

Chart validation of inpatient International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) administrative diagnosis codes for venous thromboembolism (VTE) among intravenous immune globulin (IGIV) users in the Sentinel Distributed Database

Eric M. Ammann, PhD^{a,*}, Adam Cuker, MD^b, Ryan M. Carnahan, PharmD^a, Usha S. Perepu, MBBS^{c,d}, Scott K. Winiecki, MD^e, Marin L. Schweizer, PhD^{c,f}, Charles E. Leonard, PharmD^b, Candace C. Fuller, PhD^g, Crystal Garcia, MPH^g, Cole Haskins, BA^{a,c,h}, Elizabeth A. Chrischilles, PhD^a

Abstract

The Sentinel Distributed Database (SDD) is a database of patient administrative healthcare records, derived from insurance claims and electronic health records, sponsored by the US Food and Drug Administration for evaluation of medical product outcomes. There is limited information on the validity of diagnosis codes for acute venous thromboembolism (VTE) in the SDD and administrative healthcare data more generally.

In this chart validation study, we report on the positive predictive value (PPV) of inpatient administrative diagnosis codes for acute VTE—pulmonary embolism (PE) or lower-extremity or site-unspecified deep vein thrombosis (DVT)—within the SDD. As part of an assessment of thromboembolic adverse event risk following treatment with intravenous immune globulin (IGIV), charts were obtained for 75 potential VTE cases, abstracted, and physician-adjudicated.

VTE status was determined for 62 potential cases. PPVs for lower-extremity DVT and/or PE were 90% (95% CI: 73–98%) for principal-position diagnoses, 80% (95% CI: 28–99%) for secondary diagnoses, and 26% (95% CI: 11–46%) for position-unspecified diagnoses (originating from physician claims associated with an inpatient stay). Average symptom onset was 1.5 days prior to hospital admission (range: 19 days prior to 4 days after admission).

PPVs for principal and secondary VTE discharge diagnoses were similar to prior study estimates. Position-unspecified diagnoses were less likely to represent true acute VTE cases.

Abbreviations: DVT = deep vein thrombosis, FDA = US Food & Drug Administration, HPV = human papilloma virus, ICD-9-CM = International Classification of Diseases, Ninth Revision, Clinical Modification, IGIV = intravenous immune globulin, IRB = Institutional Review Board, PE = pulmonary embolism, PPV = positive predictive value, SDD = Sentinel Distributed Database, UHDDS = uniform hospital discharge data set, VTE = venous thromboembolism.

Keywords: deep vein thrombosis, healthcare administrative data, positive predictive value, pulmonary embolism, validation, venous thromboembolism

Editor: Ting-Shuo Huang.

Funding: This work was supported by the US Food and Drug Administration and Department of Health and Human Services [HHSF223200910006]; Sentinel Coordinating Center [HHSF223201400030].

Conflict of Interest: EMA is now employed in Johnson & Johnson's medical device epidemiology research division. The analyses for this project, as well as the drafting of the project report and associated manuscripts, were completed prior to his start in that role. CH has received compensation from Biotest Pharmaceuticals for blood plasma donations used in the manufacturing of intravenous immunoglobulin products. Other workgroup members report no conflicts.

Supplemental Digital Content is available for this article.

^a College of Public Health, ^b Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, ^c Carver College of Medicine, University of Iowa, ^d University of Iowa Hospitals and Clinics, ^e Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland, ^f Iowa City VA Health Care System, ^g Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts, ^h Medical Scientist Training Program, University of Iowa, Iowa City, Iowa.

* Correspondence: Eric M. Ammann, 145 North Riverside Drive, CPHB S400, The University of Iowa, Iowa City, IA 52242 (e-mail: eric-ammann@uiowa.edu).

Copyright © 2018 the Author(s). Published by Wolters Kluwer Health, Inc.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

Medicine (2018) 97:8(e9960)

Received: 25 September 2017 / Received in final form: 30 January 2018 / Accepted: 30 January 2018

<http://dx.doi.org/10.1097/MD.00000000000009960>

1. Introduction

This paper describes the results of a chart validation study of potential cases of acute venous thromboembolism (VTE) ascertained from the Sentinel Distributed Database (SDD). The SDD is a repository of longitudinal, patient-level medical, and prescription data from large health insurers and integrated care delivery systems that are converted to a common data format.^[1] The SDD and the Sentinel program are sponsored by the US Food and Drug Administration (FDA) for active safety surveillance of marketed medical products. As of August 2015, the SDD had 351 million person-years of longitudinal patient-level data for 193 million US health plan members from 2000 to 2015.^[2]

