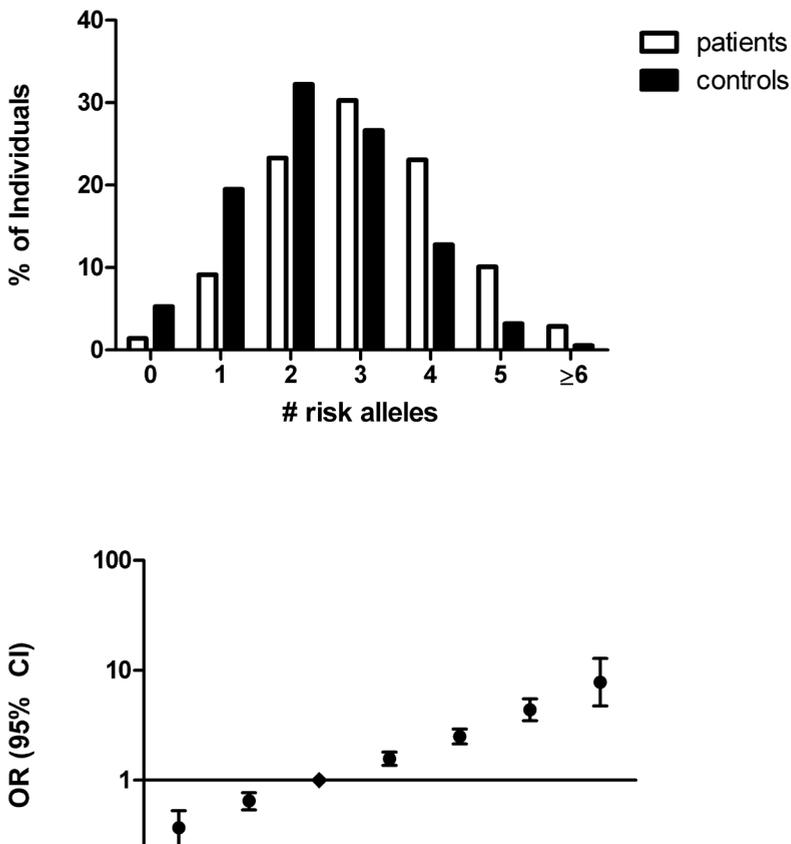


Figure 3. 5-SNP risk allele distribution in patients with venous thrombosis and control subjects (upper panel of figure) and corresponding odds ratios (lower panel).

Odds ratios (95% confidence interval) for venous thrombosis were calculated relative to the median number of risk alleles among control subjects (score 2). Individuals with 6 or more risk alleles were combined for the calculation of the odds ratio because of the low numbers of individuals with that few or many risk alleles.

No difference between the discriminative accuracy of the 5-SNP risk score in men (AUC 0.69, 95% CI 0.67-0.71) and women (AUC 0.67, 95% CI 0.65-0.69) was found. However, differences were found when we constructed and compared the 5-SNP genetic risk score in patients with DVT in the arm, patients with DVT in the leg and patients with DVT in the leg combined with PE. The AUC of the 5-SNP risk score in patients with DVT in the arm (AUC 0.62, 95% CI 0.57-0.67) was significantly lower than in patients with DVT in the leg (AUC 0.68, 95% CI 0.67-0.70) or for DVT combined with PE (AUC 0.68, 95% CI 0.67-0.70).

High-risk groups and SNP testing

To explore clinical applications of genetic profiling, we studied groups exposed to known non-genetic factors in more detail. The discriminative accuracy of the genetic risk scores in these subgroups was similar to the discriminative accuracy in the overall study population, except among cancer patients (Table 3). Sub-analysis in cancer patients according to therapy (chemotherapy, surgery, radiation) or tumor class (solid versus other) did not improve the discriminative accuracy of the weighted 5-SNP risk score (data not shown).

