Incidence of venous thrombotic events and events of special interest in a retrospective cohort of commercially insured US patients

Susan C Weller, Laura Porterfield, Gregg S Wilkinson, Lu Chen, Jacques Baillargeon

ABSTRACT

Objective To estimate the US incidence of thrombotic events and related rare diagnoses.

Design Claims-based retrospective cohort study of incidence.

Setting US commercial health insurance administrative claims database.

Participants Adults 25–64 years of age between 2015 and 2019 with a minimum of 12 consecutive thrombosis-free months of continuous enrollment beginning 2014 were selected.

Main outcomes Age (10-year intervals) and sex stratum-specific incidence rates per 100,000 person-years were determined for venous thromboembolism (VTE), cerebral venous thrombosis (CVT) and other major venous thrombotic events, and events of special interest, including immune thrombocytopenic purpura (ITP), haemolytic-uremic syndrome (HUS) and heparin-induced thrombocytopenia (HIT).

Results Of 13,249,229 enrollees (half female/male), incidence of venous thromboembolic events (deep vein thrombosis (DVT), pulmonary embolism (PE), CVT or other major venous thrombotic conditions) was 247.89 per 100,000 person-years (95% CI: 245.96 to 249.84). Incidence of VTE was 213.79 with ICD codes alone (95% CI: 211.99 to 215.59) and 129.34 (95% CI: 127.95 to 130.75) when also requiring a filled anticoagulation prescription or an inferior vena cava (IVC) filter. Incidence was 8.37 for CVT (95% CI: 6.07 to 6.69), 26.06 for ITP (95% CI: 25.44 to 26.78), 0.94 for HUS (95% CI: 0.82 to 1.06) and 4.82 for HIT (95% CI: 4.56 to 5.10). The co-occurrence of CVT with either ITP or HIT (diagnoses within 14 days of one another) was 0.090 (95% CI: 0.06 to 0.13). Incidence tended to increase with age and was higher for women under 55. Incidence for CVT, HUS and CVT with ITP or HIT was higher for women in all age groups. Incidence of PE and CVT increased significantly over the 5-year period, while DVT rates decreased.

Conclusions These results are the first US estimates for the incidence of thrombotic and rare events of interest in a large, commercially insured US population. Findings provide a critically important reference for determining excess morbidity associated with COVID-19 and more generally for vaccine pharmacovigilance.

Although interest in incidence of rare thrombotic events (eg, cerebral venous thrombosis (CVT)) has increased with recent SARS-CoV-2 virus (COVID-19) and vaccine adverse event surveillance, few recent estimates exist for the USA. This study estimates the pre-pandemic incidence of thrombotic events and rare conditions associated with thrombocytopenia that are potentially associated with COVID-19 disease and/or vaccination. Using a large, US commercial healthcare administrative database, we provide the first estimates of incidence since improvements over the past decade in prevention and detection of thromboses. These data will be useful in evaluating morbidity associated with COVID-19 disease and, more generally, for pharmacovigilance.

Incidence estimates for thrombotic events have been inconsistent and limited by population size. The focus has been on the more common diagnosis of venous thromboembolism (VTE), lower extremity deep vein thrombosis (DVT) and/or pulmonary embolism (PE), but estimates are highly variable across populations. Limited data are available on the incidence of rarer thromboses because samples were too small to detect events.


Strengths and limitations of this study

► This is the first report to provide comprehensive age-stratified and sex-stratified incidence estimates for venous thromboembolic and other rare events of interest in the USA for the 5-year pre-pandemic period (2015–2019).

► The dataset used to estimate rates is large (over 13 million people with more than a year of observation) and facilitates estimation of rare events.

► Unlike studies using national health care data, this study used a commercially insured cohort to estimate incidence and thus, may not be representative of the US population.
Recently, a report from Denmark estimated combined incidence for major venous thromboses, as well as for VTE alone, and data from the European Union (EU) and the UK estimated CVT incidence. In this study, incidence of VTE, CVT, other thromboses, immune thrombocytopenic purpura (ITP), haemolytic-uremic syndrome (HUS) and heparin-induced thrombocytopenia (HIT) was estimated using administrative data from a large, US commercial health insurance programme. Incidence was estimated stratified by sex and age, among adults aged 26–64 years for the most recent 5-year period prior to the pandemic (2015–2019).