Periodic endpoint validation studies provide an essential foundation for the design and interpretation of epidemiologic studies based on administrative healthcare data.^[3] Prior validation studies conducted outside the SDD indicate that the positive predictive value (PPV) of administrative diagnosis codes for VTE can be variable, depending on factors such as the specific endpoint definition, data source, clinical setting (e.g., inpatient or outpatient), and patient characteristics such as age and other risk factors (e.g., recent hip replacement surgery).^[4] To date, the only data on the validity of VTE diagnosis codes within the SDD come from a study of the risk of incident VTE in young women following human papilloma virus (HPV) vaccination.^[5] To provide additional information on the validity of VTE diagnoses within the SDD, we report on the PPVs associated with inpatient VTE diagnoses identified as part of an intravenous immune globulin (IGIV) safety assessment.^[6]

2. Methods

2.1. Data sources and study population

The administrative healthcare data and medical charts used to ascertain and validate potential VTE cases came from 13 SDD Data Partners (i.e., large insurers and integrated care delivery systems) who participated in the protocol-based Sentinel assessment of thromboembolic events following immunoglobulin administration.^[6] Potential cases from the years 2006 to 2012 were selected for chart review if an inpatient VTE diagnosis code was recorded in the SDD up to 1 month following a non-specific IGIV treatment episode that occurred in an outpatient care

setting. A complete description of the criteria used to select potential cases can be found in the Appendix, <http://links.lww.com/MD/C129>.

IGIV is used in the treatment of primary and secondary immunoglobulin deficiencies, and a variety of inflammatory and autoimmune disorders (e.g., chronic demyelinating polyneuropathy and immune thrombocytopenia).^[7] Additional details concerning the design and objectives of the parent study have been described previously.^[6] To provide additional context, descriptive characteristics on the potential VTE cases for whom chart review was possible are provided in Table 1. These health conditions were defined as previously described in the study protocol.^[6]

The data presented in this paper were collected as part of a public health surveillance activity conducted under the auspices of the FDA Sentinel Initiative. For this reason, the collection and analysis of these data did not qualify as human subjects research under the Common Rule and were not subject to Institutional Review Board (IRB) review.^[8–10] The study consisted of a secondary analysis of existing healthcare records, and obtaining informed consent from the included patients was not required.

2.2. Case identification and chart retrieval

The endpoint definition used to identify potential VTE cases from the SDD consisted of any of the following International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes for lower-extremity or site-unspecified deep vein thrombosis (DVT) or pulmonary embolism (PE) originating from an inpatient hospital stay: 415.1x, 451.1x, 453.1, 453.2, 453.4x, and 453.9. Associated descriptions for these diagnosis codes are listed in Appendix Table A1, <http://links.lww.com/MD/C129>. We did not include diagnosis codes for upper-extremity DVTs because they are often attributable to central lines, a potential source of time-varying confounding for the IGIV risk assessment.

Within the Sentinel Common Data Model, diagnosis codes associated with inpatient encounters are categorized as principal, secondary, or “unable to classify” (i.e., position unspecified). These classifications reflect standard coding practices and the addition of a third category to accommodate heterogeneity across Sentinel Data Partners in how encounters and coding positions

Table 1

Criteria for adjudication of potential acute venous thromboembolism cases.

	Pulmonary embolism	Deep vein thrombosis
Definite	Confirmed by pulmonary angiography, spiral CT scan, MRI scan, or pathology	Confirmed by conventional venography, compression/duplex ultrasound, CT scan/CT angiography or pathology
Probable	If above tests not performed or were indeterminate, but ventilation-perfusion scan findings were of high probability	If above tests not performed or were indeterminate, but impedance plethysmography, radionuclide venography, or radiolabeled fibrinogen scan test results were reported as positive
Possible	If all of the above tests were not performed or were indeterminate and 2 of the following criteria were satisfied: medical record indicates physician-diagnosed acute DVT, signs or symptoms of acute DVT were documented and the patient underwent therapy with anticoagulants, or an IVC filter was placed.	If all of the above tests were not performed or were indeterminate and 2 of the following criteria were satisfied: medical record indicates physician-diagnosed acute DVT, signs or symptoms of acute DVT were documented and the patient underwent therapy with anticoagulants, or an IVC filter was placed.
No VTE or VTE status unknown	Based on the judgment of the physician adjudicator, potential cases not meeting diagnostic criteria for confirmation as a definite, probable, or possible VTE were classified as no acute VTE or as having an unknown status due to incomplete documentation or ambiguity.	

CT=computed tomography; HPV=human papilloma virus; IVC=inferior vena cava; MRI=magnetic resonance imaging.