To assess whether the genetic risk score performs better than the current clinical practice of assessing family history, we compared the discriminative accuracy of the genetic risk score with a risk score with family history alone. The AUC of the 5-SNP risk score (0.68, 95% CI 0.67-0.70) was significantly higher than the AUC of family history (0.58, 95% CI 0.57-0.60), with a similar trend observed among all subgroups of high-risk individuals (Table 3).

Table 3. Risk score prediction in sub-groups of individuals exposed to known non-genetic risk factors.

High Risk group	Patients, N	Control subjects, N	Family history risk score, AUC (95%CI)	5-SNP risk score, AUC (95%CI)	Non-genetic risk score, AUC (95%CI)	Combined risk score, AUC (95%CI)
Surgery	292	111	0.60 (0.55-0.66)	0.66 (0.60-0.72)	0.67 (0.61-0.72)	0.73 (0.67-0.78)
Plaster cast	111	18	0.61 (0.48-0.73)	0.73 (0.59-0.87)	0.70 (0.56-0.84)	0.78 (0.64-0.91)
Hospitalization	278	93	0.57 (0.50-0.63)	0.66 (0.59-0.72)	0.60 (0.53-0.66)	0.66 (0.59-0.72)
Oral contraceptives*	513	327	0.58 (0.54-0.62)	0.71 (0.68-0.75)	0.73 (0.69-0.76)	0.81 (0.78-0.84)
HRT	58	90	0.59 (0.49-0.68)	0.71 (0.63-0.80)	0.74 (0.66-0.82)	0.80 (0.72-0.87)
Pregnancy/Postpartum*	67	46	0.54 (0.44-0.65)	0.70 (0.60-0.79)	0.68 (0.57-0.79)	0.76 (0.66-0.85)
Age>50 years	944	1534	0.57 (0.55-0.60)	0.68 (0.66-0.70)	0.73 (0.71-0.75)	0.79 (0.77-0.81)
Travel	379	610	0.58 (0.54-0.62)	0.70 (0.67-0.73)	0.77 (0.73-0.80)	0.82 (0.80-0.85)
Family history	659	551	-	0.68 (0.65-0.71)	0.74 (0.71-0.76)	0.81 (0.78-0.83)
Malignancies	156	65	0.57 (0.49-0.65)	0.60 (0.52-0.68)	0.71 (0.64-0.78)	0.72 (0.65-0.80)

Abbreviations: AUC=area under the receiver-operating characteristic (ROC) curve; CI=confidence interval; HRT=hormone replacement therapy
 * Women under 50 years.

Combining non-genetic and genetic risk scores

We assessed the discriminative accuracy of a non-genetic risk score based on known non-genetic risk factors for venous thrombosis (leg injury, surgery, pregnancy, plaster cast, bedridden at home, hospitalization, travel, OC use, HRT, obesity and malignancy) and family history. For the individual components the AUC ranged from 0.50 (95% CI 0.48-0.51) for recent travel to 0.67 (95% CI 0.65-0.69) for OC use by women. The AUC for the non-genetic risk score including family history was 0.77 (95% CI 0.76-0.78). When we added the genetic risk score to the non-genetic score, the AUC significantly increased to 0.82 (Figure 4: 95% CI 0.81-0.83) compared with the non-genetic risk score alone (p -value <0.0001) using either the 31-SNP or the 5-SNP risk score. In addition, 28.8% of the total variability in venous disease risk was explained by the non-genetic risk score, which significantly improved to 37.8% (Nagelkerke pseudo r^2 ; Table 2) when combining the non-genetic and genetic risk scores. Both the non-genetic and the combined risk score models performed better in women than in men (non-genetic risk score: AUC 0.81, 95% CI 0.80-0.83 for women and AUC 0.74, 95% CI 0.72-0.75 for men; combined risk score: AUC 0.85, 95% CI 0.83-0.86 for women and AUC 0.80, 95% CI 0.78-0.81 for men).