METHODS
Diagnoses were identified using the Optum Clininformatics Data Mart Database, one of the largest commercial insurance populations in the USA. The database contains de-identified inpatient and outpatient claims (medical, pharmaceutical and laboratory) for approximately 62 million unique enrollees (2007–2018); 81% of whom have at least one medical claim per year and almost half of whom have continuous enrolment for 2 or more years. The distribution of age (25–64 years) and sex in the claims population is similar to that of the US population (within 1%–3%), but more members reside in the southern US region (11% higher).7

Diagnoses were identified for: acute VTE (DVT: ICD-9: 451.1x, 451.2x, 451.81, 451.83; ICD-10: I80.1x, I80.2x, I80.3; PE: ICD-9: 415.1x, ICD-10: I12.6x, I12.9x); CVT (ICD-9: 325.0, 437.6; ICD-10: G08, 167.6, 163.6) and other major venous thrombotic conditions, excluding pregnancy-related conditions (portal vein thrombosis, hepatic vein thrombosis/Budd-Chiari syndrome, thrombophlebitis migrans, embolism or thrombosis of inferior vena cava, embolism or thrombosis of renal vein, or mesenteric thrombosis: ICD-9: 452, 453.0, 453.1, 453.2, 453.3, 557.0; ICD-10: 181, 182.0, 182.1, 182.2, 182.3, K55.0H). Additional diagnoses were ITP (ICD-9: 287.31, ICD-10: D69.3), HUS (ICD-9: 283.11, ICD-10: D59.3) and HIT (ICD-9: 289.84; ICD-10: D75.82). VTE incidence was first estimated using ICD codes alone and second requiring that an anticoagulant be filled (excluding ‘flush’ formulations of heparins), an inferior vena cava (IVC) filter be placed or disenrolment occur within 30 days of the diagnosis.

Incidence was calculated as the frequency of each diagnosis divided by disease-free person-time across a 5-year period with 95% CIs.9 Quarterly incidence was calculated to assess variation and trends across time; monotonic trends were examined using an ordinal correlation coefficient across the 20 quarterly estimates.9 All enrollees between 26 and 65 years of age with a minimum of 12 consecutive months enrolment were followed from 1 January 2014 to 31 December 2019. Rates were calculated by sex and 10-year age categories: 26–35, 36–45, 46–55 and 56–64. A 12-month event-free lookback period was required for each outcome, before follow-up began. Denominators contained time each enrollee contributed during follow-up (whether or not they used medical services), censored at time of diagnosis, turning 65 or disenrolment. Because Medicare begins at age 65, we limited follow-up to enrollees under 65. There was no patient involvement in this research. To provide a description of the cohort, prevalence of several diagnoses also was calculated (for details see online supplemental appendix 1).

Patient and public involvement
None.

RESULTS
The 5-year sample included 13249229 people with one or more years of observation; 6555995 were women and 6693634 were men, with approximately 3 million people in each of the four age categories (for details see online supplemental appendix 1). Individuals with missing sex (0.01%) or age (0.00007%) information were excluded. The most common chronic conditions (see online supplemental appendix 1) were hypertension (20.93%), followed by obesity (9.31%), tobacco use (8.58%), diabetes (8.54%), coronary artery disease (3.16%) and cancer (2.33%). Pregnancy was recorded for 14.54% of women 26–35 years and for 5.39% of women 36–45 years of age.

For enrollees 25–64 years of age, incidence of a venous thrombotic event (VTE, CVT or other major venous thrombosis) was 247.89 per 100 000 person-years (95% CI: 245.96 to 249.84; 62598 cases with 2525020 person-years). Rates increased with age and were higher for women than for men until age 55, with larger sex differences at younger ages (for all outcomes, see table 1 for rates per 100 000 for subgroups and overall; see online supplemental appendix 1 for number of cases, person-years and incidence).

Incidence of VTE was 213.79 per 100 000 person-years (95% CI: 211.99 to 215.59). Incidence of VTE, DVT without PE (47.73; 95% CI: 46.88 to 48.58) and PE with or without DVT (171.18; 95% CI: 169.58 to 172.80) had similar sex by age patterns with higher rates at older ages and higher rates for women younger than 55 (online supplemental appendix 1). With a more restrictive estimate of VTE (requiring anticoagulation or an IVC filter), incidence of VTE (VTE+r, table 1) was 129.34 (95% CI: 127.95 to 130.75); 16.91 for DVT alone (95% CI: 16.41 to 17.42) and 115.02 for PE with or without DVT (95% CI: 113.70 to 116.34). With the restricted definition, rates were higher for women younger than 46 years.