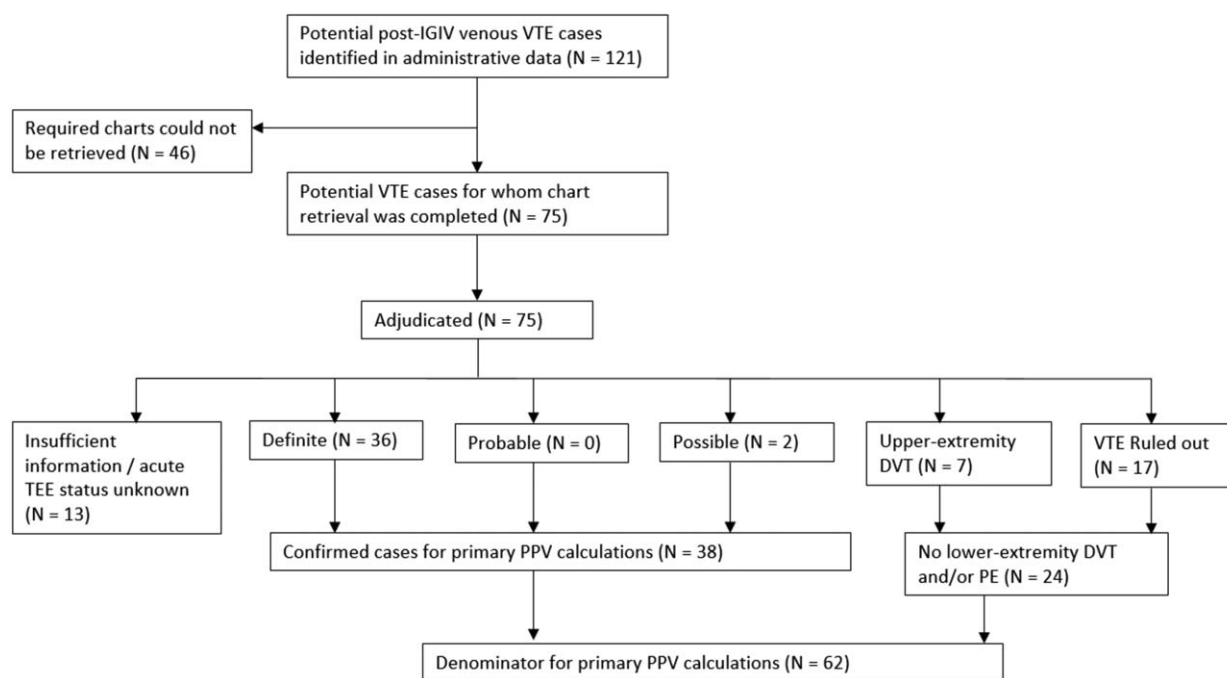


Figure 1. Disposition of potential acute venous thromboembolism (VTE) cases identified in the Sentinel Distributed Database (SDD).

are defined. Under Uniform Hospital Discharge Data Set (UHDDS) guidelines used by US hospitals and insurers,^[11] inpatient diagnoses are coded as follows:

- Principal diagnosis: The condition established after study to be chiefly responsible for occasioning the admission of the patient to the hospital
- Secondary diagnosis: A condition also present on admission, that developed during the hospital stay, or that influenced the care of the patient or length of stay

In the SDD, there are also position-unspecified diagnoses that cannot be classified as principal or secondary. These diagnoses codes may represent diagnoses originating from non-facility claims associated with an inpatient stay, e.g., a physician services claim submitted separately from the facility claim. Codes of this type generally come from claims-based Data Partners.

Qualifying inpatient encounters with a VTE diagnosis code listed in a principal or unspecified position were selected for review. Secondary inpatient VTE diagnoses were excluded because the VTE safety assessment did not focus on inpatient IGIV treatments or VTEs that developed during a hospital stay. (Secondary diagnoses are more likely to represent conditions that develop during a hospital stay, and VTE often arises during hospitalization due to immobility, injury, or surgery.) For a small number of cases with position-unspecified VTE diagnoses, a secondary VTE diagnosis was also recorded during the same index VTE hospital encounter. We report results separately for these cases because the presence of a secondary inpatient VTE diagnosis may be associated with a higher PPV for acute VTE.

For each potential VTE case meeting eligibility criteria, Sentinel Data Partners were asked to retrieve a medical chart corresponding to the hospital encounter during which the VTE diagnosis was recorded. In this validation report, we restricted the denominator for our PPV calculations to the subsample of potential cases for

whom we received a chart that was sufficiently complete to determine whether an acute VTE occurred (Fig. 1).

2.3. Chart abstraction

A trained nurse abstractor (LP or KP) reviewed the medical chart (s) associated with the index VTE hospital encounter. The abstractor recorded information concerning symptom onset, clinician notes, diagnostic testing including D-dimer and imaging results, and other factors relevant for the IGIV safety assessment.

2.4. Adjudication

Completed abstraction forms (and copies of the medical charts, if necessary) were reviewed by a board-certified hematologist (AC or USP). Adjudication criteria for acute VTE, shown in Table 1, were based on the clinical definitions used in the Worcester Venous Thromboembolism Study.^[12] Based on the documentation available in the charts, potential cases were adjudicated as a definite, probable, possible acute VTE, no acute VTE, or as unknown due to insufficient information.

For reasons discussed previously, upper-extremity DVTs were excluded from the study endpoint definition in the parent study on IGIV safety. For our primary PPV calculations presented in Table 2, confirmed upper-extremity DVTs were counted as false positives. Since upper-extremity DVTs may be of interest for other research questions, we also present PPV statistics with these events counted as true positives in Appendix Table A2, <http://links.lww.com/MD/C129>.