We also studied the discriminative accuracy of the combined risk score model in the high-risk groups. For all subgroups the AUC improved when using the combined risk score compared with the non-genetic risk score, which was significant for individuals using oral contraceptives, individuals with a positive family history of venous thrombosis and individuals over 50 years old (Table 3).

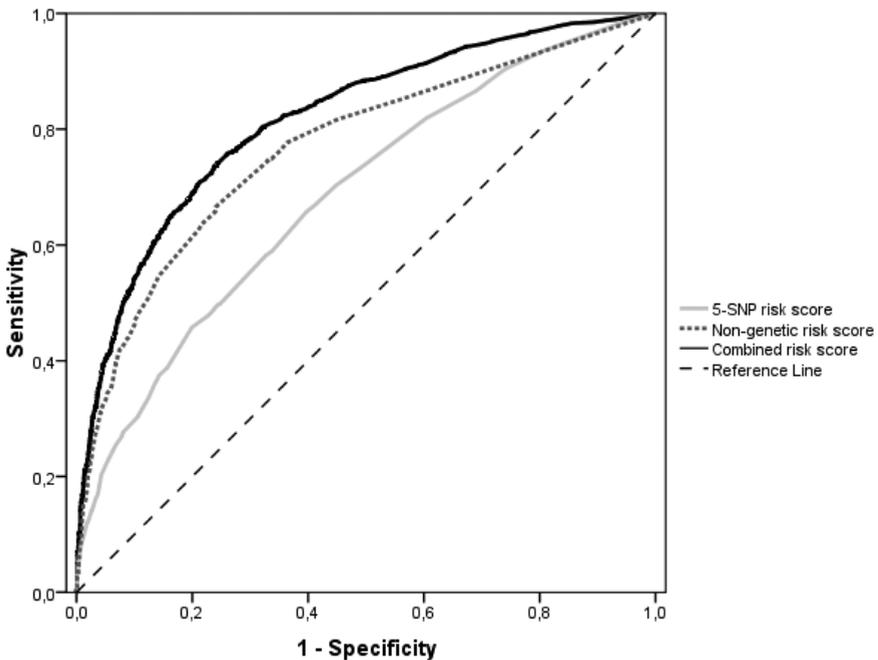


Figure 4. ROC (AUC) curves of the weighted 5-SNP risk score (light grey line), the non-genetic risk score (dotted grey line) and the combined risk score (black line).

The striped black line represents the reference line (no discrimination).

Validation of the risk scores

In order to validate the genetic, non-genetic and combined risk scores, we studied their discriminative accuracy in subjects from another population, the LETS population. As described in the Methods, LETS and MEGA are both population-based case control studies and are similar with respect to mean age at index of patients (45 years in LETS, 47 years in MEGA) or control subjects (45 years in LETS, 48 years in MEGA) and sex distribution (43% men in LETS, 47% men in MEGA). Associations between the 31 SNPs and venous thrombosis in LETS can be found in Supplemental Table 2. The discriminative accuracy of the weighted 31-SNP and 5-SNP risk scores in LETS were 0.69 (95% CI 0.65-0.72) and 0.67 (95% CI 0.64-0.71) respectively, which are similar to those found in MEGA (Table 2).

We also constructed the non-genetic risk score weighted according to the risk estimates of each risk factor from MEGA, except for malignancies as having cancer was an exclusion criterion in LETS. In addition, information of some non-genetic risk factors, i.e. HRT, recent travel, leg injury and plaster cast was not assessed in LETS or not in such detail as in MEGA. Therefore, these risk factors were excluded from the non-genetic risk score. The discriminative accuracy of the non-genetic risk score in LETS was 0.71 (95% CI 0.68-0.74) and improved to 0.77 (95% CI 0.74-0.80) when combined with the genetic risk score. Both risk scores performed slightly better in MEGA than in LETS (Table 2).

