JAMA | Original Investigation

Association of Supernumerary Sex Chromosome Aneuploidies With Venous Thromboembolism

Alexander S. F. Berry, PhD; Brenda M. Finucane, MS; Scott M. Myers, MD; Angela Abril, MD; H. Lester Kirchner, PhD; David H. Ledbetter, PhD; Christa Lese Martin, PhD; Matthew T. Oetjens, PhD

IMPORTANCE An increased risk of venous thromboembolism (VTE) has been reported in men with an additional sex chromosome. The association between other sex chromosome aneuploidies and VTE is not well characterized.

OBJECTIVE To determine if sex chromosome aneuploidy is associated with VTE.

DESIGN, SETTING, AND PARTICIPANTS Retrospective cohort study of sex chromosome aneuploidy and VTE, performed by analyzing X- and Y-chromosome dosage and VTE incidence in 642 544 individuals from 2 population-scale biobanks: the US Geisinger MyCode Community Health Initiative (N = 154 519) and the UK Biobank (N = 488 025); analysis was limited to participants self-identified as White because of inadequate sample sizes for other race and ethnicity groups. A total of 108 461 unrelated MyCode participants with electronic health record follow-up ranging from September 1996 to December 2020 and 418 725 unrelated British and Irish UK Biobank participants who attended the baseline assessment between March 2006 and October 2010, with follow-up extending to November 2020, were included in analyses of VTE.

EXPOSURES Sex chromosome aneuploidies.

MAIN OUTCOMES AND MEASURES Individuals with 1 primary inpatient VTE diagnosis, 2 primary outpatient VTE diagnoses, or a self-reported VTE diagnosis were defined as VTE cases. *P* values were adjusted for multiple comparisons.

RESULTS Identification of sex chromosome aneuploidy was undertaken among 642 544 individuals aged 18 to 90 years. Identification of a diagnosis of VTE was undertaken among 108 461 unrelated MyCode participants (65 565 [60.5%] female; mean age at last visit, 58.0 [SD, 17.6] years; median follow-up, 15.3 [IQR, 9.7] years) and among 418 725 unrelated UK Biobank participants (224 695 [53.7%] female; mean age at baseline interview, 56.9 [SD, 8.0] years; median follow-up, 12.0 [IQR, 1.6] years). Among MyCode participants, during 10 years of follow-up, 17 incident VTE events per 1353 person-years were detected among those with supernumerary sex chromosome aneuploidy (1.3% per person-year) compared with 2060 per 816 682 person-years among those with 46,XX or 46,XY (0.25% per person-year) (hazard ratio, 5.4 [95% CI, 3.4-8.7]; 10-year risk difference, 8.8% [95% CI, 4.2%-14.0%]; *P* < .001). Among UK Biobank participants, during 10 years of follow-up, 16 incident VTE events per 3803 person-years were detected among those with supernumerary sex chromosome aneuploidy (0.42% per person-year) compared with 4491 per 3 970 467 person-years among those with 46,XX or 46,XY (0.11% per person-year) (hazard ratio, 4.1 [95% CI, 2.5-6.7]; 10-year risk difference, 3.7% [95% CI, 1.4%-5.9%]; *P* < .001).

CONCLUSIONS AND RELEVANCE Adults with supernumerary sex chromosome aneuploidies compared with 2 sex chromosomes had a small but statistically significant increased risk of VTE. Further research is needed to understand the clinical implications of this association.

Supplemental content

Author Affiliations: Autism and Developmental Medicine Institute, Geisinger, Lewisburg, Pennsylvania (Berry, Finucane, Myers, Abril, Ledbetter, Martin, Oetjens); Department of Population Health Sciences, Geisinger, Danville, Pennsylvania (Kirchner).

Corresponding Author: Matthew T. Oetjens, PhD, Henry Hood Research Center, Geisinger, 100 N Academy Ave, Danville, PA 17822 (mtoetjens@geisinger.edu).

JAMA. 2023;329(3):235-243. doi:10.1001/jama.2022.23897

Individuals with sex chromosome aneuploidies, characterized by an atypical number of X or Y chromosomes, collectively comprise a common genetic group, with prevalence estimates ranging from 1 in 1400 births[1] to 1 in 650 births[2] in 2019. Sex chromosome aneuploidies are associated with anthropometric differences, such as tall stature in those with a supernumerary sex chromosome (more than 2 sex chromosomes) or short stature in those with loss of a sex chromosome, in addition to biochemical imbalances (eg, decreased testosterone levels) that can affect long-term health.[3-6] Previous sex chromosome aneuploidy prevalence estimates have been limited due to reliance on diagnosed cases, which represent a minority of total occurrences. Population-based studies have shown that sex chromosome aneuploidies, including their important medical risks, are underdiagnosed in adults.[1,6,7]

Venous thromboembolism (VTE) describes a disease spectrum that encompasses 2 subtypes of potentially fatal blood clots: (1) deep vein thrombosis, generally in the lower extremities, and (2) pulmonary embolism.[8] As of 2019, VTE affected an estimated 1 million US residents and more than 700 000 individuals in Europe annually,[9] and it is a frequent complication among intensive care inpatients and those with certain medical conditions, such as cancer and COVID-19.[10] Venous thromboembolism is the most common preventable cause of death among hospital inpatients,[11] for whom thromboprophylaxis is considered the standard of care when VTE risk is determined to be high.[12]

The association of 47,XXY and 47,XYY karyotypes with VTE has previously been described in retrospective studies[7,13] and case reports.[14,15] However, this association remains unexplored in female individuals with sex chromosome aneuploidies including trisomy X (47,XXX) and Turner syndrome (45,X). In this study, sex chromosome aneuploidies were genetically identified among adults in 2 population-based cohorts to model the association between sex chromosome aneuploidy and VTE.

Key Points

Question Is sex chromosome aneuploidy associated with venous thromboembolism (VTE)?

Findings In this retrospective multicohort study that included 642 544 adult participants, the incidence of a VTE diagnosis among those with an additional sex chromosome compared with those with 2 sex chromosomes was 1.3% per person-year compared with 0.25% per person-year, respectively, in one cohort, and 0.42% per person-year compared with 0.11% per person-year, respectively, in the other cohort. These differences were statistically significant.

