

Supplementary Appendix

Prediction of Venous Thromboembolism Based on Clinical and Genetic Factors

David A. Kolin, M.Sc.*; Scott Kulm, B.Eng.*; Olivier Elemento, Ph.D.

*Mr. Kolin and Kulm contributed equally to this article.

Address for Correspondence:

Olivier Elemento, Ph.D.
Caryl and Israel Englander Institute for Precision Medicine
The Meyer Cancer Center
Weill Cornell Medicine
1300 York Avenue
New York, NY 10065
ole2001@med.cornell.edu

Table of Contents:

- Page 3 Table S1.** Definition of Venous Thromboembolism
- Page 4 Table S2.** Single Nucleotide Polymorphisms Used to Calculate the Polygenic Risk Score
- Page 5 Table S3.** Clinical Risk Score for Venous Thromboembolism Based on Established Risk Factors
- Page 6 Table S4.** Risk Group Definitions
- Page 7 Table S5.** Characteristics of the Participants at Baseline
- Page 8 Table S6.** Risk of Venous Thromboembolism with Duration of Contraception Use
- Page 9 Table S7.** The Five Most Common Oral Contraceptive Pills Amongst Current Users
- Page 10 Table S8.** Univariate Risk Ratios for Death from Venous Thromboembolism
- Page 11 Table S9.** Risk of Venous Thromboembolism with Common Non-Cancer Illnesses, Adjusted for Common Medications
- Page 12 Figure S1.** Kaplan-Meier Curves for Four Individual Clinical Factors
- Page 13 Figure S2.** Minimally Adjusted Risk of Venous Thromboembolism with Established Clinical Risk Factors
- Page 14 Figure S3.** Minimally Adjusted Risk of Venous Thromboembolism with Common Cancer Subtypes
- Page 15 Figure S4.** Minimally Adjusted Risk of Venous Thromboembolism with Common Medications
- Page 16 Figure S5.** Minimally Adjusted Risk of Venous Thromboembolism with Common Non-Cancer Illnesses
- Page 17 Figure S6.** Minimally Adjusted Risk of Venous Thromboembolism with Common Fracture Sites
- Page 18 Figure S7.** Number of Incident Venous Thromboembolic Events Amongst Participants
- Page 19 Figure S8.** Polygenic Risk Score Stratified by the Number of Venous Thromboembolic Events
- Page 20 Figure S9.** Hazard Ratios for Sub-Classified Venous Thromboembolic Events with the Clinical and Genetic Scores
- Page 21 Figure S10.** Interactions Between Individual Covariates and the Genetic Score
- Page 22 Figure S11.** Associations Between Traits Relevant to Venous Thromboembolism, Polygenic Risk Scores of These Traits, or Interactions Between the Two
- Page 23 Figure S12.** Associations Derived from a Cox-proportional Hazard Model for Possible Frailty within the Population
- Page 24 Figure S13.** Hazard Ratios for Venous Thromboembolism at Various Time Points
- Page 25 Figure S14.** Kaplan-Meier Curve of the Combined Score

Table S1. Definition of Venous Thromboembolism

Outcome	ICD-10	ICD-9
Venous thromboembolism		
Pulmonary embolism (PE)	I26	415
Deep vein thrombosis (DVT)	I80-I82	451-453

Table S2. Single Nucleotide Polymorphisms Used to Calculate the Polygenic Risk Score

Chromosome	Position	P-value	Risk Estimate	Risk Allele	Lead SNP	Study	Study Coefficient
1	169165889	2e-12	2.02	C	rs16861990	Heit et al., 2012	0.68
1	169467654	1.2e-08	2.38	A	rs1894692	de Haan et al., 2018	-0.13
1	169481223	7.9e-09	2.06	G	rs6427196	de Haan et al., 2018	-0.13
1	169490392	9.1e-09	2.05	C	rs2420370	de Haan et al., 2018	-0.13
1	169491555	1.2e-08	2.03	A	rs2420371	de Haan et al., 2018	-0.13
1	169542517	3e-11	1.2	T	rs4524	Germain et al., 2015	-0.85
1	169544768	2e-26	2.52	A	rs1018827	Germain et al., 2011	0.05
1	169549811	4e-137	2.93	T	rs6025	Hinds et al., 2016	0.17
2	42372465	2e-07	1.36	G	rs72798544	Hinds et al., 2016	0.17
3	1004354	2e-06	1.18	C	rs6764623	Tang et al., 2013	0.07
4	154599778	2e-13	1.53	A	rs7659024	Germain et al., 2011	0.05
4	154623920	2e-19	1.22	T	rs7654093	Hinds et al., 2016	0.17
4	186266940	3e-14	1.24	G	rs4253399	Tang et al., 2013	0.07
4	186277851	1e-23	1.27	C	rs4253417	Germain et al., 2015	-0.85
9	133257522	6e-12	1.47	G	rs8176719	Heit et al., 2012	0.68
9	133279294	3e-16	1.65	T	rs495828	Heit et al., 2012	0.68
10	69458890	3e-07	1.165	G	rs17490626	Hinds et al., 2016	0.17
10	69485520	2e-16	1.31	T	rs78707713	Germain et al., 2015	-0.85
15	31501727	3e-06	1.15	A	rs7164569	Tang et al., 2013	0.07
19	10631494	3e-15	1.21	G	rs2288904	Germain et al., 2015	-0.85
19	10632450	6e-09	1.145	T	rs9797861	Hinds et al., 2016	0.17

Table S3. Clinical Risk Score for Venous Thromboembolism Based on Established Risk Factors

Clinical Risk Factors	Points (Total = 10)
Sex	Female – 0; Male – 1
Age, years	<50 – 0; ≥50 to <60 – 1; ≥60 – 2
Body mass index, kg/m²	<25 – 0; ≥25 to <30 – 1; ≥30 – 2
Cancer	No – 0; Yes – 1
Smoking	Not current – 0; Current – 1
Fracture in the last five years	No – 0; Yes – 1
Previous venous thromboembolism	No – 0; Yes – 2

Previous venous thromboembolism was defined as previous deep vein thrombosis or previous pulmonary embolism.