DISCUSSION

We calculated a genetic risk score based on SNPs consistently associated with venous thrombosis and observed a 'dose-response' relationship between this score and the risk of venous thrombosis. The more risk alleles or genotypes present, the higher the risk of venous thrombosis. A score constructed of the 5 most strongly associated SNPs appeared to differentiate between patients and control subjects equally as well as the initial genetic risk score based on 31 SNPs. The discriminative accuracy of both the 5-SNP and 31-SNP risk score was replicated in another study (LETS) suggesting robustness of the genetic models.

When preventive measures following a positive test are invasive or can have harmful side-effects, strict discrimination is required between those at high risk and low risk of developing a specific disease. In the case of venous thrombosis, indiscriminate may lead to an increased risk of thrombosis in high-risk individuals receiving insufficient prophylactic anticoagulant treatment, whereas individuals at low risk receiving treatment are at an increased risk of major bleeding. We investigated the extent to which genetic risk scores can improve the accuracy of thrombosis risk assessment by means of ROC curves. The 5-SNP genetic score performed better than family history assessment, which is the current clinical practice of risk assessment in individuals exposed to known non-genetic risk factors. However, the 5-SNP genetic risk score performed worse than a risk score of non-genetic risk factors. A recent study by Hippisley-Cox and Coupland²⁷ showed that an algorithm of non-genetic risk factors is able to discriminate between patients and control subjects with an AUC of 0.75. This is similar to the AUC observed

with our non-genetic risk score (0.77). However, the AUC may be an overestimation since we used (the logarithm of) the risk estimates from MEGA as weights.

Here, we showed that addition of the 5-SNP genetic risk score to the non-genetic risk score model significantly improved the AUC to 0.82, indicating good diagnostic accuracy. In our validation study, information on the non-genetic risk factors was less complete, which explains the lower discriminative accuracy of both the non-genetic risk score (0.71) and the combined risk score (0.77).

Identification of individuals at risk of developing venous thrombosis is most useful in high-risk populations. This is because the incidence of venous thrombosis in the general population is too low (1 per 1000 individuals a year²⁸) to justify genotyping of all individuals. In all subgroups of high-risk individuals the combined risk score performed better than the non-genetic score alone, which may indicate the potential clinical value of genetic profiling in these high-risk individuals.

We defined a basic genetic risk score that counts the total number of risk-increasing alleles in individuals. To take into account the stronger association of some SNPs with venous thrombosis, we assigned literature-based weights to each SNP, which discriminated patients better from controls than a non-weighted genetic risk score. Although the proportion of variability explained by the 5-SNP risk score is smaller than by the 31-SNP risk score, we showed that the discriminative accuracy of the 5-SNP and 31-SNP risk scores was similar. The genetic risk score is still limited though by its assumption that all SNPs act independently and in an additive manner in venous thrombosis susceptibility. An additive effect was assumed for the different genotypes, whereas we cannot exclude a multiplicative effect. Gene-gene interaction and gene-environment interaction is not taken into account, while in reality many interactions exist. Examples for venous thrombosis are the synergistic effects between factor V Leiden (rs6025) and oral contraceptive use²⁹ and between the *F13A1* Val34Leu variant (rs5985) and fibrinogen levels.³⁰ We chose to include SNPs on their contribution to risk (effect size) and gave weights corresponding to the logarithm of this effect size. This is the most relevant for an individual who has a certain genotype. One could argue that on a population level, the prevalence of risk alleles is of relevance. However, this would not be expected to improve the performance of the risk prediction model, and indeed a genetic risk model based on the 5 SNPs with the highest risk allele frequency

in MEGA performed worse than the non-weighted 5-SNP risk score which is based on the 5 SNPs with the highest effect-size (AUC 0.54, 95% CI 0.53-0.56 and AUC 0.66, 95% CI 0.64-0.67, respectively).