Meaning The presence of a supernumerary sex chromosome aneuploidy was associated with a small but statistically significant increased risk of VTE, but further research is needed to understand the clinical implications of this association.

Methods

Participants

The Geisinger MyCode Community Health Initiative is a large health care–based cohort with linked electronic health records that began consenting patients on February 8, 2007. MyCode recruits primarily adult patients. Written informed consent was obtained from adult patients and from the parents or guardians of pediatric patients. The Geisinger institutional review board approved the study. Patients were recruited into MyCode during a primary care or specialty clinic visit at Geisinger, and eligibility did not depend on a particular condition, diagnosis, or demographic characteristic. Previous studies have shown that MyCode is similar in clinical characteristics to the Geisinger adult patient population.[16] The DiscovEHR cohort is an ongoing collaborative study between Geisinger and the Regeneron Genetics Center[16] that has generated whole-exome sequences and genotype array data on the subset of MyCode participants who have provided a blood sample to be sequenced. For the purposes of this report, MyCode refers to this DiscovEHR subset (**Table 1**; eTable 1 in Supplement 1).

The UK Biobank is a large epidemiological cohort containing extensive self-reported health data with linkage to hospital inpatient records in addition to genotype array data for most participants. This study was conducted from January 2021 to September 2022 under UK Biobank project number 49945. The UK Biobank has ethical approval from the North West Multicentre Ethics Committee. All participants provided informed consent to participate in UK Biobank projects.

To address relatedness in the cohorts, analyses of sex chromosome aneuploidy and VTE were restricted to unrelated individuals (first degree) and stratified by fixed categories of race and ethnicity. Race and ethnicity categories recorded in Geisinger's electronic health record system and self-identification from a fixed-answer questionnaire were used for this variable in MyCode and the UK Biobank, respectively. Due to the limited number of non-White participants in both cohorts, and the proportionally limited number of non-White participants with a sex chromosome aneuploidy, only the White participant subsets of each cohort had sufficient power to evaluate the association between sex chromosome aneuploidy and VTE (eTable 2 in Supplement 1). The White participant subset of the MyCode cohort was filtered to include only individuals who were at least 18 years old at last visit and who self-reported as non-Hispanic White in their electronic health record. The White participant subset of the UK Biobank cohort was filtered to include only individuals who self-reported as White British or Irish at the baseline interview. Genetic kinship was calculated using PLINK (version 1.9) and 1 randomly selected individual was removed from each first-degree kinship pair using the ukbtools R package.[17,18] Genetically inferred male sex was defined by the presence of 1 or more Y chromosomes, while genetically inferred female sex was defined by the absence of a Y chromosome, regardless of X chromosome dosage. Individuals whose reported sex did not match their genetically inferred sex were removed from each cohort.

Exposures

Sex chromosome aneuploidies were identified using an array-based approach by extracting log R ratios at single nucleotide

variants across the X and Y chromosomes (eAppendix in Supplement 1). To estimate the diagnosis rates for each sex chromosome aneuploidy, a broad group of *International Classification of Diseases, Ninth Revision* (*ICD-9*) and *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* (*ICD-10*) codes that include all sex chromosome anomalies in addition to general chromosome anomalies were used (eTable 3 in Supplement 1). Prevalence and diagnosis rates of sex chromosome aneuploidies were calculated for the whole data set before filtering for race and ethnicity and relatedness.

Outcomes

MyCode participants with a primary diagnosis of VTE were identified using inpatient and outpatient encounters. In all analyses, VTE codes used for classifying outcomes were restricted to those recorded as primary diagnoses. A confirmed VTE diagnosis in MyCode was defined as having 1 inpatient diagnosis and/or 2 independent outpatient diagnoses. Accuracy of the VTE phenotype was determined by comparing cases and controls with VTE diagnoses confirmed with manual chart review (eAppendix in Supplement 1).

A confirmed VTE primary diagnosis in the UK Biobank was defined as having 1 inpatient and/or 1 self-reported VTE diagnosis. The *ICD-9* and *ICD-10* codes for venous thrombosis of the lower extremities (deep vein thrombosis), pulmonary embolism, cerebral vein thrombosis, portal vein thrombosis, and other venous embolism or thrombosis were selected from a prior study of VTE[13] and were used to identify individuals with a diagnosis of VTE in both cohorts (eTable 3 in Supplement 1). Self-reported VTE diagnoses in the UK Biobank were based on responding to the question "Has a doctor ever told you that you have had any of the following conditions? (You can select more than one answer)" with the answers "Blood clot in the leg (deep vein thrombosis)" and/or "Blood clot in the lung." Additionally, self-reported illness codes for pulmonary embolism and/or deep vein thrombosis were included (eTable 3 in Supplement 1).

The association between sex chromosome aneuploidy and incident VTE diagnoses was assessed in MyCode and the UK Biobank. Incident VTE risk was assessed during the 10 years immediately following the baseline time point among participants without a VTE event at baseline. Baseline in MyCode was defined as 2 years after the initial visit in the health care system or 2 years after the patient's 18th birthday for those younger than 18 years at the initial visit. The 2-year postbaseline window allowed for the determination and exclusion of VTE occurring before baseline. Baseline in the UK Biobank was defined as the date of attendance at the assessment center. For analyses of prevalent VTE, events occurring at the time of last visit in MyCode and events occurring before the date of attending the assessment center in the UK Biobank were considered prevalent VTE.

Mediation of VTE Risk

Supernumerary sex chromosome aneuploidy has broad phenotypic effects that may mediate the observed association with VTE. For example, tall stature is a hallmark of supernumerary sex chromosome aneuploidy and also a risk factor for VTE.[19] To determine if the increased risk of VTE in individuals with supernumerary sex chromosome aneuploidy is explained by anthropometry and biomarkers, a mediation analysis was performed to quantify the proportion of VTE risk explained by indirect factors (ie, anthropometry and biomarkers) and the direct effect of the aneuploidy diagnosis. Because MyCode did not include many of these relevant measurements, the mediation analysis was limited to the UK Biobank cohort.