Table S4. Risk Group Definitions

Risk Group	Genetic Score (%)	Clinical Score (%)
Low	-3.59 to -0.80 (19.9)	0 to 2 (38.1)
Medium	-0.80 to 0.74 (80.0)	3 to 4 (49.3)
High	0.74 to 6.19 (20.1)	5 to 10 (12.6)

Table S5. Characteristics of the Participants at Baseline

Variable	Total (N = 502,536)
Sex – no. (%)	
Female	273,402 (54.4)
Male	229,134 (45.6)
Age, years – mean (95% CI)	56.53 (56.51-56.55)
Body mass index, kg/m² – mean (95% CI)	27.43 (27.42-27.45)
Previous cancer diagnosis – no. (%)	
No	459,974 (91.5)
Yes	41,700 (8.3)
Unknown	862 (0.2)
Smoking status – no. (%)	
Never	273,537 (54.4)
Previous	173,030 (34.4)
Current	52,979 (10.5)
Declined to answer	2059 (0.4)
Unknown	891 (0.2)
Alcohol intake frequency – no. (%)	
Never	40,648 (8.1)
Special occasions only	58,012 (11.5)
One to three times a month	55,858 (11.1)
Once or twice a week	129,297 (25.7)
Three or four times a week	115,445 (23.0)
Daily or almost daily	101,774 (20.3)
Prefer not to answer	605 (0.1)
Unknown	897 (0.2)
Ever use oral contraceptive pill – no. (%)	
No	51,532 (18.8)
Yes	220,446 (80.6)
Do not know	444 (0.2)
Prefer not to answer	497 (0.2)
Unknown	483 (0.2)
Ever use hormone replacement therapy – no. (%)	
No	167,903 (61.4)
Yes	103,919 (38.0)
Do not know	800 (0.3)
Prefer not to answer	297 (0.1)
Unknown	483 (0.2)
Fracture in the last 5 years – no. (%)	
No	451,266 (90.0)
Yes	47,466 (9.4)
Do not know	2,533 (0.5)
Declined to answer	340 (<0.1)
Unknown	931 (0.2)

Table S6. Risk of Venous Thromboembolism with Duration of Contraception Use

Duration of Contraception Use (Years)	No. of Patients	No. of Events	Hazard Ratio (95% CI)	P-value
Other	77,900	820	1	0.29
>0 to <10	99,601	779	0.93 (0.83-1.04)	
≥10 to <20	66,377	408	0.87 (0.76-1.00)	
≥20	29,524	151	0.81 (0.67-0.98)	

Cox regression models were adjusted for sex, age, body mass index, previous cancer diagnosis, smoking status, alcohol intake frequency, use of hormone replacement therapy, fracture in the last five years, previous deep vein thrombosis, previous pulmonary embolism, and the first four principal components of ancestry. The p-value was calculated with duration of contraception use coded as a continuous variable.

Table S7. The Five Most Common Oral Contraceptive Pills Amongst Current Users

Oral Contraceptive Pill	Total (N = 2,598)
Cerazette (75 µg tablet)	935
Micronor tablet	558
Microgynon 30 tablet	518
Femulen tablet	305
Noriday tablet	282