Quarterly incidence estimates with the restricted definition for DVT, PE and VTE indicated that DVT decreased significantly across the 5-year period (rhowo=0.95 for women’s total rates, n=20, p<0.0001; rhowo=0.98 for men’s total rates, n=20, p<0.0001; figure 1 top panel), while quarterly PE estimates increased significantly (rhowo=0.89 for women’s total rates, n=20, p<0.0001; rhowo=0.68 for
The combined, VTE quarterly rates showed a significant increase across 5 years for women (\( r_{ho} = +0.77 \) total rates, \( n=20, p<0.0001 \)) and a small, non-significant increase for men (\( r_{ho} = +0.11 \) total rates, \( n=20, p<0.66 \); figure 1 middle panel). Rates did not appear to be affected by the change from ICD-9 to ICD-10 codes in 2015.

Five-year incidence per 100 000 person-years was 6.37 for CVT (95% CI: 6.07 to 6.69), increasing significantly across time for women (\( r_{ho} = +0.66 \) total rates, \( n=20, p<0.002 \)) and men (\( r_{ho} = +0.58 \) total rates, \( n=20, p<0.01 \); figure 1 bottom panel, numeric detail in online supplemental appendix 1). Rates did not appear to be affected by the change from ICD-9 to ICD-10 codes in 2015.

Five-year incidence for ITP was 26.06 (95% CI: 25.44 to 26.78), 4.82 for HIT (95% CI: 4.56 to 5.10) and 0.94 for HUS (95% CI: 0.84 to 1.06). Incidence was higher in women of all ages for CVT and HUS. Rates for the co-occurrence of CVT with other diagnoses (within 14 days) was 0.04 with ITP (95% CI: 0.02 to 0.06), 0.06 with HIT (95% CI: 0.03 to 0.09) and 0.09 with either ITP or HIT (95% CI: 0.06 to 0.13). For the co-occurrence of CVT with other events (online supplemental appendix 1), incidence was higher in women under 55 years, but was higher in women of all ages for CVT with ITP or HIT.

Although it was not part of our trend analysis, rates are shown for the first quarter of 2020 (figures 1 and 2, and online supplemental appendix 1), because there has been speculation whether the SARS-CoV-2 coronavirus was circulating in winter 2020 and whether it might elevate

**Table 1** Incidence by age and sex per 100 000 person-years

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>26–35</td>
<td>36–45</td>
</tr>
<tr>
<td>VTE (DVT OR PE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>95.26*</td>
<td>161.16*</td>
</tr>
<tr>
<td>Male</td>
<td>51.31</td>
<td>105.99</td>
</tr>
<tr>
<td>Total</td>
<td>72.43</td>
<td>132.87</td>
</tr>
<tr>
<td>VTE+a (DVT+a OR PE+a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>51.68*</td>
<td>90.02*</td>
</tr>
<tr>
<td>Male</td>
<td>29.78</td>
<td>65.90</td>
</tr>
<tr>
<td>Total</td>
<td>40.31</td>
<td>77.66</td>
</tr>
<tr>
<td>CVT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>8.61</td>
<td>8.93</td>
</tr>
<tr>
<td>Male</td>
<td>2.54</td>
<td>2.70</td>
</tr>
<tr>
<td>Total</td>
<td>5.46</td>
<td>5.74</td>
</tr>
<tr>
<td>Any major venous thrombosis (DVT, CVT and other)†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>113.72</td>
<td>187.86</td>
</tr>
<tr>
<td>Male</td>
<td>61.13</td>
<td>123.55</td>
</tr>
<tr>
<td>Total</td>
<td>86.40</td>
<td>154.87</td>
</tr>
<tr>
<td>ITP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>30.38</td>
<td>27.09</td>
</tr>
<tr>
<td>Male</td>
<td>9.73</td>
<td>13.06</td>
</tr>
<tr>
<td>Total</td>
<td>19.66</td>
<td>19.90</td>
</tr>
<tr>
<td>HUS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.96</td>
<td>1.08</td>
</tr>
<tr>
<td>Male</td>
<td>0.38</td>
<td>0.56</td>
</tr>
<tr>
<td>Total</td>
<td>0.66</td>
<td>0.81</td>
</tr>
<tr>
<td>HIT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.52</td>
<td>1.92</td>
</tr>
<tr>
<td>Male</td>
<td>0.58</td>
<td>1.99</td>
</tr>
<tr>
<td>Total</td>
<td>1.03</td>
<td>1.96</td>
</tr>
</tbody>
</table>