In the UK Biobank, 30 biochemical markers were measured from blood samples collected at recruitment for all participants. Additionally, several anthropometric measurements were made at assessment, including standing height, seated height, body mass index (BMI; calculated as weight in kilograms divided by height in meters squared), waist circumference, and hip circumference. Waist-to-hip ratio was calculated by dividing waist circumference by hip circumference. Leg length, which has previously been identified as an independent risk factor for VTE,[20] was calculated by subtracting seated height from standing height. All 30 blood biomarkers and 4 anthropometric measurements (BMI, standing height,

Table 1. Demographics of the MyCode Community Health Initiative and UK Biobank Cohorts

Characteristics	MyCode[a]	UK Biobank[b]
Age, mean (SD), y	56.4 (18.5)	56.5 (8.1)
Sex, No. (%)		
Female	93 785 (60.7)	264 714 (54.2)
Male	60 734 (39.3)	223 311 (45.8)
Height, mean (SD), cm	168.0 (9.9)	168.5 (9.3)
Body mass index, mean (SD)[c]	31.4 (7.6)	27.4 (4.8)
Follow-up time, mean (IQR), y	13.8 (9.9)	11.4 (1.5)
Race and ethnicity, No. (%)[d]		
Asian	569 (0.4)	11 773 (2.4)
Black	2907 (1.9)	3749 (0.8)
Latino	4011 (2.6)	NA
White	146 332 (94.7)	459 908 (94.2)
Other	293 (0.2)	10 289 (2.1)
Unknown	407 (0.3)	2306 (0.5)

Abbreviation: NA, not applicable.

[a] Values for MyCode are as follows: age at last visit; median height across visits after 18 years of age; BMI calculated using median weight and height across visits after 18 years of age. Height and BMI exclude participants with median height less than 120 cm.

[b] All values for the UK Biobank are from baseline interviews.

[c] Calculated as weight in kilograms divided by height in meters squared.

[d] Asian participant race in the UK Biobank includes Pakistani, Indian, Bangladeshi, Chinese, Asian or Asian British, White and Asian (mixed), and any other Asian background. Black participant race in MyCode includes Black or African American; Black participant race in the UK Biobank includes White and Black African (mixed), African, Black or Black British, and any other Black background. Latino participant ethnicity in MyCode includes Hispanic or Latino. The UK Biobank does not contain data on Hispanic or Latino ethnicity. White participant race in the UK Biobank includes White, White British, White Irish, or any other White background. Other participant race in MyCode includes American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, and other race; other participant race in the UK Biobank includes White and Black Caribbean, Caribbean, mixed, any other mixed background, and other ethnic group.

leg length, and waist-to-hip ratio) were converted to *z* scores and residualized for age and sex.

Statistical Analyses

To calculate the risk of 10-year incident VTE among individuals with an extra X or Y chromosome compared with those with a typical sex chromosome complement, a Cox proportional hazard regression model was fit, stratified by age binned in 10-year increments and controlling for sex and the first 4 principal components of ancestry, using the survival package[21] in R version 4.1.1.[22] Individuals with more than 1 extra chromosome (eg, 48,XXYY) were not included in analyses of VTE. Individuals with any VTE diagnosis before the baseline interview were excluded. Individuals with a supernumerary sex chromosome (47,XXX; 47,XXY; or 47,XYY) were compared as a group with those with 46,XX and 46,XY for risk of incident VTE. No variables used in prevalent or incident analyses of sex chromosome aneuploidy and VTE had missing values. The proportionality assumption of the model was tested with the cox.zph function and the results were nonsignificant. Kaplan-Meier plots were generated using the survminer package in R[23] to visualize the rate of VTE diagnosis over the 10 years immediately following baseline. Incident event rates per person-year were calculated by dividing the number of events by the sum of follow-up time during the 10-year postbaseline window. Absolute risk of 10-year incident VTE was calculated following the method described by Austin,[24] whereby individuals without sex chromosome aneuploidy were matched (2:1) to all individuals with a sex chromosome aneuploidy on age and sex. Confidence intervals were calculated using bootstrapping.

In analyses of sex chromosome aneuploidy and prevalent VTE, odds ratios for VTE diagnosis were calculated using logistic regression, controlling for age and the first 4 principal components of ancestry, in R. Absolute prevalence differences were calculated based on the sex-, age-, and ancestry-adjusted regression models. In analyses of primary VTE, individuals with only a secondary VTE diagnosis or with a single outpatient diagnosis were excluded. Individuals with 47,XXX and 45,X were compared with female individuals with a typical sex chromosome complement (46,XX). Individuals with 47,XYY and 47,XXY were compared with male individuals with a typical sex chromosome complement (46,XY). Confidence intervals for absolute prevalence differences were calculated using bootstrapping. *P* values from 4 tests per cohort were adjusted using Benjamini-Hochberg false discovery rate correction.[25] Adjusted $P < .05$ was considered significant. All reported *P* values are 2-sided.

To test the mediation of VTE risk in the UK Biobank, associations between incident VTE and each of the 34 quantitative factors among individuals without aneuploidies were determined by performing independent generalized linear modeling for each factor controlling for age, sex, and BMI. Incident VTE was treated as a binary outcome, defined as any primary VTE diagnosis occurring during the 10 years immediately following baseline. The association between incident VTE and BMI was performed controlling only for age and sex. *P* values from 34 tests for each outcome were adjusted using Benjamini-Hochberg false discovery rate correction. Similarly, the association between each quantitative factor and having a supernumerary sex chromosome was determined by performing independent generalized linear modeling for each factor controlling for age, sex, and BMI. For each test, participants with missing values were omitted from the model (eTable 4 in Supplement 1).

A mediation analysis was performed using structural equation modeling using the lavaan package in R.[26] After removing individuals with a primary or secondary prevalent VTE diagnosis, 10-year incident VTE was used as a binary endogenous variable in a structural equation model, the presence of a supernumerary sex chromosome aneuploidy was used as an exogenous variable, the 9 quantitative factors that were independently associated with both VTE and having a supernumerary sex chromosome were used as endogenous mediators, and age and sex were used as exogenous control variables. Model fit was assessed using root mean square error of approximation, a comparative fit index, and the Tucker-Lewis Index.[27] The model had a close fit, with a root mean square error of approximation less than 0.05, a comparative fit index greater than 0.95, and a Tucker-Lewis Index greater than 0.95. For each mediator, the direct effect of supernumerary sex chromosome aneuploidy on the mediator was multiplied by the direct effect of the mediator on incident VTE to calculate its indirect contribution to the association between supernumerary sex chromosome aneuploidy and incident VTE. *P* values from the 9 supernumerary sex chromosome aneuploidy indirect effect estimates were adjusted using Benjamini-Hochberg false discovery rate corrections.

