Multiple SNP testing improves risk prediction of first venous thrombosis


1Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; 2Department of Health Technology & Services Research, MIRA Institute for Biomedical Technology and Technical Medicine, Enschede, The Netherlands; 3Department of Medical Statistics, Leiden University Medical Center, Leiden, The Netherlands; 4Eindhoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, The Netherlands; 5Department of Thrombosis and Hemostasis, Leiden University Medical Center, Leiden, The Netherlands; 6Celera, Alameda, CA; and 7Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands

Introduction

Venous thrombosis is the result of innate thrombotic tendency and nongenetic triggers. Many common genetic variants, mainly single nucleotide polymorphisms (SNPs), with modest effects on risk of venous thrombosis have been reported. Individual SNPs have little predictive value because of 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 persons 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.

Genetic profiling has become a popular aim in epidemiologic studies of many common diseases because a large amount of data from genome-wide association studies has become available. For recurrent venous thrombosis, we previously investigated the potential clinical utility of multiple SNP testing for recurrent events. 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, persons undergoing surgery, persons requiring a plaster cast, or those subject to prolonged immobilization.

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. The MEGA analysis included 2712 patients and 4634 controls (Multiple Environmental and Genetic Assessment). 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% confidence interval [CI], 0.25-0.53) for persons with 0 risk alleles to 7.48 (95% CI, 4.49-12.46) for persons with more than or equal to 6 risk alleles. The AUC of a risk model based on known nongenetic risk factors was 0.77 (95% CI, 0.76-0.78). Combining the nongenetic and genetic risk models improved the AUC to 0.82 (95% CI, 0.81-0.83), indicating good diagnostic accuracy. To become clinically useful, subgroups of high-risk persons must be identified in whom genetic profiling will also be cost-effective. (Blood. 2012;120(3):656-663)
SNPs were added from the Group Health study\textsuperscript{13,14}; these SNPs were obtained from the FARIVE study and the MEGA study.\textsuperscript{17} Among the 40 SNPs in the initial follow-up study of a genome-wide association study and replicated in the literature,\textsuperscript{1,13} rs2069552 \textsuperscript{12} was explained by rs1799809 in the Factor V gene (rs6050 and rs2070006 in F5; and rs3753305 in F3) were 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.

### 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.\textsuperscript{1,13} Twelve SNPs were added from the Group Health study,\textsuperscript{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.\textsuperscript{15,16} Another added SNP was recently identified in a follow-up study of a genome-wide association study and replicated in the FARIVE study and the MEGA study.\textsuperscript{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, rs1800738 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 persons. 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 the 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 ratios (ORs) in the literature and assessed whether adding SNPs to the risk score improved the area under the curve (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.

### Nongenetic risk factors

We constructed a nongenetic risk score, which included the following risk factors: recent (within 3 months before the index date) leg injury, surgery, pregnancy or postpartum, immobilization (ie, plaster cast, bedridden at home, hospitalization), travel for more than 4 hours in 24 hours, cancer diagnosis between 5 years before and 6 months after the index date, and a cancer diagnosis between 5 years before and 6 months after the index date. The index date was defined as date of first diagnosis for patients and their partner controls and the date of completing the questionnaire for random controls. We also included family history in the nongenetic risk score. Family history was defined as positive when a parent or sibling had experienced venous thrombosis, or when the participant was not aware of venous thrombosis in the family.
corresponding to the logarithm of the risk estimates in MEGA (supplementary Table 1, available on the Blood Web site; see the Supplemental Materials link at the top of the online article) 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 nongenetic risk score using a logistic regression model.

Application of genetic profiling may be most useful in high-risk groups (ie, persons exposed to known nongenetic risk factors). We therefore studied the discriminative accuracy of our genetic risk score as well as the combined score in high-risk situations of surgery, plaster cast, hospitalization, young women (younger than 50 years) using OCs, women using HRT, pregnancy or postpartum, middle-aged persons (older than 50 years) and travel. We also studied persons with a positive family history and persons with malignant disorders.