2.5. Positive predictive value (PPV) calculation

We calculated the PPV of the VTE diagnoses codes identified in the SDD by dividing the number of confirmed VTE cases (definite, probable, or possible) by the total number of cases for whom a

Table 2**Characteristics of potential venous thromboembolism (VTE) cases (N = 75) identified in the Sentinel Distributed Database.**

Covariate	N (%)
Demographics	
Age	
0–19 y	2 (3%)
20–39 y	9 (12%)
40–59 y	33 (44%)
60–79 y	25 (33%)
80+ y	6 (8%)
Sex	
Female	33 (44%)
Male	42 (56%)
Possible indication for IGIV use*	
Autoimmune/inflammatory condition	52 (69%)
Immune deficiency	49 (65%)
Infection	19 (25%)
Bone marrow or hematopoietic stem cell transplant	11 (15%)
Other indication	23 (31%)
Factors associated with VTE risk*	
History of VTE	18 (24%)
Oral anticoagulant use	14 (19%)
Hospitalization	40 (53%)
Condition associated with impaired mobility	43 (57%)
Cancer	42 (56%)
Major cardiovascular risk factors*	
Myocardial infarction	9 (12%)
Angina	14 (19%)
Atrial fibrillation or flutter	6 (8%)
Ischemic stroke	7 (9%)
Peripheral vascular disease	6 (8%)
Hypertension, uncomplicated	41 (55%)
Hypertension, complicated	10 (13%)
Diabetes mellitus	24 (32%)
Data partner type	
Insurer/claims-based	72 (96%)
Integrated healthcare delivery system	3 (4%)

* Possible indications, cardiovascular risk factors, and venous thromboembolism risk factors were assessed using diagnoses and procedures recorded in administrative data during the 183 days prior to the proximate IGIV date. The indication indicator variables are not mutually exclusive, so indication percentages may sum to >100%.

sufficiently complete chart was obtained for the index VTE hospitalization. Potential cases adjudicated as having an unknown VTE status due to insufficient information were removed from the denominator for the PPV calculation (Fig. 1). Exact binomial 95% confidence intervals (Clopper–Pearson) were calculated for the PPV estimates to quantify their precision.

2.6. Comparison of diagnosis date and VTE onset

For confirmed VTE cases (definite, probable, or possible), we also evaluated the degree of correspondence between the VTE diagnosis date recorded in the administrative data (in the SDD, this would be the admission date associated with the hospital encounter) and the date of initial onset of VTE signs/symptoms. VTE onset can be insidious and may not prompt all patients to seek immediate medical care; in addition, the initial diagnosis and treatment of VTE may first occur in an outpatient setting for some patients. For these reasons we also report on the distribution of the difference in days between the admission date for the index VTE hospital encounter, as recorded in the SDD, and the initial onset of VTE signs and/or symptoms, as recorded in the patient's chart.

3. Results

One hundred twenty-one post-IGIV VTE cases were identified in the SDD in 2006 to 2012; required charts could be obtained for 75 (62%) of these patients. Common reasons that charts were unavailable included an inability to map the encounter record in the SDD to patient and provider identifiers required for chart requests, an inability to locate the medical chart corresponding to the requested encounter, and refusal by the healthcare provider. (See Appendix Table A3, <http://links.lww.com/MD/C129> for a complete list of reasons that charts were unobtainable.)

Of the 75 cases for which charts were available, 72 were from claims-based Data Partners, and 3 from integrated care delivery systems. The median patient age was 56 years; 44% were female. As described previously, all of the patients were identified as potential post-IGIV VTE cases; possible indications for IGIV are described in Table 2. Based on administrative data records from the 6 months prior to the index VTE hospital encounter, these patients appeared to have a high burden of VTE risk factors, including history of VTE (24%), recent hospitalization (53%), conditions associated with limited mobility (57%), and cancer (56%). Additional descriptive information on these patients is provided in Table 2.

Acute VTE status could be determined for 62 potential cases, of which 38 were confirmed by physician adjudicators (36 definite VTEs and 2 possible; see Fig. 1). For the 2 possible cases, definitive diagnostic testing results were not available in the charts, but a physician diagnosis was documented and anticoagulation therapy was initiated. The PPVs for the inpatient VTE diagnoses recorded in SDD were 61% overall (38/62, 95% CI: 48–73%), 90% (27/30, 95% CI: 73–98%) for principal-position diagnoses, 80% (4/5, 95% CI: 28–99%) for secondary diagnoses, and 26% (7/27, 95% CI: 11–46%) for position-unspecified diagnoses. Additional PPV statistics stratified by coding position, ICD-9-CM diagnosis code, Data Partner type, prior acute VTE diagnosis, and prior oral anticoagulant use are provided in Table 3. PPVs were higher among patients without a history of VTE or anticoagulant use recorded prior to the index hospital encounter.