Results

Sex Chromosome Aneuploidy Prevalence and Diagnosis Rates

The MyCode Community Health Initiative comprised 154 519 participants, and the UK Biobank comprised 488 025 participants. Individuals with sex chromosome aneuploidy were identified among all participants by examining dosage of the X and Y chromosomes inferred from whole-genome genotyping microarrays. Using a genome-first approach, an overall sex chromosome aneuploidy prevalence of 1 in 528 in MyCode and 1 in 832 in the UK Biobank was found (**Table 2**). Based on *ICD-9/10* codes (eTable 3 in Supplement 1), only 17.7% and 10.4% of individuals with sex chromosome aneuploidy had a documented clinical diagnosis for a chromosome anomaly in MyCode and the UK Biobank, respectively. Twenty-five individuals with sex chromosome aneuploidy were identified among non-White participants across both cohorts.

Association Between Sex Chromosome Aneuploidy and Incident VTE

Manual chart review showed that using primary VTE diagnoses in the electronic health record was more accurate (0.94) than using both primary and secondary diagnoses (0.88) (eTable 5 in Supplement 1). After removing sex mismatches (MyCode, n = 23; UK Biobank, n = 195), 108 461 unrelated White MyCode participants (65 565 [60.5%] female; mean age at first

Table 2. Prevalence and Diagnosis Rates for Individuals With SCA[a]

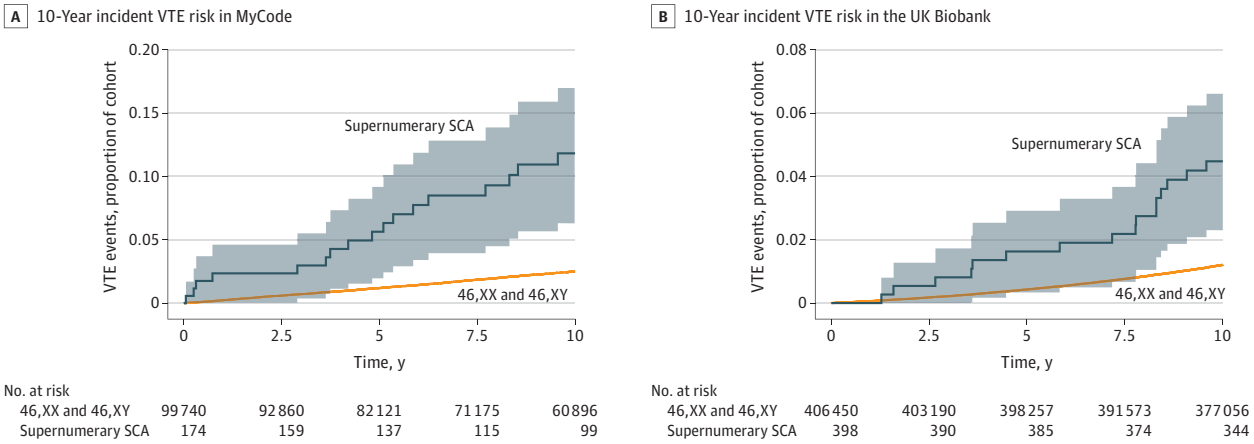
Sex chromosome complement	MyCode Community Health Initiative (N = 154 519)			UK Biobank (N = 488 025)		
	No. with SCA/ No. of male or female individuals (%)	No. with SCA per 100 000	No. diagnosed/ No. genetically ascertained (%)	No. with SCA/ No. of male or female individuals (%)	No. with SCA per 100 000	No. diagnosed/ No. genetically ascertained (%)
47,XXY	91/60 734 (0.15)	150	25/91 (27.5)	226/223 495 (0.10)	101	37/226 (16.4)
47,XYY	59/60 734 (0.10)	97	3/59 (5.1)	152/223 495 (0.07)	68	0/152
47,XXX	82/93 785 (0.09)	88	2/82 (2.4)	123/264 530 (0.05)	46	0/123
45,X	34/93 785 (0.04)	36	16/34 (47.1)	63/264 530 (0.02)	24	22/63 (34.9)
48,XXYY	2/60 734 (<0.01)	3	0/2	2/223 495 (<0.01)	1	0/2
48,XXXY	2/60 734 (<0.01)	3	2/2 (100.0)	0/223 495	0	NA
48,XXXX	1/93 785 (<0.01)	1	0/1	0/264 530	0	NA
47,XXY; 47,XYY; 48,XXYY; 48,XXXY	154/60 734 (0.25)	254[b]	30/154 (19.5)	380/223 495 (0.17)	170	37/380 (9.7)
45,X; 47,XXX; 48,XXXX	117/93 785 (0.12)	125	18/117 (15.4)	186/264 530 (0.07)	70	22/186 (11.8)

Abbreviations: NA, not applicable; SCA, sex chromosome aneuploidy.

[a] Prevalences of 47,XXY; 47,XYY; 48,XXYY; and 48,XXXY are among all male individuals, and prevalences of 45,X; 47,XXX; and 48,XXXX are among all female individuals. The number and percentage of each SCA with an electronic health record–documented diagnosis of a chromosome anomaly is shown across both cohorts.

[b] Combined prevalence of male SCA per 100 000 is higher than the sum of male SCA per 100 000 due to rounding of the individual male SCA prevalence to the nearest whole number.

Figure 1. 10-Year Incident VTE Risk in the MyCode Community Health Initiative and the UK Biobank



SCA indicates sex chromosome aneuploidy; VTE, venous thromboembolism. Kaplan-Meier plots and risk tables are shown for 10-year incident VTE risk in (A) MyCode and (B) the UK Biobank for those without SCA (46,XX and 46,XY karyotypes) and those with supernumerary SCA (47,XXX; 47,XXY; and 47,XYY karyotypes). 95% confidence intervals are shaded in the plots.

visit, 44.4 [SD, 17.7] years; mean age at baseline visit, 47.0 [SD, 16.6] years; mean age at last visit, 58.0 [SD, 17.6] years; median years of follow-up, 15.3 [IQR, 9.7]) and 418 725 unrelated White UK Biobank participants (224 695 [53.7%] female; mean age at attending baseline interview, 56.9 [SD, 8.0] years; median years of follow-up, 12.0 [IQR, 1.6]) were included in the study.