Table S8. Univariate Risk Ratios for Death from Venous Thromboembolism

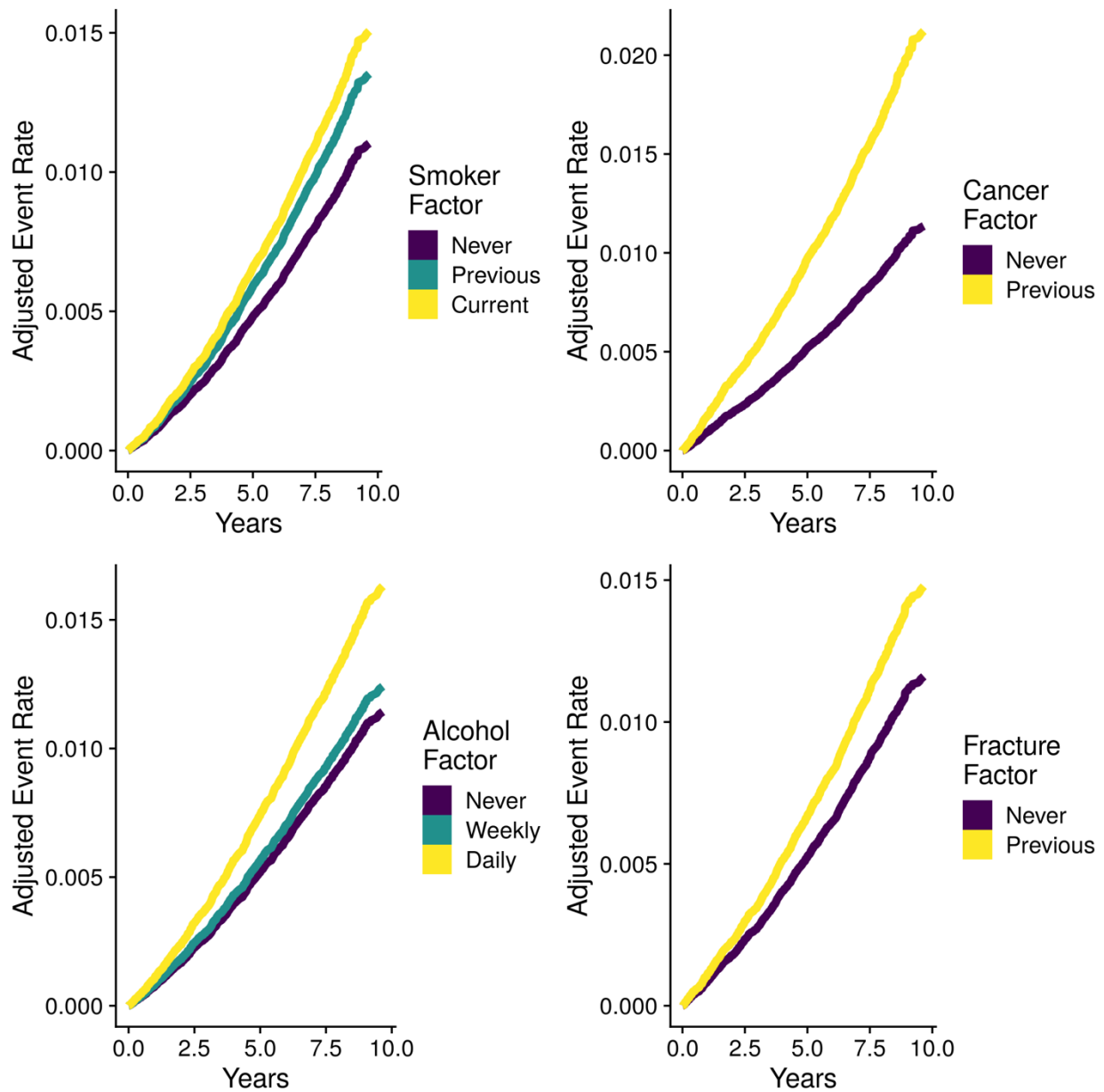
Variable	No. of Patients	No. of Events	Risk Ratio (95% CI)
Sex			
Female	273,402	102	1
Male	229,134	133	1.56 (1.20-2.01)
Age, years			
<50	117,885	21	1
≥50 to <60	167,158	60	2.01 (1.23-3.31)
≥60	217,493	154	3.97 (2.52-6.27)
Body-mass index, kg/m²			
<18.5	2,626	3	5.46 (1.68-17.76)
≥18.5 to <25	162,423	34	1
≥25 to <30	212,127	79	1.78 (1.19-2.66)
≥30 to <35	87,557	60	3.27 (2.15-4.99)
≥35 to <40	24,994	26	4.97 (2.98-8.28)
≥40	9,704	28	13.78 (8.36-22.72)
Previous cancer diagnosis			
No	459,974	211	1
Yes	41,700	23	1.20 (0.78-1.85)
Ever smoked			
Never	273,537	113	1
Previous	173,070	80	1.12 (0.84-1.49)
Current	52,979	38	1.74 (1.20-2.51)
Alcohol intake frequency			
Never	40,648	26	1
Special occasions only	58,012	47	1.27 (0.78-2.04)
One to three times a month	55,858	28	0.78 (0.46-1.34)
Once or twice a week	129,297	53	0.64 (0.40-1.02)
Three or four times a week	115,445	43	0.58 (0.36-0.95)
Daily or almost daily	101,774	37	0.57 (0.34-0.94)
Ever use oral contraceptive pill			
No	51,532	31	1
Yes	220,446	69	0.52 (0.34-0.79)
Ever use hormone-replacement therapy			
No	167,903	52	1
Yes	103,919	49	1.52 (1.03-2.25)
Fracture in the last 5 years			
No	451,266	214	1
Yes	47,466	20	0.89 (0.56-1.40)
Previous deep veins thrombosis			
No	492,582	219	1
Yes	9,323	16	3.86 (2.32-6.41)
Previous pulmonary embolism			
No	498,417	221	1
Yes	3,955	14	7.98 (4.66-13.69)

Table S9. Risk of Venous Thromboembolism with Common Non-Cancer Illnesses, Adjusted for Common Medications

Common Non-Cancer Illness	Additional Covariates	Hazard Ratio (95% CI)
Hypercholesterolemia	Atorvastatin, simvastatin, and pravastatin	1.00 (0.90-1.11)
Asthma	Beclomethasone, beconase, and becotide	1.14 (1.04-1.25)
Osteoarthritis	Aspirin, ibuprofen, and diclofenac	1.17 (1.06-1.29)
Depression	Amitriptyline, citalopram, and fluoxetine	1.19 (1.03-1.37)
Diabetes	Metformin, insulin products, and glibenclamide	0.86 (0.71-1.04)

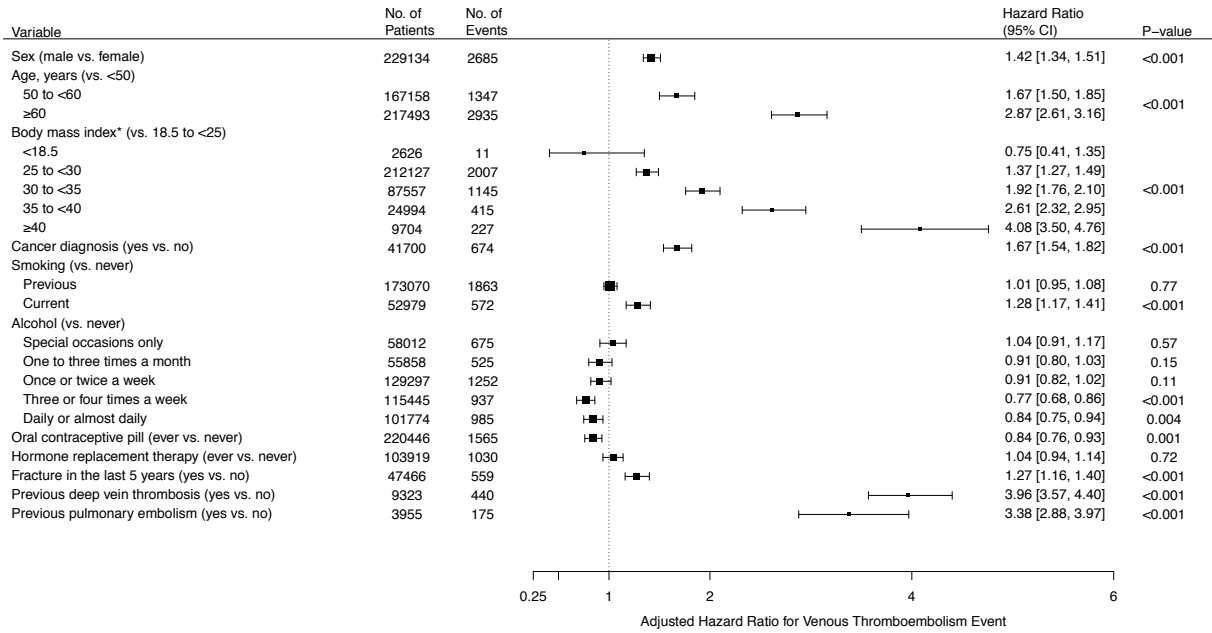
Cox regression models were adjusted for sex, age, body mass index, previous cancer diagnosis, smoking status, alcohol intake frequency, use of oral contraceptive pills, use of hormone replacement therapy, fracture in the last five years, previous deep vein thrombosis, previous pulmonary embolism, and the first four principal components of ancestry. Common illness-specific medications were added to each model to test mediation by medication use.