Upper-extremity DVTs were not included in our study endpoint definition because they are often attributable to central lines, a potential source of time-varying confounding for the IGIV risk assessment. Seven patients had upper-extremity DVTs, which are counted as false positives in the estimates above. If these cases were counted as confirmed events, the overall PPVs for the endpoint definition would be 73% (45/62, 95% CI: 60–83%) overall, 93% (28/30, 95% CI: 78–99%) for principal diagnoses, 80% (4/5, 95% CI: 28–99%) for secondary diagnoses, and 48% (13/27, 95% CI: 29–68%) for position-unspecified diagnoses. PPVs for PE codes improved only slightly (from 68% to 71% overall), whereas PPVs for DVT codes improved from 54% to 75% overall with the inclusion of the upper-extremity DVTs as confirmed cases. A more detailed breakdown by subgroup is provided in Appendix Table A2, <http://links.lww.com/MD/C129>.

For a significant number of the confirmed VTE cases, VTE onset occurred prior to the day of hospital admission. Figure 2 shows the difference in days between the VTE symptom onset date (i.e., the first date on which clinical sign/symptoms consistent with VTE were reported) and the index hospital admission date for 32 of the 38 confirmed VTE cases. For the other 6 confirmed cases, the onset date was indeterminate. On average (median value), VTE onset was 1.5 days prior to the admission date

Table 3**Positive predictive values (PPVs)* associated with inpatient administrative diagnosis codes for venous thromboembolism (VTE) by position.**

	PPVs for all potential VTE cases (N=62)	PPVs for principal position VTE diagnoses (N=30)	PPVs for secondary VTE diagnoses (N=5) [†]	PPVs for position-unspecified VTE diagnoses (N=27) [†]
All VTE codes	61% (38/62, 95% CI: 48–73%)	90% (27/30, 95% CI: 73–98%)	80% (4/5, 95% CI: 28–99%)	26% (7/27, 95% CI: 11–46%)
By diagnosis code recorded in administrative data				
Deep vein thrombosis codes	54% (15/28, 95% CI: 34–72%)	90% (9/10, 95% CI: 55–100%)	100% (1/1, 95% CI: 3–100%)	29% (5/17, 95% CI: 10–56%)
451.11	0% (0/1, 95% CI: 0–98%)	–	–	0% (0/1, 95% CI: 0–98%)
451.19	25% (1/4, 95% CI: 1–81%)	–	–	25% (1/4, 95% CI: 1–81%)
451.2	0% (0/1, 95% CI: 0–98%)	–	–	0% (0/1, 95% CI: 0–98%)
451.9	–	–	–	–
453.1	–	–	–	–
453.2	–	–	–	–
453.40	44% (4/9, 95% CI: 14–79%)	50% (1/2, 95% CI: 1–99%)	–	43% (3/7, 95% CI: 10–82%)
453.41	88% (7/8, 95% CI: 47–100%)	100% (5/5, 95% CI: 48–100%)	100% (1/1, 95% CI: 3–100%)	50% (1/2, 95% CI: 1–99%)
453.42	100% (3/3, 95% CI: 29–100%)	100% (3/3, 95% CI: 29–100%)	–	–
453.9	0% (0/2, 95% CI: 0–84%)	–	–	0% (0/2, 95% CI: 0–84%)
Pulmonary embolism codes	68% (23/34, 95% CI: 49–83%)	90% (18/20, 95% CI: 68–99%)	75% (3/4, 95% CI: 19–99%)	20% (2/10, 95% CI: 3–56%)
415.11	–	–	–	–
415.12	0% (0/1, 95% CI: 0–98%)	–	–	0% (0/1, 95% CI: 0–98%)
415.13	–	–	–	–
415.19	70% (23/33, 95% CI: 51–84%)	90% (18/20, 95% CI: 68–99%)	75% (3/4, 95% CI: 19–99%)	22% (2/9, 95% CI: 3–60%)
Data partner type				
Insurer/claims-based	59% (35/59, 95% CI: 46–72%)	89% (24/27, 95% CI: 71–98%)	80% (4/5, 95% CI: 28–99%)	26% (7/27, 95% CI: 11–46%)
Integrated care delivery systems	100% (3/3, 95% CI: 29–100%)	100% (3/3, 95% CI: 29–100%)	–	–
Acute VTE code observed in 183 days prior to index hospitalization				
No prior VTE	70% (35/50, 95% CI: 55–82%)	93% (26/28, 95% CI: 76–99%)	100% (3/3, 95% CI: 29–100%)	32% (6/19, 95% CI: 13–57%)
Prior VTE	25% (3/12, 95% CI: 5–57%)	50% (1/2, 95% CI: 1–99%)	50% (1/2, 95% CI: 1–99%)	13% (1/8, 95% CI: 0–53%)
Oral anticoagulant prescription filled in 183 days prior to index hospitalization				
No use	69% (35/51, 95% CI: 54–81%)	93% (25/27, 95% CI: 76–99%)	100% (4/4, 95% CI: 40–100%)	30% (6/20, 95% CI: 12–54%)
Oral anticoagulant use	27% (3/11, 95% CI: 6–61%)	67% (2/3, 95% CI: 9–99%)	0% (0/1, 95% CI: 0–98%)	14% (1/7, 95% CI: 0–58%)

* Statistics reported in this table reflect the PPV of administrative ICD-9-CM VTE diagnosis codes for a confirmed (definite, probable, or possible) acute VTE. Patients with a classification of insufficient information/acute VTE status unknown were removed from the denominator for the PPV calculations and not included in this table.

