

# Positive predictive value and misclassification of diagnosis of pulmonary embolism and deep vein thrombosis in Swedish patient registries

Ludvig Öhman  
Magdalena Johansson  
Jan-Håkan Jansson  
Marcus Lind  
Lars Johansson

Department of Public Health and  
Clinical Medicine, Skellefteå Research  
Unit, Umeå University, Umeå,  
Västerbotten, Sweden

**Purpose:** To validate diagnoses of pulmonary embolism (PE) and deep vein thrombosis (DVT) in administrative registries. We also estimated the frequency of misclassified PE and DVT events.

**Patients and methods:** A registry search for ICD codes representing PE and DVT was performed between 1985 and 2014 in a large population-based cohort in northern Sweden. An additional search using an extended set of ICD codes was performed to identify misclassified events. Diagnoses were validated manually by reviewing medical records and radiology reports.

**Results:** Searching ICD codes in the National Patient Registry and Cause of Death Registry identified 2,450 participants with a first-time diagnosis of PE or DVT. The positive predictive value (PPV) for a diagnosis of PE or DVT was 80.7% and 59.2%, respectively. For the period of 2009 to 2014, the PPV was higher for PE (85.8%) but lower for DVT (54.1%). Misclassification occurred in 16.4% of DVT events and 1.1% of PE events.

**Conclusion:** Registry-based data on PE, especially in recent years, are of acceptable quality and can be considered for use in registry-based studies. For DVT, we found that data were of low quality in regards to both PPV and misclassification and should not be used without validation.

**Keywords:** pulmonary embolism, deep vein thrombosis, positive predictive value, International Classification of Diseases, validation

## Introduction

Administrative healthcare data, such as patient registries, are often used for research purposes. In studies based on national patient registries, large quantities of data can be collected and analyzed at relatively low cost. A limitation of registry-based studies is the risk of systematic errors when the studied disease does not correspond with the intended event. Therefore, validation of registries and their predictive value is essential for all events intended to be studied in this way. The positive predictive value (PPV) of the event to be studied has implications for the study design and whether the diagnosis needs to be validated. In addition, the degree to which the disease under study has been coded incorrectly indicates the risk of not identifying patients with the disease.

Validation studies have shown that the PPV can differ widely between different diagnoses; for example, the PPV of atrial fibrillation is 96% compared to 35% for herpes simplex encephalitis.<sup>1,2</sup> Previous studies on the PPV of a diagnosis of pulmonary embolism (PE) and deep vein thrombosis (DVT) have reported a range of 31% to 91% depending on which ICD codes are used and if diagnoses made at hospital wards or emergency departments are included.<sup>3-7</sup> PE or DVT events may be classified erroneously as another event (eg, other thromboembolic events) and be missed when

Correspondence: Lars Johansson  
Department of Public Health and Clinical  
Medicine, Skellefteå Research Unit,  
Department of Medicine, Skellefteå  
County Hospital, SE-931 86 Skellefteå,  
Sweden  
Tel +46 91 077 1000  
Fax +46 91 077 1657  
Email lars.johansson@vll.se

collecting data from registries. The extent and importance of such misclassification of disease is largely unknown, but may be considerable.

The methods for investigating and diagnosing venous thromboembolism (VTE) have improved over the last decades with the advent of better and more accessible radiological imaging, such as computer tomography (CT) angiography, improved sonographic imaging with Doppler, and the widespread use of diagnostic aids, such as Wells Score and D-dimer. However, whether these improvements have influenced the accuracy of diagnosis is unknown. The primary aim of the present study was to estimate the PPV for ICD diagnosis codes of first-time PE and DVT and to investigate whether restricting the search to recent years increases the validity. The secondary aim was to estimate the proportion of valid PE and DVT events that are misclassified as other diseases.

## Methods

### Study population

The Venous Thromboembolism in Northern Sweden (VEINS) study is a prospective study on the risk markers for VTE.<sup>8,9</sup> The cohort comprises 108,413 participants, aged 30–60 years, who were invited to undergo a health examination between January 1, 1985, and September 5, 2014 as part of a large health screening program. The participation rate has been reported to be 65%.<sup>10</sup> The VEINS study has yielded a large and comprehensive population-based cohort that can be used for epidemiological research.<sup>11</sup>

### ICD code search

The Swedish National Patient Register (SNPR) is an administrative diagnosis registry containing information on all hospital inpatients, outpatients, and emergency department visits. The SNPR includes a main diagnosis, the primary reason for the hospitalization or visit, and secondary diagnoses determined by the treating physician.<sup>12</sup> Diagnoses are coded according to the ICD system. ICD-8 codes were used from 1985–1987, ICD-9 codes between 1988 and 1997, and ICD-10 codes from 1998 onwards. To find individuals whose death certificates stated that the cause of death was PE and/or DVT, we searched the Cause of Death Register (CDR). This registry contains information on all deceased Swedish citizens, including the main cause of death and contributing causes,<sup>13</sup> and includes deceased outside of hospital and diagnoses determined at autopsy. The diagnosis is determined by the treating physician or the pathologist performing the autopsy.