In the future, adding newly discovered predictive SNPs to the model may further improve discrimination. In a simulation study, Janssens et al² showed that the AUC depends on the number of SNPs included, and their OR and risk allele frequency. The heritability of a disease determines the maximum obtainable AUC. For venous thrombosis the heritability is estimated to be about 60%.^{31,32} The simulation study indicated that at this level high AUCs (>0.90) can be obtained, given that all genetic contributors are in the prediction model. Identification of new genetic predictors and validation of the genetic risk score in other study populations will reveal whether genetic profiling is useful in venous thrombosis.

In summary, we demonstrated that addition of a 5-SNP risk score to a risk scoring system based on non-genetic risk factors significantly improved the risk prediction of venous thrombosis. Although additional predictive markers may be required for a risk score to be clinically useful in the general population, the 5-SNP risk score may aid the management of subgroups of high-risk individuals.

REFERENCES

1. Bezemer ID, Rosendaal FR. Predictive genetic variants for venous thrombosis: what's new? *Semin Hematol.* 2007; 44(2):85-92.
2. Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM. Predictive testing for complex diseases using multiple genes: fact or fiction? *Genet Med.* 2006; 8(7):395-400.
3. Janssens AC, Moonesinghe R, Yang Q, Steyerberg EW, van Duijn CM, Khoury MJ. The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases. *Genet Med.* 2007; 9(8):528-535.
4. Brautbar A, Ballantyne CM, Lawson K, et al. Impact of adding a single allele in the 9p21 locus to traditional risk factors on reclassification of coronary heart disease risk and implications for lipid-modifying therapy in the Atherosclerosis Risk In Communities Study. *Circ Cardiovasc Genet.* 2009; 2(3):279-285.
5. Paynter NP, Chasman DI, Pare G, et al. Association between a Literature-Based Genetic Risk Score and Cardiovascular Events in 19,313 Women. *JAMA.* 2010; 303(7): 631-637.
6. Davies RW, Dandona S, Stewart AFR, et al. Improved prediction of cardiovascular disease based on a panel of single nucleotide polymorphisms identified through genome-wide association studies. *Circ Cardiovasc Genet.* 2010; 3(5): 468-474.
7. Chen H, Poon A, Yeung C, et al. A genetic risk score combining ten psoriasis risk loci improves disease prediction. *PLoS One.* 2011; 6(4): e19454.
8. Ripatti S, Tikkanen E, Orho-Melander M, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet.* 2010; 376(9750):1393-1400.
9. van Hylckama Vlieg A, Baglin CA, Bare LA, Rosendaal FR, Baglin TP. Proof of principle of potential clinical utility of multiple SNP analysis for prediction of recurrent venous thrombosis. *J Thromb Haemost.* 2008; 6(5):751-754.
10. Blom JW, Doggen CJM, Osanto S, Rosendaal FR. Malignancies, prothrombotic mutations, and the risk of venous thrombosis. *JAMA.* 2005; 293(6):715-722.
11. van Stralen KJ, Rosendaal FR, Doggen CJM. Minor injuries as a risk factor for venous thrombosis. *Arch Intern Med.* 2008; 168(1):21-26.
12. Koster T, Rosendaal FR, De Ronde H, Briët E, Vandenbroucke JP, Bertina RM. Venous thrombosis due to poor anticoagulant response to activated protein C: Leiden Thrombophilia Study. *Lancet.* 1993; 342(8886-8887):1503-1506.
13. Smith NL, Hindorf LA, Heckbert SR, et al. Association of genetic variations with nonfatal venous thrombosis in postmenopausal women. *JAMA.* 2007; 297(5):489-498.