[†] Though the venous VTE endpoint definition included only principal or position-unspecified VTE diagnosis codes, data for the 5 potential cases with both secondary and position-unspecified diagnosis codes are reported separately as secondary diagnoses in this table. In the SDD, position-unspecified diagnoses generally originate from physician/provider claims rather than institutional claims/records and may have lower PPVs than principal or secondary inpatient diagnoses.

(range: 19 days prior to 4 days after admission). It should be noted that the more remote symptom onset dates (more than a few days prior to hospital admission) may be less precise, as they are based on patient recall and self-report.

4. Discussion

In this chart validation study, which relied on data from a protocol-based assessment of the risk of thromboembolic events following IGIV treatment,^[6] we evaluated the validity of inpatient administrative diagnosis codes for PE or lower-extremity or position-unspecified DVT within the SDD. We found that PPVs were high for principal diagnoses (90%, 95% CI: 73–98%), somewhat lower for secondary diagnoses (80%, 95% CI: 28–99%) though uncertain due to data sparseness, and low for position-unspecified diagnoses (26%, 95% CI: 11–46%). We also found that the PPV of acute VTE diagnosis codes was considerably lower among patients with a prior VTE diagnosis code (recorded in any setting during the prior 6 months) than in patients without such a diagnosis (25% vs 70%). This finding suggests that an extended washout period for VTE outcome codes prior to the start of follow-up may be preferable in future SDD studies of VTE risk if chart confirmation of study outcomes is not possible.

The high PPV of principal-position inpatient ICD-9-CM codes for DVT and pulmonary embolism is consistent with the results of prior chart validation studies, which have reported PPVs of

80% to 100%.^[4,13] Secondary inpatient DVT and PE codes have generally been associated with lower PPVs (30–80%).^[13–17] Our PPV point estimate for secondary diagnoses falls within the upper end of that range; however, the confidence intervals for that estimate are wide. Between-study variation in PPV estimates^[12,18–28] may be explained by differences in the coding algorithms used to ascertain potential cases, chart-validation criteria, patient populations, the nature of the administrative data used to ascertain potential cases, and sampling error. In a previous Sentinel safety assessment that evaluated the association between the Gardasil vaccine and VTE,^[5] the reported PPV for inpatient VTE diagnoses was 64%, similar to our overall VTE PPV estimate of 61%.

An important finding from our study was that position-unspecified inpatient VTE diagnoses appear to have a lower PPV for acute VTEs. In the SDD, a position-unspecified diagnosis can represent a diagnosis associated with a hospital stay that was not recorded on a facility claim, such as a diagnosis recorded on a professional claim. In the absence of an acute VTE discharge diagnosis recorded on an inpatient facility claim, it appears that a large number of these diagnoses reflect one or more of the following: a history of VTE, chronic VTE, consideration of acute VTE as part of a differential diagnosis, and/or diagnostic testing for acute VTE. In the future, researchers may consider excluding position-unspecified inpatient diagnosis codes (and their analogs outside the SDD) from their VTE endpoint definitions, particularly if chart review for case confirmation is not possible.

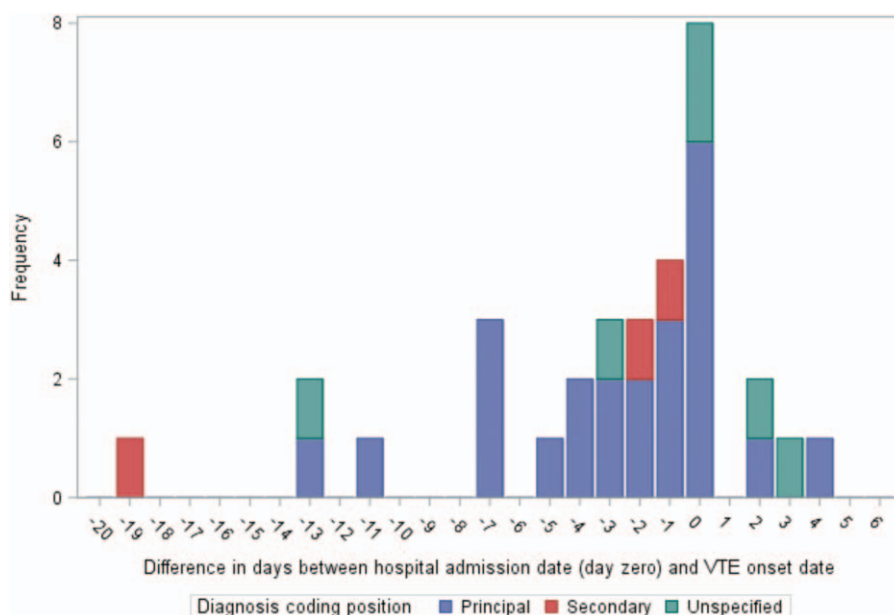


Figure 2. Distribution of the difference in days* between acute venous thromboembolism (VTE) symptom onset and the index hospital admission date among confirmed VTE cases. *A negative number indicates that VTE symptom onset occurred prior to the hospital admission date (day 0). VTE symptom onset is shown above for the 32 confirmed VTE cases for whom the date of onset could be determined; for 6 confirmed cases, the date of VTE symptom onset was indeterminate.