Figure S1. Kaplan-Meier Curves for Four Individual Clinical Factors



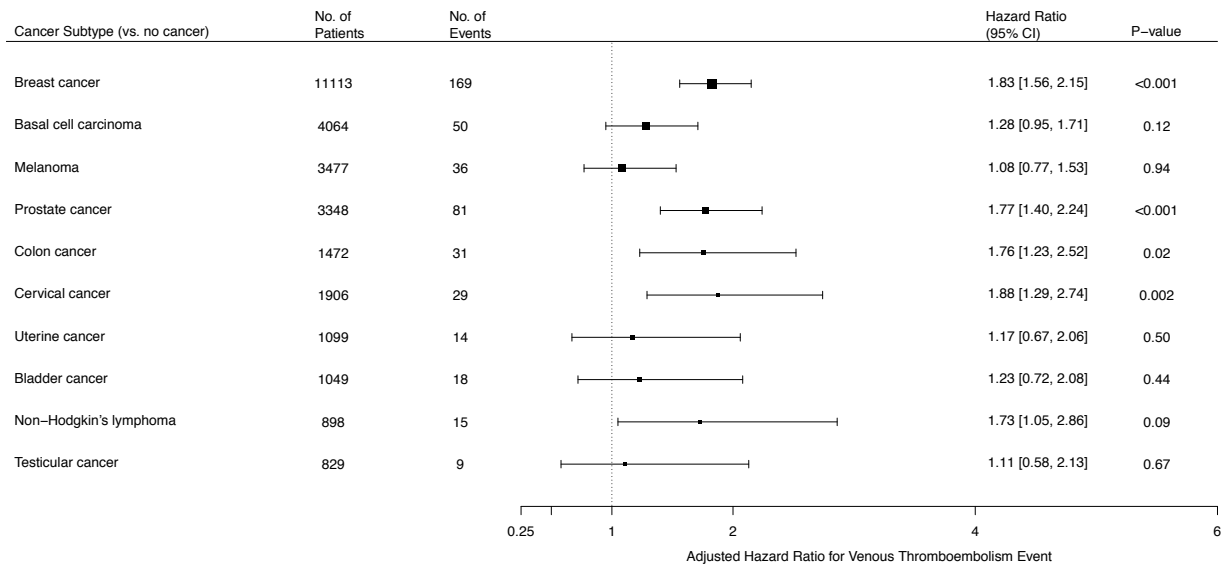
The Kaplan-Meier curves were adjusted for the first four principal components of ancestry, age, sex, body mass index, previous venous thromboembolism, smoking status, previous cancer diagnosis, alcohol intake frequency, and fracture in the last five years.

Figure S2. Minimally Adjusted Risk of Venous Thromboembolism with Established Clinical Risk Factors



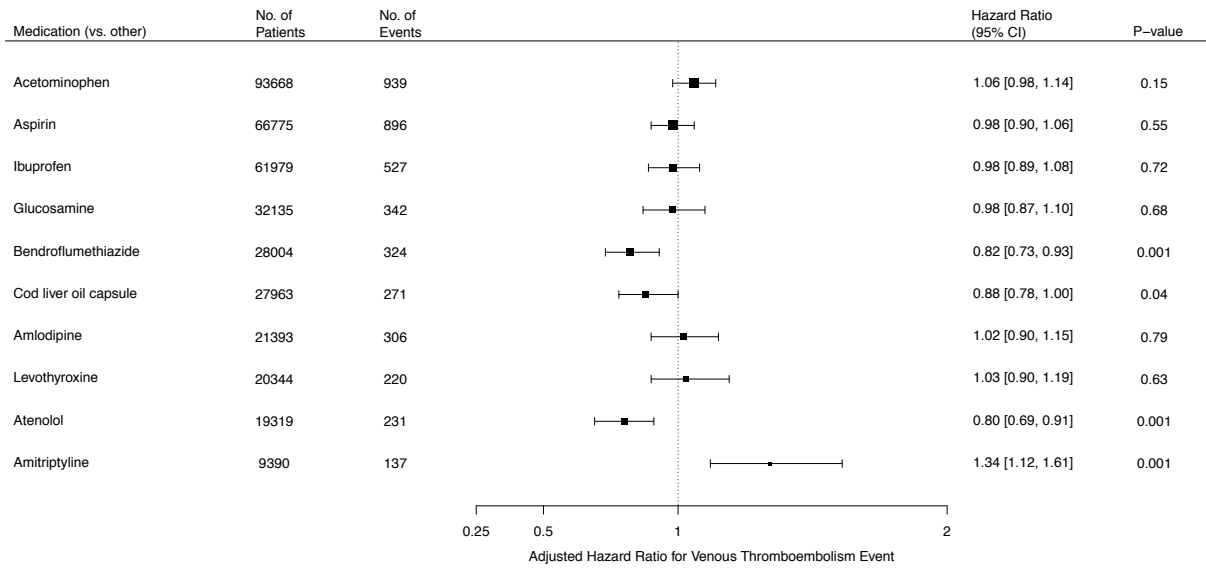
Minimally adjusted Cox regression models were adjusted for sex, age, body mass index, and smoking status.