14. Smith NL, Rice KM, Bovill EG, et al. Genetic variation associated with plasma von Willebrand factor levels and the risk of incident venous thrombosis. *Blood*. 2011;117(22):6007-6011.
15. Bezemer ID, Bare LA, Doggen CJM, et al. Gene variants associated with deep vein thrombosis. *JAMA*. 2008; 299(11):1306-1314.
16. Li Y, Bezemer I, Rowland CM, et al. Genetic variants associated with deep vein thrombosis: the F11 locus. *J Thromb Haemost*. 2009; 7(11):1802-1808.
17. Morange PE, Bezemer I, Saut N, et al. A follow-up study of a genome-wide association scan identifies a susceptibility locus for venous thrombosis on chromosome 6p24.1. *Am J Hum Genet*. 2010;86(4):592-595.
18. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983; 148(3):839-843.
19. Delluc A, Gourhant L, Lacut K, et al. Association of common genetic variations and idiopathic venous thromboembolism. Results from EDITH, a hospital-based case-control study. *Thromb Haemost*. 2010; 103(6):1161-1169.
20. Emmerich J, Rosendaal FR, Cattaneo M, 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):809-816.
21. Jick H, Slone D, Westerholm B, et al. Venous thromboembolic disease and ABO blood type. A cooperative study. *Lancet*. 1969; 15;1(7594):539-542.
22. Grünbacher G, Weger W, Marx-Neuhold E, et al. The fibrinogen gamma (FGG) 10034C>T polymorphism is associated with venous thrombosis. *Thromb Res*. 2007; 121(1):33-36.
23. Uitte de Willige S, Pyle ME, Vos HL, et al. Fibrinogen gamma gene 3'-end polymorphisms and risk of venous thromboembolism in the African-American and Caucasian population. *Thromb Haemost*. 2009; 101(6):1078-1084.
24. Smith NL, Wiggins KL, Reiner AP, et al. Replication of findings on the association of genetic variation in 24 hemostasis genes and risk of incident venous thrombosis. *J Thromb Haemost*. 2009; 7(10):1743-1746.
25. Austin H, De Staercke C, Lally C, et al. New gene variants associated with venous thrombosis: a replication study in white and black Americans. *J Thromb Haemost*. 2011; 9(3):489-495.
26. Wells PS, Anderson JL, Scarvelis DK, et al. Factor XIII Val34Leu variant is protective against venous thromboembolism: a HuGe review and meta-analysis. *Am J Epidemiol*. 2006; 164:101-109.
27. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithm (QThrombosis) to estimate future risk of venous thromboembolism: prospective cohort study. *BMJ*. 2011; 343: d4656.

28. Naess IA, Christiansen SC, Romundstad P, Cannegieter SC, Rosendaal FR, Hammerström J. Incidence and mortality of venous thrombosis: a population-based study. *J Thromb Haemost.* 2007; 5(4):692-699.
29. Vandenbroucke JP, Koster T, Briët E, Reitsma PH, Bertina RM, Rosendaal FR. Increased risk of venous thrombosis in oral-contraceptive users who are carriers of factor V Leiden mutation. *Lancet.* 1994; 344(8935):1453-1457.
30. Vossen CY, Rosendaal FR. The protective effect of the factor XIII Val34Leu mutation on the risk of deep venous thrombosis is dependent on the fibrinogen level. *J Thromb Haemost.* 2005; 3(5):1102-1103.
31. Souto JC, Almasy L, Borrell M, 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):1452-1459.
32. Larsen TB, Sorensen HT, Skytthe A, Johnsen SP, Vaupel JW, Christensen K. Major genetic susceptibility for venous thromboembolism in men: a study of Danish twins. *Epidemiology.* 2003; 14(3):328-332.

SUPPLEMENTAL TABLES**Table S1.** Associations between non-genetic risk factors and venous thrombosis risk in MEGA.