The association between supernumerary sex chromosome aneuploidy and the rate of 10-year incident VTE was investigated among those without a previous VTE diagnosis at baseline. During the 10-year follow-up in MyCode, 2078 incident VTE diagnoses (2.1% of the cohort) were observed (eTable 6 in Supplement 1). Of the incident diagnoses, 1186 were in women (1.9% of women) and 892 were in men (2.2% of men). There were no incident VTE events among individuals with 45,X in MyCode (n = 29) or in the UK Biobank (n = 61), and therefore, only the association between the presence of a supernumerary sex chromosome aneuploidy and VTE was explored in downstream analyses. The presence of a supernumerary sex chromosome was associated with increased risk of VTE during the 10-year follow-up in MyCode (1.3 events per 100 person-years) compared with individuals with 2 sex chromosomes (0.25 events per 100 person-years) after controlling for age, sex, and principal components of ancestry (hazard ratio, 5.4 [95% CI, 3.4-8.7]; 10-year risk difference, 8.8%; [95% CI, 4.2%-14.0%]; $P < .001$) (**Figure 1**A; eTable 7 in Supplement 1).

During the 10-year follow-up in the UK Biobank, 4507 incident VTE diagnoses (1.1% of the cohort) were observed. Of the incident diagnoses, 2004 were in female individuals (0.9% of females) and 2503 were in male individuals (1.3% of males). The presence of a supernumerary sex chromosome was asso-

Table 3. VTE Diagnoses Among MyCode Community Health Initiative and UK Biobank Participants Grouped by Sex Chromosome Complement[a]

Sex chromosome complement	MyCode				UK Biobank			
	Total No.	No. (%) with VTE diagnosis	Odds ratio (95% CI)[b]	Adjusted *P* value	Total No.	No. (%) with VTE diagnosis	Odds ratio (95% CI)[b]	Adjusted *P* value
46,XX	61 607	2292 (3.7)	NA	NA	224 197	6489 (2.9)	NA	NA
46,XY	40 004	1680 (4.2)	NA	NA	193 351	4609 (2.4)	NA	NA
47,XXY	72	16 (22.2)	7.78 (4.29-13.33)	<.001	204	23 (11.3)	5.58 (3.51-8.46)	<.001
47,XYY	47	6 (12.8)	3.95 (1.50-8.68)	.002	137	19 (13.9)	7.23 (4.30-11.50)	<.001
47,XXX	71	13 (18.3)	6.34 (3.30-11.28)	<.001	112	13 (11.6)	4.56 (2.43-7.86)	<.001
45,X	29	0	NA	.91	61	0	NA	.89

Abbreviations: NA, not applicable; VTE, venous thromboembolism.

[a] This table includes only unrelated White participants without exclusively secondary VTE diagnoses or a single outpatient VTE diagnosis; therefore, the sample size of each group is smaller than in Table 2.

[b] Odds ratios and 95% CIs for 45,X and 47,XXX were calculated relative to 46,XX, while those for 47,XXY and 47,XYY were calculated relative to 46,XY. Only prevalent VTE events occurring before baseline assessment in the UK Biobank are shown.

ciated with increased risk of VTE in the 10 years following initial assessment (0.42 events per 100 person-years) compared with individuals with 2 sex chromosomes (0.11 events per 100 person-years) after controlling for age, sex, and principal components of ancestry (hazard ratio, 4.1 [95% CI, 2.5-6.8]; 10-year risk difference, 3.6% [95% CI, 1.6%-5.8%]; $P < .001$) (Figure 1B; eTable 7 in Supplement 1). A sensitivity analysis including both primary and secondary diagnoses as cases of incident VTE yielded results similar to the analysis including primary diagnoses only (eTable 8 in Supplement 1).

Association Between Sex Chromosome Aneuploidy and Prevalent VTE

In MyCode, there were 4010 cases (3.7%) of VTE at the last electronic health record encounter after excluding 6614 individuals with exclusively secondary VTE diagnoses or a single outpatient diagnosis. A total of 1739 cases (43.4%) were diagnosed exclusively during outpatient visits. In the UK Biobank, there were 11 154 cases (2.7%) of prevalent VTE at the baseline interview (**Table 3**; eTable 6 in Supplement 1) after removing 588 individuals with exclusively secondary VTE diagnoses prior to attending the baseline interview. A total of 8190 prevalent VTE events in the UK Biobank were based on self-reported VTE only (70.5% of all prevalent VTE events in the UK Biobank).

Relative to individuals with 2 sex chromosomes, supernumerary sex chromosome aneuploidy was associated with VTE diagnoses in MyCode. Controlling for age and principal components of ancestry, for 47,XXX the odds ratio was 6.3 (95% CI, 3.3-11.3), with a prevalence difference of 14.6% (95% CI, 4.5%-26.2%; adjusted $P < .001$); for 47,XYY, the odds ratio was 4.0 (95% CI, 1.5-8.7), with a prevalence difference of 8.6% (95% CI, −1.6% to 20.4%; adjusted $P = .002$); and for 47,XXY, the odds ratio was 7.8 (95% CI, 4.3-13.3), with a prevalence difference of 18.0% (95% CI, 11.1%-29.1%; adjusted $P < .001$). Replication in the UK Biobank yielded similar results, whereby supernumerary sex chromosome aneuploidy was associated with having a VTE diagnosis before the baseline interview. For 47,XXX, the odds ratio was 4.6 (95% CI, 2.4-7.9), with a prevalence difference of 8.7% (95% CI, 2.8%-14.8%; adjusted $P < .001$); for 47,XYY, the odds ratio was 7.2 (95% CI, 4.1-11.5), with a prevalence difference of 11.5% (95% CI, 6.0%-17.8%; adjusted $P < .001$); and for 47,XXY, the odds ratio was 5.6 (95% CI, 3.5-8.5), with a prevalence difference of 8.9% (95% CI, 5.8%-13.4%; adjusted $P < .001$). There were no prevalent VTE events among individuals with 45,X in MyCode or the UK Biobank. A sensitivity analysis including both primary and secondary diagnoses as cases of prevalent VTE yielded results similar to the analysis including primary diagnoses only (eTable 9 in Supplement 1).