Figure S3. Minimally Adjusted Risk of Venous Thromboembolism with Common Cancer Subtypes



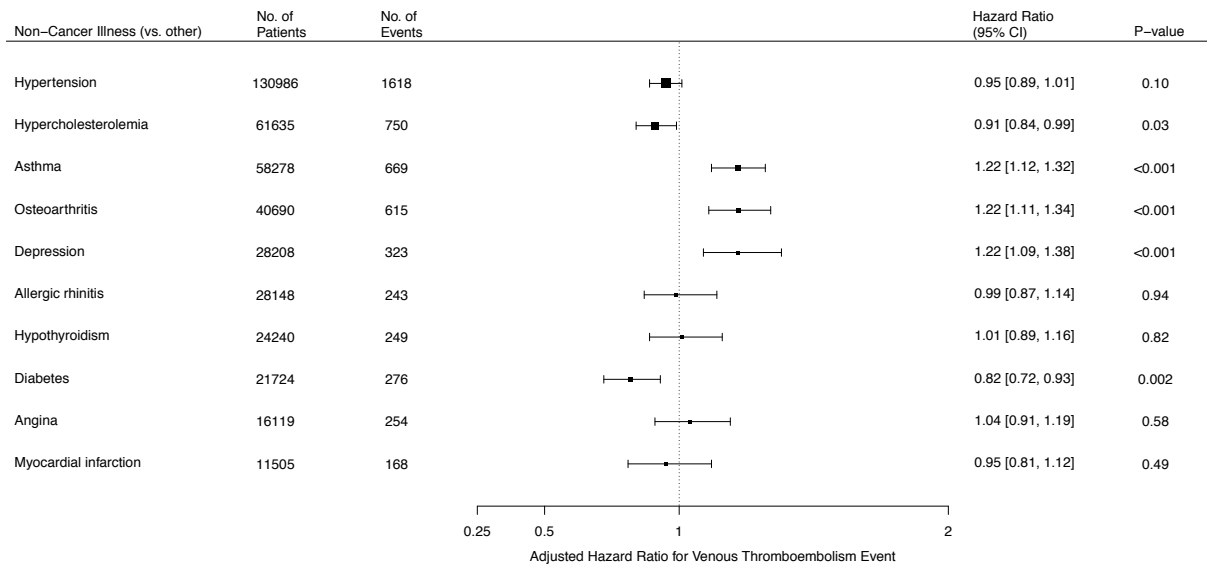
Minimally adjusted Cox regression models were adjusted for sex, age, body mass index, and smoking status.

Figure S4. Minimally Adjusted Risk of Venous Thromboembolism with Common Medications



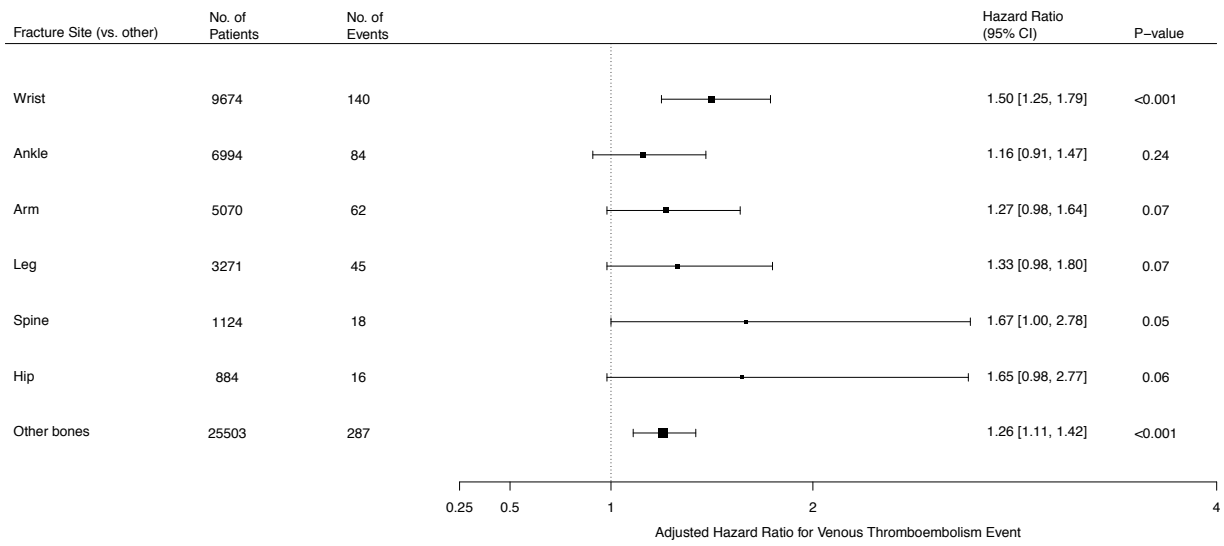
Minimally adjusted Cox regression models were adjusted for sex, age, body mass index, and smoking status.

Figure S5. Minimally Adjusted Risk of Venous Thromboembolism with Common Non-Cancer Illnesses



Minimally adjusted Cox regression models were adjusted for sex, age, body mass index, and smoking status.

Figure S6. Minimally Adjusted Risk of Venous Thromboembolism with Common Fracture Sites



Minimally adjusted Cox regression models were adjusted for sex, age, body mass index, and smoking status.