Non-genetic risk factors	MEGA		OR	95% CI
	Risk factor frequency, %			
	cases	Controls		
Plaster cast	5	1	5.35	(3.21-8.92)
Leg injury	14	3	5.11	(4.01-6.51)
Malignancy	8	2	4.91	(3.64-6.62)
OC use	28	11	3.98	(3.44-4.62)
Surgery	16	3	3.48	(2.66-4.55)
Hospitalization	15	3	2.69	(2.01-3.60)
Family history	32	17	2.68	(2.34-3.06)
Bedridden at home	12	4	2.29	(1.81-2.90)
Pregnancy or postpartum	3	1	2.23	(1.50-3.32)
Obesity	21	14	1.83	(1.57-2.13)
HRT	3	3	1.11	(0.77-1.60)
Travel	18	17	1.05	(0.90-1.22)

OR odds ratio; CI confidence interval; OC oral contraceptive; HRT hormone replacement therapy

Table S2. 31 SNP associations with venous thrombosis in LETS

Gene	SNP	Chr	Position	LETS			
				Risk allele frequency, %		OR	95% CI
				cases	Controls		
<i>F5</i>	rs6025	1	167.785.673	11	2	7.19	(4.05-12.76)
<i>F2</i>	rs1799963	11	46.717.631	3	1	2.99	(1.43-6.23)
<i>ABO</i>	rs8176719	9	136.132.908	44	36	1.43	(1.18-1.75)
<i>FGG</i>	rs2066865	4	155.744.726	34	26	1.45	(1.18-1.78)
<i>F11</i>	rs2036914	4	187.429.475	60	54	1.27	(1.05-1.53)
<i>PROCR</i>	rs2069951	20	33.227.425	7	5	1.40	(0.96-2.04)
<i>F11</i>	rs2289252	4	187.444.375	47	43	1.19	(0.99-1.42)
<i>F9</i>	rs4149755	X	138.451.778	7	7	0.96	(0.70-1.31)
<i>PROCR</i>	rs2069952	20	33.227.612	64	61	1.13	(0.94-1.37)
<i>SERPINC1</i>	rs2227589	1	172.152.839	12	9	1.42	(1.04-1.94)
<i>HIVEP1</i>	rs196713	6	11.920.517	23	20	1.02	(0.88-1.18)
<i>F2</i>	rs3136516	11	46.717332	51	50	1.03	(0.86-1.25)
<i>F5</i>	rs1800595	1	167.776.972	5	4	1.38	(0.89-2.16)
<i>PROC</i>	rs1799809	2	127.892.345	47	43	1.19	(0.99-1.43)
<i>PROCR</i>	rs867186	20	33.228.215	15	13	1.22	(0.94-1.60)
<i>VWF</i>	rs1063856	12	6.153.534	35	36	0.96	(0.79-1.17)
<i>GP6</i>	rs1613662	19	60.228.407	85	80	1.36	(1.07-1.74)
<i>F2</i>	rs3136520	11	46.699.808	3	3	1.06	(0.62-1.79)
<i>F8</i>	rs1800291	X	153.811.479	85	82	1.15	(0.93-1.42)
<i>STXBP5</i>	rs1039084	6	147.635.413	44	40	1.19	(0.98-1.43)
<i>NAT8B</i>	rs2001490	2	73.781.606	43	38	1.22	(1.01-1.49)
<i>F13B</i>	rs6003	1	195.297.644	10	8	1.32	(0.96-1.83)
<i>RGS7</i>	rs670659	1	239.228.398	70	64	1.27	(1.04-1.54)
<i>F9</i>	rs6048	X	138.460.946	73	67	1.21	(1.02-1.44)
<i>F5</i>	rs4524	1	167.778.379	80	74	1.36	(1.09-1.69)
<i>F13A1</i>	rs5985	6	6.263.794	79	76	1.19	(0.95-1.49)
<i>F3</i>	1208 indel	1	94.780.000	43	49	0.78	(0.65-0.94)
<i>TFPI</i>	rs8176592	2	188.040.937	67	69	0.89	(0.73-1.10)
<i>F11</i>	rs3822057	4	187.425.146	55	51	1.19	(0.99-1.42)
<i>NR1I2</i>	rs1523127	3	120.983.729	42	33	1.43	(1.19-1.73)
<i>CPB2</i>	rs3742264	13	45.546.095	71	67	1.22	(1.00-1.50)

SNP single nucleotide polymorphism; Chr chromosome; OR odds ratio; CI confidence interval