Mediators of VTE Risk in Supernumerary Sex Chromosome Aneuploidy

Of the 4 anthropometric and 30 blood biomarker baseline measurements tested, 18 were significantly associated with incident VTE among UK Biobank participants without sex chromosome aneuploidy using independent generalized linear models (adjusted $P < .05$) (eTable 10 in Supplement 1). Twenty-two of these quantitative measurements were independently associated with supernumerary sex chromosome aneuploidy (eTable 11 in Supplement 1). Nine anthropometric and blood biomarker measurements were associated with both incident VTE and supernumerary sex chromosome aneuploidy (**Figure 2**A). Structural equation modeling was performed among individuals without a previous VTE diagnosis (n = 406 848). We excluded 57 407 individuals (14.1%) with 1 or more missing predictors. Six of the 9 quantitative measurements (cystatin C, standing height, C-reactive protein, BMI, albumin, and leg length) significantly mediated the association between supernumerary sex chromosome aneuploidy and 10-year incident VTE risk (adjusted $P < .05$) (eTables 12 and 13 in Supplement 1). The indirect effect of anthropometry and biomarkers mediated 29.5% of this association, while 70.5% was unexplained by the quantitative factors tested (Figure 2B). A sensitivity analysis removing standing height, a correlate of leg length, from the structural equation model increased the direct effect of leg length (5.5%) and supernumerary sex chromosome aneuploidy (74.3%) on incident VTE, but otherwise the results were unchanged (eTable 14 in Supplement 1).

Discussion

The presence of a supernumerary sex chromosome aneuploidy was associated with a small but statistically significant increased risk of VTE, independent of the tested clinical bio-

Figure 2. Quantitative Factors Associated With VTE and Supernumerary Sex Chromosome Aneuploidy That Mediate Some of the Association Between 47,XXX, 47,XYY, and 47,XXY and 10-Year Incident VTE Risk in the UK Biobank

A Quantitative factors and biomarkers associated with incident VTE and supernumerary sex chromosome aneuploidy

Increased risk of VTE	Supernumerary sex chromosome aneuploidy	Increased risk of VTE and supernumerary sex chromosome aneuploidy
Positive association	**Positive association**	**Positive association**
Creatinine	Glucose	Body mass index
Urate	Total protein	Waist-to-hip ratio
Urea	Hemoglobin A_{1c}	Leg length
Negative association	**Negative association**	Standing height
Aspartate aminotransferase	Testosterone	Alkaline phosphatase
Calcium	Bilirubin	C-reactive protein
Triglycerides	Urate	Cystatin C
Glucose	Vitamin D	Sex hormone–binding globulin
Total protein	Apolipoprotein a	**Negative association**
Alanine aminotransferase	Apolipoprotein b	Albumin
	Cholesterol	
	High-density lipoprotein	
	Insulin-like growth factor 1	
	Low-density lipoprotein	

B Effect of quantitative factors and supernumerary sex chromosome aneuploidy on incident VTE

Significantly mediating quantitative factors	Estimated total effect mediated, %
Cystatin C	8.5
Standing height	7.2
C-reactive protein	5.6
Body mass index	3.1
Albumin	3.0
Leg length	2.1
Sum of indirect effects	**29.5**
Supernumerary sex chromosome aneuploidy	**70.5**
Incident VTE	**100**

VTE indicates venous thromboembolism. Among the 4 anthropometric factors and 30 blood biomarkers measured by the UK Biobank, panel A lists which ones were associated with incident VTE in 46,XX and 46,XY, with supernumerary sex chromosome aneuploidy, and with both. All quantitative factors with adjusted $P < .05$ are shown. In panel B, the quantitative factors significantly associated with both VTE risk and with supernumerary sex chromosome aneuploidy were included in a mediation model in addition to age and sex. The quantitative factors that significantly mediated (adjusted $P < .05$) the association between 47,XXX, 47,XYY, and 47,XXY and 10-year incident VTE are shown, each with the percentage of the total effect mediated. The direct effect of supernumerary sex chromosome aneuploidy on incident VTE is also shown.

markers, whereas this increased risk was not present for female individuals with 45,X. The risk estimates were in strong agreement across the 2 large biobanks included in the analysis.

An association with VTE was recently reported among other findings in male individuals with sex chromosome anomalies in the UK Biobank.[7] The current study identified and explored incidence of VTE in male and female individuals with sex chromosome aneuploidies in 2 independent cohorts. The mediation analysis suggested that most of the relationship between supernumerary sex chromosome aneuploidy and VTE was independent of clinical factors, including BMI, height, leg length, cystatin C, C-reactive protein, and albumin. The potential clinical utility of identifying supernumerary sex chromosome aneuploidy for VTE risk stratification is unknown.

The sex chromosome aneuploidy prevalence in MyCode was 1 in 528, similar to previous newborn screening estimates of 1 in 450.[28] The lower frequency of sex chromosome aneuploidies in the UK Biobank (1 in 832) may reflect a difference in ascertainment. A healthy volunteer bias has been shown to skew ascertainment of participants into the UK Biobank relative to the general population.[29] The deleterious effects of sex chromosome aneuploidy on cognition and medical health may reduce participation in the UK Biobank.[3-6]

Despite differences in ascertainment, the relative increase in 10-year incident VTE risk associated with supernumerary sex chromosome aneuploidy was similar in MyCode and the UK Biobank (hazard ratios, 5.4 and 4.1, respectively). However, the absolute risk of incident VTE is notably higher in MyCode (8.8%) compared with the UK Biobank (3.6%). This difference in risk of incident VTE may partially be explained by how events were captured in the 2 cohorts during follow up. In the UK Biobank, incident VTE was captured from inpatient encounters, while in MyCode, incident VTE was captured from both inpatient and outpatient encounters. Of all VTE diagnoses in MyCode, 43.4% were captured exclusively during outpatient visits. At the time of this study, the UK Biobank provided limited primary care data after the baseline interview, restricting 10-year incident VTE identification to inpatient diagnoses. Therefore, complete health records found in MyCode provide a more accurate and real-world estimate of absolute risk of incident VTE in individuals with supernumerary sex chromosome aneuploidies.