Figure S7. Number of Incident Venous Thromboembolic Events Amongst Participants

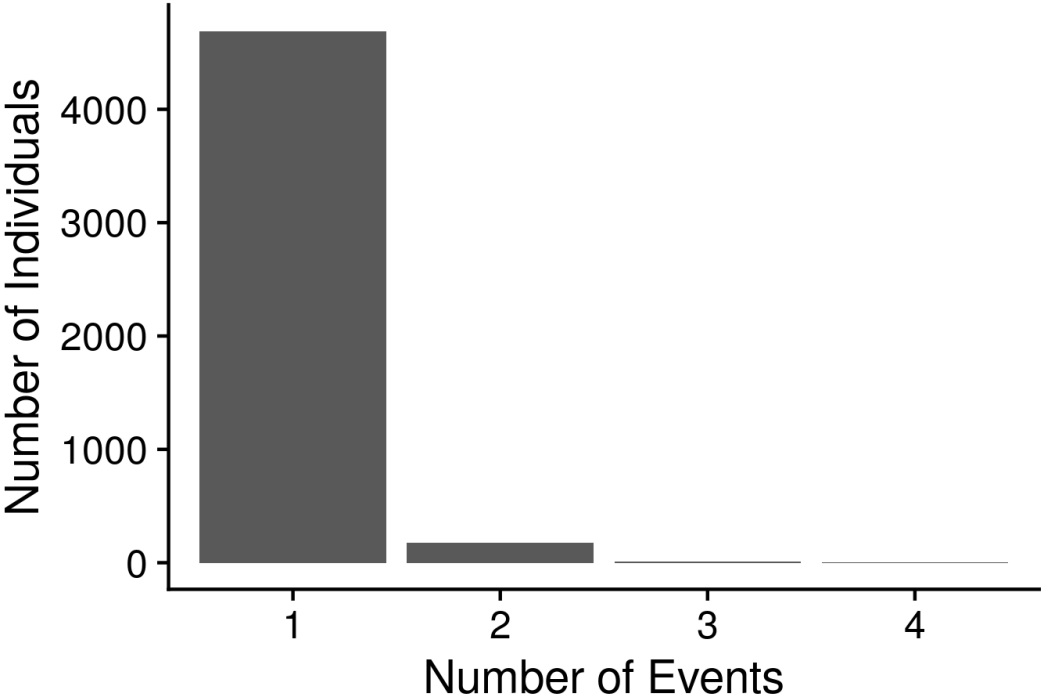


Figure S8. Polygenic Risk Score Stratified by the Number of Venous Thromboembolic Events

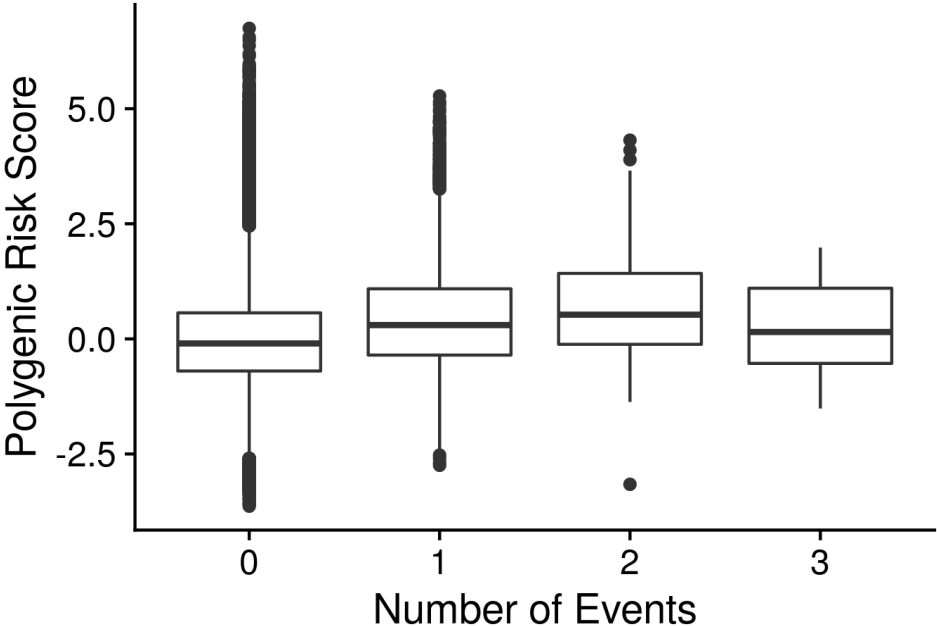
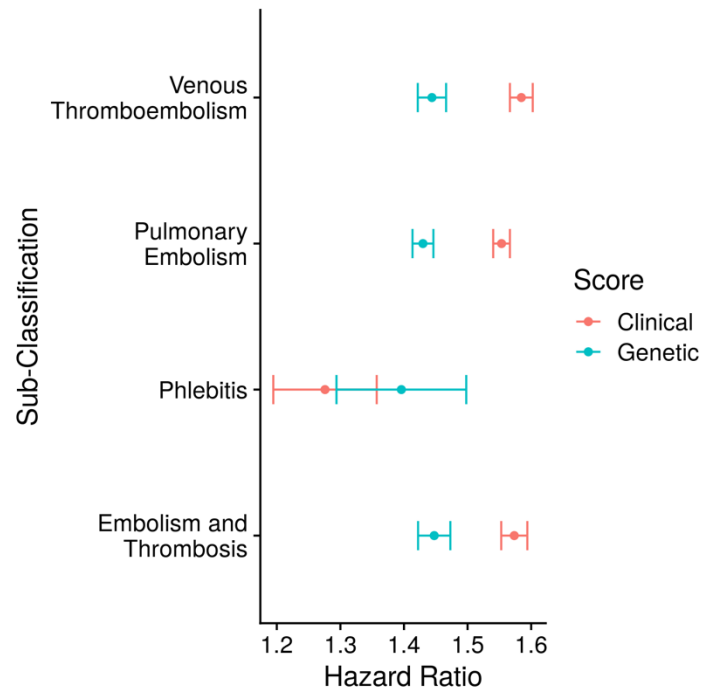
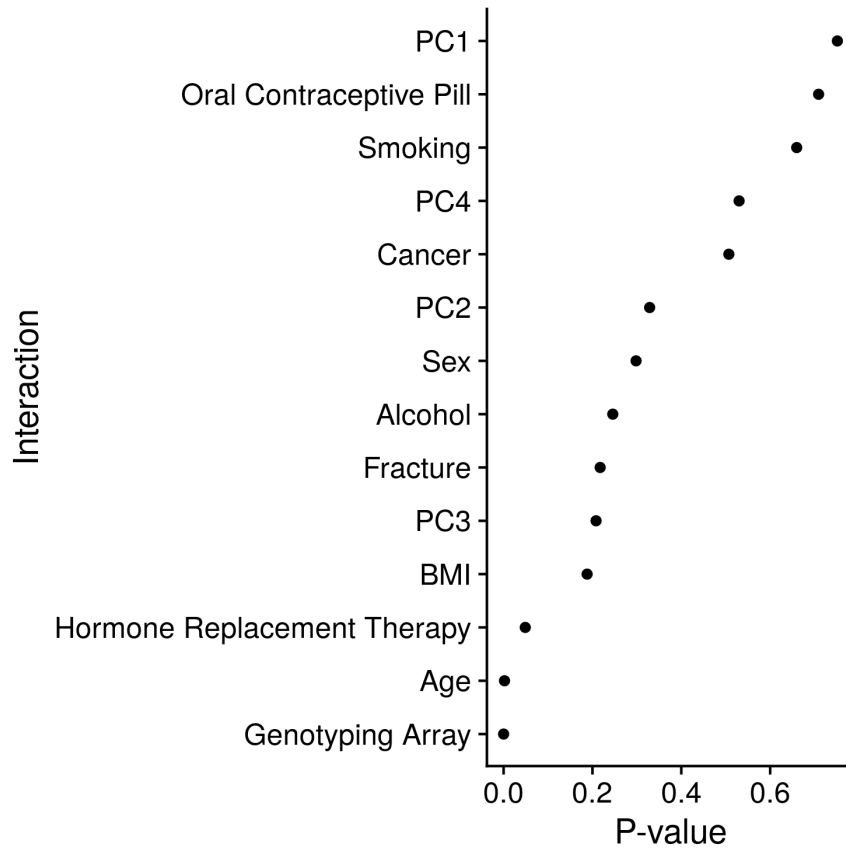