**Statistical analyses**

Crude and sex-adjusted (in case SNPs were located on the X chromosome) ORs and 95% CIs were calculated by logistic regression for individual SNPs and the genetic, nongenetic, 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 nongenetic risk models according to the method of Hanley and McNeil.\(^{18}\) Nagelkerke 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 Version 17.0.2 for Windows (SPSS Inc).

**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 the 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.

**Genetic risk score**

We first included all 31 SNPs in the genetic risk score. For each person, 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 among control subjects (24 risk alleles; bottom panel). Persons with 15 or less and 36 or more risk alleles were combined for the calculation of the OR because of the low numbers of persons with that few or many risk alleles (bottom panel).
in the 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; ie, 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 nonweighted 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 pseudo-$r^2$; Table 2).

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 ABO 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 3 most well-known prothrombotic polymorphisms (ie, 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 variable, was 9.50 (95% CI, 7.92-11.39). The 5-SNP risk score explained 13.5% of the total variability (Nagelkerke pseudo-$r^2$; Table 2).

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 nonweighted model based on number of risk alleles (AUC = 0.66, 95% CI, 0.64-0.67).

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 pulmonary embolism. 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 pulmonary embolism (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 nongenetic 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). Subanalysis in cancer patients according to therapy (chemotherapy, surgery, radiation) or tumor class (solid vs 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 persons (Table 3).

### Table 2. Venous thrombosis prediction using genetic, nongenetic, and combined risk scores

<table>
<thead>
<tr>
<th></th>
<th>MEGA (N = 7092)</th>
<th>LETS (N = 881)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC (95% CI)</td>
<td>Nagelkerke pseudo $r^2$</td>
</tr>
<tr>
<td>31-SNP risk score</td>
<td>0.71 (0.69-0.72)</td>
<td>0.161</td>
</tr>
<tr>
<td>5-SNP risk score</td>
<td>0.69 (0.67-0.70)</td>
<td>0.135</td>
</tr>
<tr>
<td>Nongenetic risk score</td>
<td>0.77 (0.76-0.78)</td>
<td>0.288</td>
</tr>
<tr>
<td>Combined risk score</td>
<td>0.82 (0.81-0.83)</td>
<td>0.378</td>
</tr>
</tbody>
</table>

The LETS study was used as a validation set.
we added the genetic risk score to the nongenetic score, the AUC significantly increased to 0.82 (Figure 4: 95% CI, 0.81-0.83) compared with the nongenetic risk score alone (P < .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 nongenetic risk score, which significantly improved to 37.8% (Nagelkerke pseudo $r^2$; Table 2) when combining the nongenetic and genetic risk scores. Both the nongenetic and the combined risk score models performed better in women than in men (nongenetic 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 nongenetic risk score, which was significant for persons using OCs, persons with a positive family history of venous thrombosis, and persons older than 50 years old (Table 3).

### Validation of the risk scores

To validate the genetic, nongenetic, and combined risk scores, we studied their discriminative accuracy in subjects from another population, the LETS population. As described in “Study populations,” 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 nongenetic 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 nongenetic risk factors (ie, 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 nongenetic risk score. The discriminative accuracy of the nongenetic 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).

### Combining nongenetic and genetic risk scores

We assessed the discriminative accuracy of a nongenetic risk score based on known nongenetic 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 nongenetic risk score, which was significant for persons using OCs, persons with a positive family history of venous thrombosis, and persons older than 50 years old (Table 3).

To validate the genetic, nongenetic, and combined risk scores, we studied their discriminative accuracy in subjects from another population, the LETS population. As described in “Study populations,” 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 nongenetic 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 nongenetic risk factors (ie, 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 nongenetic risk score. The discriminative accuracy of the nongenetic 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).

### Table 3. Risk score prediction in subgroups of persons exposed to known nongenetic risk factors

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

NA indicates not applicable.