Several potentially confounding issues should be considered when interpreting the validity of the prevalence analyses. Prevalence analyses do not account for time at risk. Therefore, the strength of the association can be biased by differences between risk groups, including the baseline age distributions at study enrollment and the rates of entering and leaving the study for reasons other than death. Prevalence analyses are also susceptible to survivor and volunteer biases, which can affect ascertainment of participants with some medical conditions and clinical risk factors into research studies. Additionally, the analysis of prevalent VTE in the UK Biobank includes self-reported VTE. The accuracy of self-reported health information in the UK Biobank, including VTE, is not well characterized. The incident analyses presented herein address many of the limitations of prevalence analyses by accounting for time at risk in a previously established cohort using a highly accurate phenotype definition validated with manual chart review.

Most participants in both cohorts used in the present study were of White race; thus, the prevalence of supernumerary sex chromosome aneuploidy, VTE, and the association between supernumerary sex chromosome aneuploidy and VTE in non-White populations could not be evaluated. Genetic VTE research in non-White populations has been limited, so impor-

tant genetic causes of VTE in these groups may be yet undiscovered. More research is necessary to determine whether supernumerary sex chromosome aneuploidy is an important genetic risk factor for VTE in non-White populations.

Whether knowledge of a sex chromosome aneuploidy would affect clinical management of VTE is unclear. At this time, the American College of Medical Genetics does not provide a consensus regarding the role of genotype-guided treatment of VTE.[30] As VTE is a major public health concern, more research into benefits and harms of genetic thrombophilia testing for management and prevention of VTE is needed.[31] Guidelines may evolve in response to new data on VTE risk factors, including inherited forms and sex chromosome aneuploidies, generated from large biobanks.

Limitations

This study has several limitations. First, detection of sex chromosome imbalances with genotype array data may not be sensitive enough to provide precise estimates of mosaicism. Validation of aneuploidies with more sensitive techniques such as whole-genome sequencing or cytogenetic confirmation (eg, fluorescence in situ hybridization) may be able to determine if mosaicism of the sex chromosome aneuploidy influences the risk of VTE. Second, at this time, the UK Biobank does not have a mechanism for researchers to recontact individual participants to collect or confirm phenotypic data. Therefore, limited options for validating phenotypic definitions may be a persistent limitation of the UK Biobank. Third, recurrent VTE was not considered as an outcome. Future studies are needed to investigate if the risk of recurrent VTE is different in those with supernumerary sex chromosome aneuploidy. Fourth, due to the lack of race and ethnicity diversity in both cohorts, the primary genetic analyses performed in this study were limited to White participants, precluding the ability to fully understand if supernumerary sex chromosome aneuploidy is associated with VTE across populations. Fifth, analyses of prevalent VTE may be influenced by ascertainment bias, differences in time at risk between risk groups, and reliance on self-report in the UK Biobank, the accuracy of which is not well characterized.

Conclusions

Adults with supernumerary sex chromosome aneuploidies compared with those having 2 sex chromosomes had a small but statistically significant increased risk of VTE. Further research is needed to understand the clinical implications of this association.

ARTICLE INFORMATION

Accepted for Publication: December 7, 2022.

Author Contributions: Dr Oetjens had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Berry, Myers, Martin, Oetjens.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Berry, Finucane, Oetjens.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Berry, Kirchner, Oetjens.
Obtained funding: Finucane, Ledbetter, Martin.
Administrative, technical, or material support: Martin, Oetjens.
Supervision: Ledbetter, Martin, Oetjens.

Conflict of Interest Disclosures: Ms Finucane reported receipt of grants from the National Institutes of Health (NIH)/National Institute of Mental Health (NIMH) outside the submitted work and personal fees from the National Fragile X Foundation and Labcorp. Dr Myers reported receipt of grants from the NIH/NIMH outside the submitted work and grants from the NIH/National Human Genome Research Institute (NHGRI) and the AllOne Foundation. Dr Abril reported receipt of grants from the NIH/NIMH outside the submitted work. Dr Ledbetter reported being an employee of and receipt of personal fees from the Unified Patient Network, receipt of personal fees from X-Therma and Singular Genomics, and being a scientific consultant to Natera and MyOme. Dr Martin reported receipt of grants from the NIH/NIMH outside the submitted work and receipt of grants from the NIH/NHGRI and the Simons Foundation. Dr Oetjens reported receipt of grants from the NIH/NIMH outside the submitted work. No other disclosures were reported.

Funding/Support: This study was supported by NIMH grants R01MH074090 and U01MH119705 (Drs Ledbetter and Martin).

Role of the Funder/Sponsor: The NIMH had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Data Sharing Statement: See Supplement 2.

Additional Contributions: We thank David Rolston, MD, Internal Medicine, Geisinger, and Samuel Gidding, MD, Genomic Medicine Institute, Geisinger, for providing thoughtful feedback on the manuscript and Yirui Hu, PhD, Department of Population Health Sciences, Geisinger, for guidance and providing code for statistical analyses. Drs Rolston, Gidding, and Hu were not compensated for their contributions. We thank all patients engaged in the MyCode Community Health Initiative and the MyCode research team. We also acknowledge the Geisinger-Regeneron DiscovEHR collaboration contributors, who have been critical in the generation of the data used for this study.