*Does not include pregnancy-related events.
†Any major venous thrombosis: VTE, CVT, portal vein thrombosis, hepatic vein thrombosis/Budd-Chiari syndrome, thrombophlebitis migrans, embolism or thrombosis of vena cava (intrahepatic), embolism or thrombosis of renal vein or mesenteric thrombosis.
CVT, cerebral venous thrombosis; DVT, deep vein thrombosis; HIT, heparin-induced thrombocytopenia; HUS, haemolytic-uremic syndrome; ITP, immune thrombocytopenic purpura; PE, pulmonary embolism; VTE, venous thromboembolism.
rates. Population incidence for the first quarter of 2020 was higher than the previous four quarters for PE, VTE and CVT, but these rates were part of a series that was increasing over time and the increase occurred in rates for men and not for women. Population incidence for the first quarter of 2020 was not higher than the previous four quarters for DVT (for men and women).

**DISCUSSION**

This is the first report to provide comprehensive age-stratified and sex-stratified incidence estimates for venous thromboembolic and other rare events of interest in the USA. Estimates cover the 5 years preceding the COVID-19 pandemic. Rates increased with older age for most outcomes and tended to be higher for women under 55 years. Incidence for CVT, HUS and CVT with ITP or HIT was higher for women at all ages.

Incidence for VTE ranged from 129 per 100,000 person-years when a restricted case definition was used (based on ICD codes plus an anticoagulant or IVC filter) to 214 when the case definition was based exclusively on ICD codes. The higher estimate includes cases, as well as suspected cases, and likely overestimates the true rate. It is, however, the method used in other studies. The lower estimate may underestimate the true incidence as anyone receiving—but not filling—their anticoagulant

---

**Figure 1** Venous thromboembolism events by year and quarter. DVT, deep vein thrombosis; PE, pulmonary embolism; IVC, inferior vena cava filter.
prescription would be excluded, which may occur as often as in 40% of cases. Thus, results may be limited by the use of administrative-level data and without detailed patient records. However, a recent cohort study of Kaiser-Permanente patients (insured, with complete patient records) compared COVID-19-positive with COVID-19-negative patients and estimated the incidence per 100,000 of VTE in COVID-19-negative patients (n=200,000) as 160 for outpatients and 300 for hospital-associated diagnoses (inpatient or post-hospital). Thus, these estimates, which are also higher than European rates, suggest that our estimates may provide upper and lower bounds for true VTE incidence in the USA.

Population differences in rates are notable, as CIs between studies often do not overlap. In the past decade, VTE incidence per 100,000 person-years ranged from 150 in the UK to 200 in Spain for adults of all ages (figure 3; for study details see online supplemental appendix 1). A recent report from Denmark using the same diagnostic codes without restrictions estimated incidence (2010–2018) for VTE at 170 and 176 for a broader category of venous thromboses among those 18–99 years of age in a population base of 5 million. For those 18–64 years of age, incidence was 91 for VTE and 95 for the more inclusive category. In this study, rates for similar age and time ranges were twice as high (214 for VTE and 248 with other major venous thrombotic events).

CVT estimates may appear more consistent across populations, but also may have non-overlapping CIs. In the past decade, CVT incidence per 100,000 ranged from 0.12 in the UK to 2.00 in New York/Florida in the

Figure 2 Cerebral venous thrombosis events by year and quarter.

Figure 3 Venous thromboembolism incidence across countries and time.
USA (for study details see online supplemental appendix 1). In this analysis, CVT incidence (6.37) was several times higher than that reported in European countries and three times higher than in New York/Florida, although the previous US study included pregnancy-related events and this study and the Denmark study did not. In contrast, the ITP rate in this study (26.06) fell within the range of 2019 European rates from 20.12 in the UK to 95.71 in Spain. The previous US study used hospital discharge data and estimated state populations to estimate CVT incidence, while this study used diagnoses occurring in any setting in a well-defined cohort; but both studies found increased incidence over time. Validation studies indicate high positive predictive value (87%–100%) for the CVT ICD-9 and ICD-10 codes used in this study, with the exception of 437.6 (33%), but additionally indicate some cases were outpatient diagnoses.

Trends through time indicated that in the past 5 years, US rates for DVT decreased significantly and PE rates increased significantly for both sexes, but the aggregate VTE rates showed an increase only for women. There has been speculation that the SARS-CoV-2 coronavirus may have elevated thrombotic events as it began circulating in winter 2020. Although this is outside of our original protocol and one data point is not sufficient to discern if there has been a change in trend, we presented the rates for the first quarter in 2020. The first quarter rates for 2020 appeared elevated only for men for PE (and VTE) and CVT, but these rates appear within a series of rates for PE and CVT that have been increasing across time. More data would be necessary to discern whether the rate of increase (slope) was greater in 2020.