*Women younger than 50 years.
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 after 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, indiscrimination may lead to an increased risk of thrombosis in high-risk persons receiving insufficient prophylactic anticoagulant treatment, whereas persons 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 ROC curves. The 5-SNP genetic risk score performed better than family history assessment, which is the current clinical practice of risk assessment in persons exposed to known nongenetic risk factors. However, the 5-SNP genetic risk score performed worse than a risk score of nongenetic risk factors. A recent study by Hippisley-Cox and Coupland showed that an algorithm of nongenetic 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 nongenetic risk score (0.77). However, the AUC may be an overestimation because 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 nongenetic risk score model significantly improved the AUC to 0.82, indicating good diagnostic accuracy. In our validation study, information on the nongenetic risk factors was less complete, which explains the lower discriminative accuracy of both the nongenetic risk score (0.71) and the combined risk score (0.77).

Identification of persons 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 persons a year) to justify genotyping of all persons. In all subgroups of high-risk persons, the combined risk score performed better than the nongenetic score alone, which may indicate the potential clinical value of genetic profiling in these high-risk persons.

We defined a basic genetic risk score that counts the total number of risk-increasing alleles in persons. 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 nonweighted 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 are not taken into account, although in reality many interactions exist. Examples for venous thrombosis are the synergistic effects between factor V Leiden (rs6025) and OC use and between the F13A1 Val34Leu variant (rs5985) and fibrinogen levels.

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 a person 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 nonweighted 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 al 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 approximately 60%. 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 conclusion, we demonstrated that addition of a 5-SNP risk score to a risk scoring system based on nongenetic 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 persons.

Acknowledgments

The Leiden Thrombophilia Study was supported by The Netherlands Heart Foundation (grant 98.113), the Dutch Cancer Foundation (RUL99/1992), and The Netherlands Organization for Scientific Research (grant 912-03-0331 2003).

Authorship

Contribution: C.Y.V. had full access to all of the data in the study, takes full responsibility for the integrity of the data and the accuracy of the data analysis, performed statistical analysis, and supervised the study; F.R.R. had full access to all of the data in the study, takes full responsibility for the integrity of the data and the accuracy of the data analysis, conceived and designed the study, acquired the data, performed statistical analysis, and supervised the study; L.A.B. conceived and designed the study, acquired the data, drafted the manuscript, and performed statistical analysis; J.J.D. conceived, designed, provided administrative, technical, or material support, and supervised the study; P.H.R. conceived, designed, and supervised the study; A.R.A. acquired the data and provided administrative, technical, or material support; H.G.d.H. drafted the manuscript, performed statistical analysis; S.L.C. performed statistical analysis; C.J.M.D. conceived and designed the study, acquired the data, provided administrative, technical, or material support and supervised the study; C.H.T. provided administrative, technical, or material support; L.A.B. provided administrative, technical, or material support and supervised the study; and all authors analyzed and interpreted the data and critically revised the manuscript for important intellectual content.

Conflict-of-interest disclosure: L.A.B., J.J.D., P.H.R., I.D.B., and F.R.R. hold or have applied for patents related to SNPs in this manuscript (notably rs6025, rs2066865, and rs2036914). A.R.A., C.H.T., J.J.D., and L.A.B. are employees of Celera USA. The remaining authors declare no competing financial interests.

Correspondence: Frits R. Rosendaal, Department of Clinical Epidemiology, C7-P, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, The Netherlands; e-mail: f.r.rosendaal@lumc.nl.

References


Multiple SNP testing improves risk prediction of first venous thrombosis

Hugoline G. de Haan, Irene D. Bezemer, Carine J. M. Doggen, Saskia Le Cessie, Pieter H. Reitsma, Andre R. Arellano, Carmen H. Tong, James J. Devlin, Lance A. Bare, Frits R. Rosendaal and Carla Y. Vossen