It is worth noting that the estimated PPV for position-unspecified VTE diagnoses increased meaningfully, from 26% to 48%, when upper-extremity DVT cases were counted as confirmed events. One possible explanation is that coding may be less precise for physician/provider claims compared with hospital/institutional claims.

A limitation of our study was that medical charts were unobtainable for 38% of potential post-IGIV cases identified in the SDD. However, the typical reasons that charts were unavailable (e.g., unable to link SDD records to patient or provider identifiers) did not give us reason to suspect that our analyzable sample was systematically different than the total set of potential cases identified. Another limitation was that our sample of potential post-IGIV VTE cases represents a unique patient group with a variety of indications for IGIV use, including a number of rare autoimmune/inflammatory conditions and immune disorders. This may limit the generalizability of our results to patient populations that differ significantly in terms of health status, age, comorbidity, history of VTE, and other VTE risk factors such as recent hospitalization and impaired mobility.

To our knowledge, a novel aspect of this validation report is the inclusion of data concerning how well the VTE diagnosis date recorded in the administrative data corresponded to the date of initial VTE symptom onset recorded in the patient’s medical chart. Among the 32 confirmed VTE cases with information on event timing, initial symptom onset occurred on the hospital admission date in 8 cases, but occurred days or weeks prior to admission in a significant number of cases (median date of onset = 1.5 days prior to admission). In a traditional cohort study with sufficiently long follow-up, this discrepancy in dates may be unimportant. However, event timing is particularly important in self-controlled studies and other assessments of acute, transient risk following exposure. For such studies, the fact that the hospital admission date and symptom onset date coincided exactly for only 25% of confirmed cases could be an important

consideration. In future validation studies of VTE and other health outcomes of interest with the potential for an insidious onset and/or delay between the onset of symptoms and initial presentation for medical care, we would encourage investigators to report on how well the diagnosis dates recorded in the administrative data aligned with the true event start date or initial onset of signs and symptoms.

Our study indicates that principal-position inpatient ICD-9-CM diagnosis codes for DVT and PE recorded in the SDD have a high PPV (90%) for acute VTE, similar to what past validation studies have reported for VTE administrative diagnosis codes in other populations. Position-unspecified inpatient diagnosis codes (i.e., diagnoses from professional claims associated with an inpatient encounter) may be less useful for the identification of acute VTE unless chart confirmation is feasible. To inform future VTE surveillance activities, additional research is needed to evaluate the validity of secondary diagnosis codes in the SDD, and how often principal and secondary diagnoses reflect out-of-hospital versus in-hospital events.

Acknowledgments

The authors thank contributors from The University of Iowa (Michael Mueller, Nicholas Rudzianski, and James Torner), Harvard Medical School and Harvard Pilgrim Health Care Institute (Madelyn Pimentel, Meghan Baker, and Casey Covarrubias), Telligen-West Des Moines (Lois Pedelty and Kim Price), and Kaiser Permanente (Bruce Fireman).

References

[1] Curtis LH, Weiner MG, Boudreau DM, et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. *Pharmacoepidemiol Drug Saf* 2012;21(Suppl):23–31.
 [2] Sentinel Coordinating Center. Snapshot of Database Statistics; 2017. Available at: <https://www.sentinelinitiative.org/sentinel/snapshot-data-base-statistics>. Accessed September 10, 2017.