REFERENCES

1. Berglund A, Viuff MH, Skakkebæk A, Chang S, Stochholm K, Gravholt CH. Changes in the cohort composition of turner syndrome and severe non-diagnosis of Klinefelter, 47,XXX and 47,XYY syndrome: a nationwide cohort study. *Orphanet J Rare Dis*. 2019;14(1):16. doi:10.1186/s13023-018-0976-2

2. Oetjens MT, Kelly MA, Sturm AC, Martin CL, Ledbetter DH. Quantifying the polygenic contribution to variable expressivity in eleven rare genetic disorders. *Nat Commun*. 2019;10(1):4897. doi:10.1038/s41467-019-12869-0

3. Berglund A, Stochholm K, Gravholt CH. The comorbidity landscape of 47,XXX syndrome: a nationwide epidemiologic study. *Genet Med*. 2022;24(2):475-487. doi:10.1016/j.gim.2021.10.012

4. Gravholt CH, Chang S, Wallentin M, Fedder J, Moore P, Skakkebæk A. Klinefelter syndrome: integrating genetics, neuropsychology, and endocrinology. *Endocr Rev*. 2018;39(4):389-423. doi:10.1210/er.2017-00212

5. Stochholm K, Juul S, Gravholt CH. Diagnosis and mortality in 47,XYY persons: a registry study. *Orphanet J Rare Dis*. 2010;5:15. doi:10.1186/1750-1172-5-15

6. Tuke MA, Ruth KS, Wood AR, et al. Mosaic Turner syndrome shows reduced penetrance in an adult population study. *Genet Med*. 2019;21(4):877-886. doi:10.1038/s41436-018-0271-6

7. Zhao Y, Gardner EJ, Tuke MA, et al. Detection and characterization of male sex chromosome abnormalities in the UK Biobank study. *Genet Med*. 2022;24(9):1909-1919. doi:10.1016/j.gim.2022.05.011

8. Perrier A. Deep vein thrombosis and pulmonary embolism: a single disease entity with different risk factors? *Chest*. 2000;118(5):1234-1236. doi:10.1378/chest.118.5.1234

9. Benjamin EJ, Muntner P, Alonso A, et al; American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2019 update: a report from the American Heart Association. *Circulation*. 2019;139(10):e56-e528. doi:10.1161/CIR.0000000000000659

10. Middeldorp S, Coppens M, van Haaps TF, et al. Incidence of venous thromboembolism in hospitalized patients with COVID-19. *J Thromb Haemost*. 2020;18(8):1995-2002. doi:10.1111/jth.14888

11. Geerts WH, Pineo GF, Heit JA, et al. Prevention of venous thromboembolism: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest*. 2004;126(3)(suppl):338S-400S. doi:10.1378/chest.126.3_suppl.338S

12. Skeik N, Westergard E. Recommendations for VTE prophylaxis in medically ill patients. *Ann Vasc Dis*. 2020;13(1):38-44. doi:10.3400/avd.ra.19-00115

13. Zöller B, Ji J, Sundquist J, Sundquist K. High risk of venous thromboembolism in Klinefelter syndrome. *J Am Heart Assoc*. 2016;5(5):e003567. doi:10.1161/JAHA.116.003567

14. Boos CJ, Matfin G. Klinefelter's syndrome manifesting as an acute pulmonary embolus in a 52-year-old man. *Endocr Pract*. 2002;8(1):68-69. doi:10.4158/EP.8.1.68

15. Murray FE. Mesenteric vein thrombosis associated with Klinefelters syndrome—a case report. *Angiology*. 1988;39(1 pt 1):45-48. doi:10.1177/000331978803900107

16. Carey DJ, Fetterolf SN, Davis FD, et al. The Geisinger MyCode Community Health Initiative: an electronic health record-linked biobank for precision medicine research. *Genet Med*. 2016;18(9):906-913. doi:10.1038/gim.2015.187

17. Hanscombe KB, Coleman JRI, Traylor M, Lewis CM. Ukbtools: an R package to manage and query UK Biobank data. *PLoS One*. 2019;14(5):e0214311. doi:10.1371/journal.pone.0214311

18. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(7):s13742-015-0047-0048. doi:10.1186/s13742-015-0047-8

19. Lutsey PL, Virnig BA, Durham SB, et al. Correlates and consequences of venous thromboembolism: the Iowa Women's Health Study. *Am J Public Health*. 2010;100(8):1506-1513. doi:10.2105/AJPH.2008.157776

20. Lutsey PL, Cushman M, Heckbert SR, Tang W, Folsom AR; Longitudinal Study of Thromboembolism Etiology. Longer legs are associated with greater risk of incident venous thromboembolism independent of total body height. *Thromb Haemost*. 2011;106(1):113-120.

21. Therneau TM. *A Package for Survival Analysis in R*. Published 2021. Accessed November 2, 2021. https://CRAN.R-project.org/package=survival

22. R Core Team. *R: A Language and Environment for Statistical Computing*. Published 2020. Accessed November 1, 2021. https://www.R-project.org/

23. Kassambara A, Kosinski M, Biecek P. *Survminer: Drawing Survival Curves Using "ggplot2."* Published 2021. Accessed November 2, 2021. https://CRAN.R-project.org/package=survminer

24. Austin PC. Absolute risk reductions and numbers needed to treat can be obtained from adjusted survival models for time-to-event outcomes. *J Clin Epidemiol*. 2010;63(1):46-55. doi:10.1016/j.jclinepi.2009.03.012

25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

26. Rosseel Y. Lavaan: an R package for structural equation modeling. *J Stat Softw*. 2012;48(2):1-36. doi:10.18637/jss.v048.i02

27. Xia Y, Yang Y. RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: the story they tell depends on the estimation methods. *Behav Res Methods*. 2019;51(1):409-428. doi:10.3758/s13428-018-1055-2

28. Nielsen J, Wohlert M. Sex chromosome abnormalities found among 34,910 newborn children: results from a 13-year incidence study in Arhus, Denmark. *Birth Defects Orig Artic Ser*. 1990;26(4):209-223.

29. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol*. 2017;186(9):1026-1034. doi:10.1093/aje/kwx246

30. Zhang S, Taylor AK, Huang X, et al; ACMG Laboratory Quality Assurance Committee. Venous thromboembolism laboratory testing (factor V Leiden and factor II c.*97G>A), 2018 update: a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2018;20(12):1489-1498. doi:10.1038/s41436-018-0322-z

31. Cushman M, Barnes GD, Creager MA, et al; American Heart Association Council on Peripheral Vascular Disease, Council on Arteriosclerosis, Thrombosis and Vascular Biology, Council on Cardiovascular and Stroke Nursing, Council on Clinical Cardiology, and Council on Epidemiology and Prevention; International Society on Thrombosis and Haemostasis. Venous thromboembolism research priorities: a scientific statement from the American Heart Association and the International Society on Thrombosis and Haemostasis. *Circulation*. 2020;142(6):e85-e94. doi:10.1161/CIR.0000000000000818