Figure S9. Hazard Ratios for Sub-Classified Venous Thromboembolic Events with the Clinical and Genetic Scores



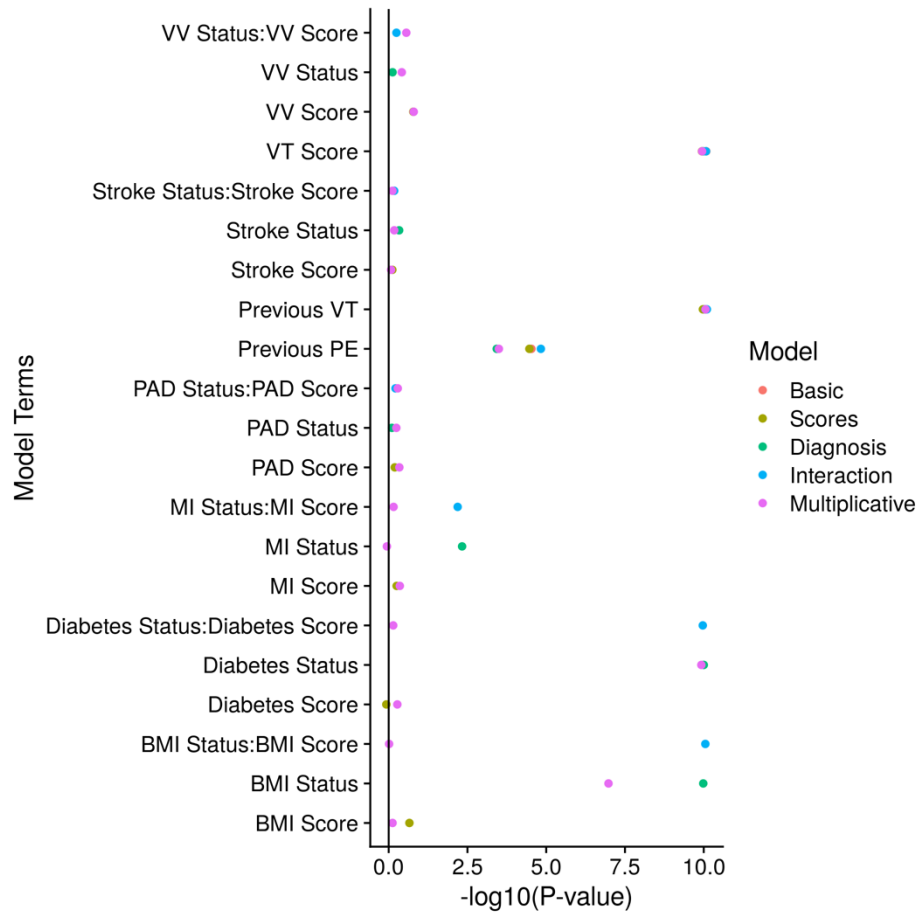
Event sub-classifications were coded as follows: pulmonary embolism (I26), phlebitis (I80), and embolism and thrombosis (I82). Hazard ratios were calculated using Cox-proportional hazard models including the factor indicated and adjusted for the first four principal components of ancestry. Hazard ratios were calculated with the clinical and genetic scores coded on a continuous scale, with a range from 0 to 9 for the clinical score and -3.59 to 6.19 for the genetic score.

Figure S10. Interactions Between Individual Covariates and the Genetic Score



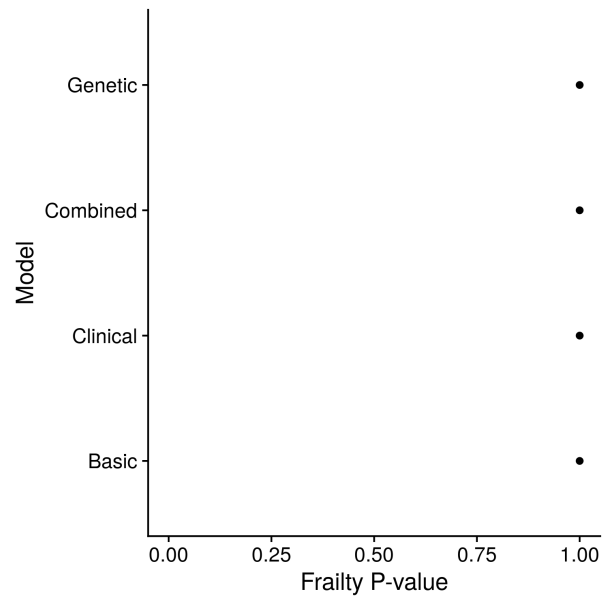
Interactions were tested between the individual covariates indicated and the genetic score. The p-values for interactions were computed using a Cox-proportional hazard model adjusted for the first four principal components of ancestry (PC1 to PC4), age, sex, previous venous thromboembolism, body mass index (BMI), previous cancer diagnosis, smoking status, alcohol intake frequency, and fracture in the last five years.