- [3] Carnahan RM. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative data: summary of findings and suggestions for future research. *Pharmacoepidemiol Drug Saf* 2012;21(Suppl):90–9.
- [4] Tamariz L, Harkins T, Nair V. A systematic review of validated methods for identifying venous thromboembolism using administrative and claims data. *Pharmacoepidemiol Drug Saf* 2012;21(Suppl):154–62.
- [5] Yih WK, Greene SK, Zichittella L, et al. Mini-Sentinel medical product assessment: Evaluation of the risk of venous thromboembolism after Gardasil vaccination; 2015. Available at: http://mini-sentinel.org/work_products/Assessments/Mini-Sentinel_PRISM_Gardasil-and-Venous-Thromboembolism-Report.pdf. Accessed September 21, 2015.
- [6] Thromboembolic Events After Immunoglobulin Administration Workgroup. Mini-Sentinel Assessment Protocol: Thromboembolic Events After Immunoglobulin Administration: Version 3.0; 2015. Available at: https://www.sentinelinitiative.org/sites/default/files/Drugs/Assessments/Mini-Sentinel_Thromboembolic-Events-After-Immunoglobulin-Administration-Protocol_0.pdf. Accessed September 10, 2017.
- [7] Orange JS, Hossny EM, Weiler CR, et al. Use of intravenous immunoglobulin in human disease: a review of evidence by members of the Primary Immunodeficiency Committee of the American Academy of Allergy, Asthma and Immunology. *J Allergy Clin Immunol* 2006;117(Suppl):S525–53.
- [8] McGraw D, Rosati K, Evans B. A policy framework for public health uses of electronic health data. *Pharmacoepidemiol Drug Saf* 2012;21(Suppl):18–22.
- [9] Rosati K, Evans B, McGraw D. HIPAA and Common Rule Compliance in the Mini-Sentinel Pilot. Unpublished White Paper; 2012.
- [10] Forrow S, Campion DM, Herrinton LJ, et al. The organizational structure and governing principles of the Food and Drug Administration's Mini-Sentinel pilot program. *Pharmacoepidemiol Drug Saf* 2012;21(Suppl):12–7.
- [11] Health information policy council. 1984 revision of the Uniform Hospital Discharge Data Set—HHS. Notice. *Fed Regist* 1985;50:31038–40.
- [12] Spencer FA, Emery C, Lessard D, et al. The Worcester Venous Thromboembolism study: a population-based study of the clinical epidemiology of venous thromboembolism. *J Gen Intern Med* 2006;21:722–7.
- [13] White RH, Garcia M, Sadeghi B, et al. Evaluation of the predictive value of ICD-9-CM coded administrative data for venous thromboembolism in the United States. *Thromb Res* 2010;126:61–7.
- [14] Leibson CL, Needleman J, Buerhaus P, et al. Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology Project VTE cohort. *Med Care* 2008;46:127–32.
- [15] Henderson KE, Recktenwald A, Reichley RM, et al. Clinical validation of the AHRQ postoperative venous thromboembolism patient safety indicator. *Jt Comm J Qual Patient Saf* 2009;35:370–6.
- [16] Zhan C, Battles J, Chiang YP, et al. The validity of ICD-9-CM codes in identifying postoperative deep vein thrombosis and pulmonary embolism. *Jt Comm J Qual Patient Saf* 2007;33:326–31.
- [17] Fang MC, Fan D, Sung SH, et al. Validity of using inpatient and outpatient administrative codes to identify acute venous thromboembolism: The CVRN VTE Study. *Med Care* 2017;55:e137–43.
- [18] Gerstman BB, Freiman JP, Hine LK. Use of subsequent anticoagulants to increase the predictive value of Medicaid deep venous thromboembolism diagnoses. *Epidemiology* 1990;1:122–7.
- [19] Willey VJ, Bullano MF, Hauch O, et al. Management patterns and outcomes of patients with venous thromboembolism in the usual community practice setting. *Clin Ther* 2004;26:1149–59.
- [20] Spencer FA, Emery C, Joffe SW, et al. Incidence rates, clinical profile, and outcomes of patients with venous thromboembolism. The Worcester VTE study. *J Thromb Thrombolysis* 2009;28:401–9.
- [21] Cushman M, Tsai AW, White RH, et al. Deep vein thrombosis and pulmonary embolism in two cohorts: the longitudinal investigation of thromboembolism etiology. *Am J Med* 2004;117:19–25.
- [22] Spyropoulos AC, Hurley JS, Ciesla GN, et al. Management of acute proximal deep vein thrombosis: pharmacoeconomic evaluation of outpatient treatment with enoxaparin vs inpatient treatment with unfractionated heparin. *Chest* 2002;122:108–14.
- [23] White RH, Gettner S, Newman JM, et al. Predictors of rehospitalization for symptomatic venous thromboembolism after total hip arthroplasty. *N Engl J Med* 2000;343:1758–64.
- [24] White RH, Romano PS, Zhou H, et al. Incidence and time course of thromboembolic outcomes following total hip or knee arthroplasty. *Arch Intern Med* 1998;158:1525–31.
- [25] McCarthy EP, Iezzoni LI, Davis RB, et al. Does clinical evidence support ICD-9-CM diagnosis coding of complications? *Med Care* 2000;38:868–76.
- [26] White RH, Brickner LA, Scannell KA. ICD-9-CM codes poorly identified venous thromboembolism during pregnancy. *J Clin Epidemiol* 2004;57:985–8.
- [27] Anderson FA Jr, Wheeler HB, Goldberg RJ, et al. A population-based perspective of the hospital incidence and case-fatality rates of deep vein thrombosis and pulmonary embolism. The Worcester DVT Study. *Arch Intern Med* 1991;151:933–8.
- [28] Lawthers AG, McCarthy EP, Davis RB, et al. Identification of in-hospital complications from claims data. Is it valid? *Med Care* 2000;38:785–95.