Inconsistencies across studies may be due to methodological differences between studies and to differences in underlying population characteristics (age, smoking, obesity and comorbidities), lack of standardisation, differences in included ICD codes or detection bias. To be valid, a comparison of population-based rates must at a minimum be similar in several critical areas such as time periods and be standardised to the same reference population on those characteristics for which data are available such as age and sex. Use of a commercially insured cohort to estimate incidence has limited generalisability to the entire US population (unlike studies using national healthcare system data). People in the cohort tend to be employed and may be healthier than the general population, although exclusion of the uninsured could result in slightly higher estimates of non-life-threatening health service usage. Also, estimates based on administrative data not linked to patient records could overestimate rates by including suspected/probable cases among the definite cases. However, VTE estimates appeared to be similar to those from an insured population with access to complete patient records and CVT codes had good positive predictive value. A further limitation is that this study focused on non-pregnancy-related outcomes, but did not exclude women who might be pregnant from the denominators, consistent with EU studies (eg, Østergaard et al). This could underestimate VTE incidence for women in the youngest age group, but would have only a very small effect on total incidence for women. Future
research might attempt to validate pregnancy codes and estimate population rates for pregnant and post-partum women of reproductive age. Also, future research may attempt meta-analyses of studies with similar methodologies, as was done in Canada.18

This is the first comprehensive US report on the incidence of these conditions, since advances have been made in the past decade in thrombophrophylaxis and detection. Although generalisations to the US population from this dataset may be limited, given the restriction to commercially insured persons, there are few longitudinal US databases with sufficient clinical and diagnostic data for those younger than 65 years. Thus, the Optum database is one of the few data sources that permit estimation of rare health events in the US population and has been used for the detection of vaccine adverse events.25 Furthermore, this report contains some of the first population-based rates for the incidence of rare events associated with thrombocytopenia. Rates may be useful when determining excess morbidity linked to COVID-19 and more generally for vaccine adverse event surveillance.25–27

Author affiliations
1Preventive Medicine and Population Health, The University of Texas Medical Branch at Galveston, Galveston, Texas, USA
2Family Medicine, The University of Texas Medical Branch at Galveston, Galveston, Texas, USA
3Sealy Institute for Vaccine Sciences, University of Texas Medical Branch, Galveston, Texas, USA
4Population Health Sciences MD/PhD Program, The University of Texas Medical Branch at Galveston, Galveston, Texas, USA
5Epidemiology and Biostatistics Consulting Unit, Office of Biostatistics, University of Texas Medical Branch, Galveston, Texas, USA

Contributors SCW acquired grant funding for this project, contributed significantly to the drafting of the manuscript, was responsible for the design and implementation of the project, and is the guarantor of the study. JB and GSW contributed to the design and implementation of the project. LP provided clinical expertise on the project and helped to draft the manuscript. JD helped to draft manuscript, contributed clinical expertise and systematically reviewed and abstracted all previous studies with incidence estimates. LC was the primary data coder and was responsible for incidence estimation.

Funding This work was funded in part by a grant from the Texas Academy of Family Physicians Foundation (SCW: Does Statin Therapy Reduce VTE Due To Hormone Therapy In Post-Menopausal Women?) and in part by the Sealy Institute for Vaccine Sciences (SCW, LP: Incidence of Thrombotic Events in a Large, Insured US Population).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the University of Texas Medical Branch at Galveston’s Institutional Review Board (IRB #20-0313). De-identified data in large insurance claims database.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. All data relevant to the study are included in the article or uploaded as supplementary information. No further data are available. Data are proprietary of Optum Health Systems. However, requests for re-analyses will be considered.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID ids
Susan C Weller http://orcid.org/0000-0002-6095-736X
Laura Porterfield http://orcid.org/0000-0001-9144-144X
John Davis http://orcid.org/0000-0002-8137-0323
Gregg S Wilkinson http://orcid.org/0000-0002-2657-3902
Lu Chen http://orcid.org/0000-0002-6867-2618
Jacques Baillargeon http://orcid.org/0000-0002-3297-653X

REFERENCES


23 Handley JD, Emsley HC. Validation of ICD-10 codes shows intracranial venous thrombosis incidence to be higher than previously reported. *Health Inf Manag* 2020;49:58–61.