Figure S11. Associations Between Traits Relevant to Venous Thromboembolism, Polygenic Risk Scores of These Traits, or Interactions Between the Two



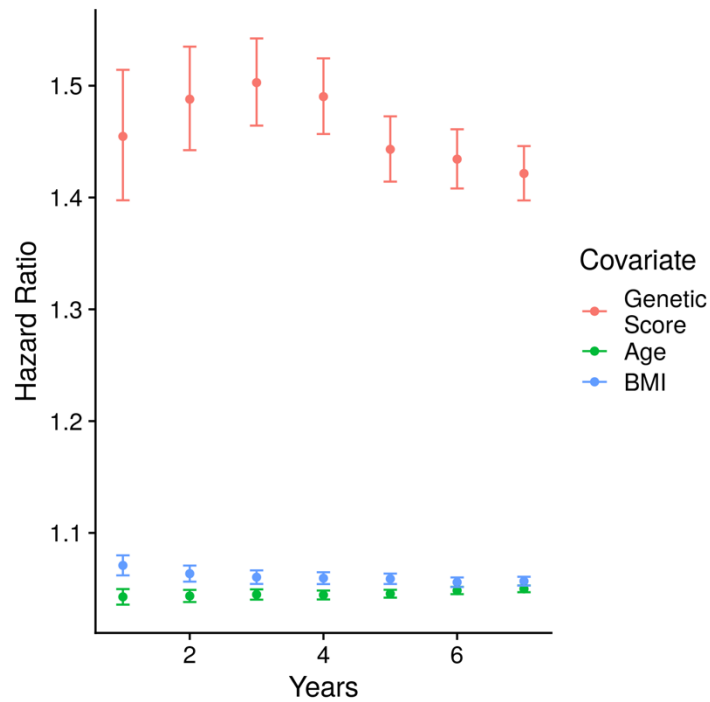
The relevance of traits associated with venous thromboembolism was analyzed through both genetics, via a polygenic risk score calculated to best predict the relevant trait, and outcomes of the relevant trait determined through related ICD codes. Specifically, the associated traits included venous thromboembolism (VT), pulmonary embolism (PE), body mass index (BMI), diabetes, myocardial Infarction (MI), pulmonary artery disease (PAD), stroke, and varicose veins (VV). Five Cox-proportional hazard models were constructed for this analysis: basic, which included principal components, age, sex, and previous venous thromboembolism; scores, which included the basic model with the polygenic risk scores for all traits; diagnosis, which is the basic model with the binary outcome for each trait; interaction, which is the basic model with the interaction terms between each trait's outcome and score; and multiplicative, which is the basic model, trait scores, trait outcomes, and the interaction terms between the trait outcomes and scores. The negative \log_{10} p-value for each covariate in these models is reported, with values above 10 rounded to 10.

Figure S12. Associations Derived from a Cox-proportional Hazard Model for Possible Frailty within the Population



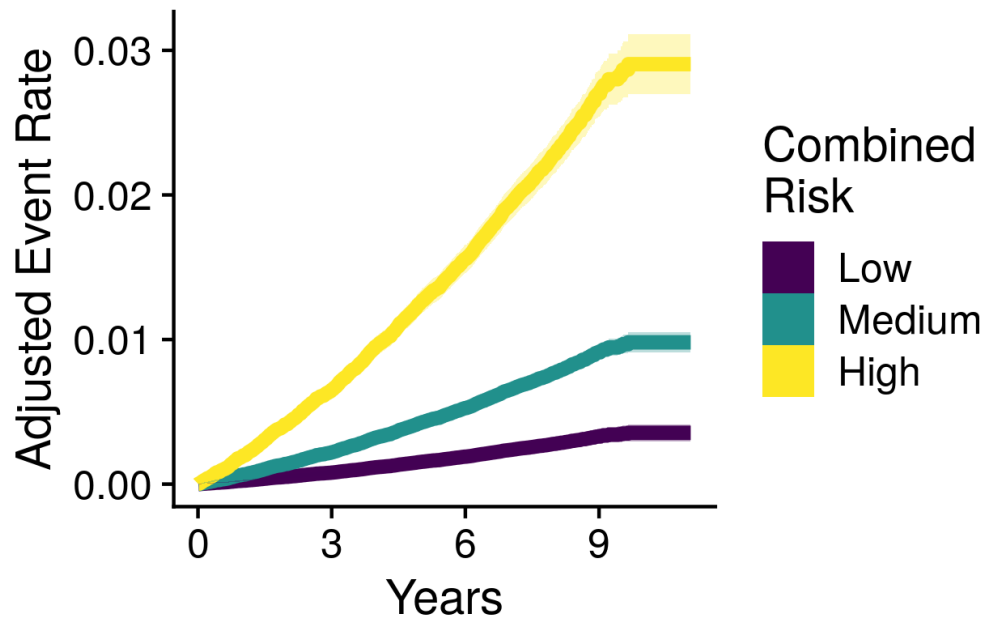
A frailty analysis was conducted to check for additional clustering among the participants. Four Cox-proportional hazard models were computed. The genetic model included the first four principal components of ancestry, age, sex, previous venous thromboembolism and the genetic score. The combined model included the first four principal components of ancestry and the combined score. The clinical model included the first four principal components of ancestry and the clinical score. The basic model included the first four principal components of ancestry, age, sex, and previous venous thromboembolism.

Figure S13. Hazard Ratios for Venous Thromboembolism at Various Time Points



A time dependent analysis was conducted to check for any strong temporal relationships between three covariates and venous thromboembolic events. Three of the covariates (genetic score, age, and body mass index [BMI]) were chosen due to their strong association with venous thromboembolism. The maximum study length data frame of events was truncated at each year. The Cox-proportional hazard model was adjusted for the first four principal components of ancestry, age, sex, BMI, previous venous thromboembolism, previous cancer diagnosis, alcohol intake frequency, smoking status, the genetic score, and fracture in the last five years. Hazard ratios were calculated with each covariate of interest coded on a continuous scale.

Figure S14. Kaplan-Meier Curve of the Combined Score



The Kaplan-Meier curve for the combined score was adjusted for the first four principal components of ancestry.