A search was performed in the SNPR and CDR for ICD codes corresponding to DVT and PE (see Appendix S1) using a unique 12-digit personal identifier, the Swedish civil registration number, for all participants in the study cohort. Information was obtained on the primary and secondary diagnostic codes for PE, DVT, or both, as well as the dates of these diagnoses. Only first-time PE or DVT events were included, and participants with a PE or DVT diagnosis prior to the health examination were excluded.

### Validation of diagnoses

All PE and DVT diagnoses were validated manually. Medical records, radiology reports, and/or autopsy reports were reviewed by three physicians with experience in the field of VTE. The reviewers were not blinded to the registered diagnoses. In the case of uncertainty, a consensus agreement was reached. PE was considered valid when confirmed by CT, pulmonary angiography, MRI, ventilation-perfusion lung scan, or autopsy. DVT was considered valid when confirmed by CT, venography, ultrasonography, MRI, or autopsy. Radiology reports stating suspected or possible PE or DVT were not considered valid. A diagnosis of PE or DVT from the CDR was confirmed only if an autopsy was performed. If participants had symptoms of PE and a verified DVT, they were considered to have both a valid PE and DVT event. Participants without conclusive radiology or autopsy reports were considered free of PE and DVT.

### Search for misclassified events

Misclassification occurred when a valid PE or DVT event was incorrectly coded as another ICD diagnosis. In a pilot study in 2005 (unpublished data), we searched for misclassified PE and DVT events using two additional registries, an anticoagulant treatment registry and the radiology registry (ultrasound or CT angiography). We identified patients without an ICD code for PE or DVT but with anticoagulant treatment for VTE or radiology reports indicating thromboembolism. We compiled a list of the most frequently used codes when PE and DVT were misclassified. To estimate the number of misclassified PE and DVT events in the present study, we searched the SNPR and CDR using this extended set of ICD codes (see Appendix S2). All additional diagnoses identified were manually reviewed and validated using the same protocol by the same reviewers as described above.

### Statistical analysis

We calculated the PPV as the ratio of the number of participants with a confirmed diagnosis after the validation to the

number of participants with a specific diagnosis in the registries. Valid PE or DVT events found through the extended ICD code search were defined as misclassified events. Misclassified events were presented as the proportion of all valid events found by searching the registries. The estimation of the 95% confidence interval (CI) assumed an approximation to the binominal distribution.

A separate analysis restricted to the time period between January 1, 2009, and September 5, 2014, was performed to estimate the potential effects of recent enhanced diagnostic accuracy and coding on PPV and misclassification.

## Ethics

All participants provided written informed consent at the time of their health examination. Ethic approval was granted for the project as part of the VEINS study by the Regional Ethics Review Board, Umeå, Sweden (Dnr: 06–162M §157/06. 2006-12-05).

## Results

Between January 1, 1985, and September 5, 2014, a total of 108,413 participants underwent health examinations and constitute the study cohort. A total of 355 individuals had a diagnosis of PE or DVT prior to their health examination and were excluded (Figure 1A). The present study population included the remaining 108,058 participants (51% female). The mean age at inclusion was 47 years and participants were followed for a median of 15.5 years.

The ICD code search identified 2,462 participants with a code for PE and/or DVT. Medical records were retrieved and reviewed for 2,450 (99.5%) of the participants. Of these, 955 (38.8%) had an ICD code for PE, 1,292 (52.5%) for DVT, and 203 (8.2%) for both PE and DVT. Validation confirmed a PE or DVT event in 1,771 participants, resulting in a PPV of 72.3% (95% CI: 70.3–74.1, Table 1 and Figure 1B). When analyzed separately, a PE event was confirmed in 934 of the 1,158 participants and a DVT event in 885 of the 1,495 participants, corresponding to a PPV of 80.7% (95% CI: 78.4–82.9) for a diagnosis of PE and 59.2% (95% CI: 56.7–61.7) for a diagnosis of DVT. The CDR alone identified 43 valid events, 41 of which were PE events. A separate analysis restricted to participants with diagnoses between January 1, 2009, and September 5, 2014, identified 1,261 participants with a first-time diagnosis of PE or DVT. The PPV for a diagnosis of PE or DVT was 71.1% (95% CI: 68.6–73.6), for PE 85.8% (95% CI: 83.1–88.5) and for DVT 54.1% (95% CI: 50.5–57.7).

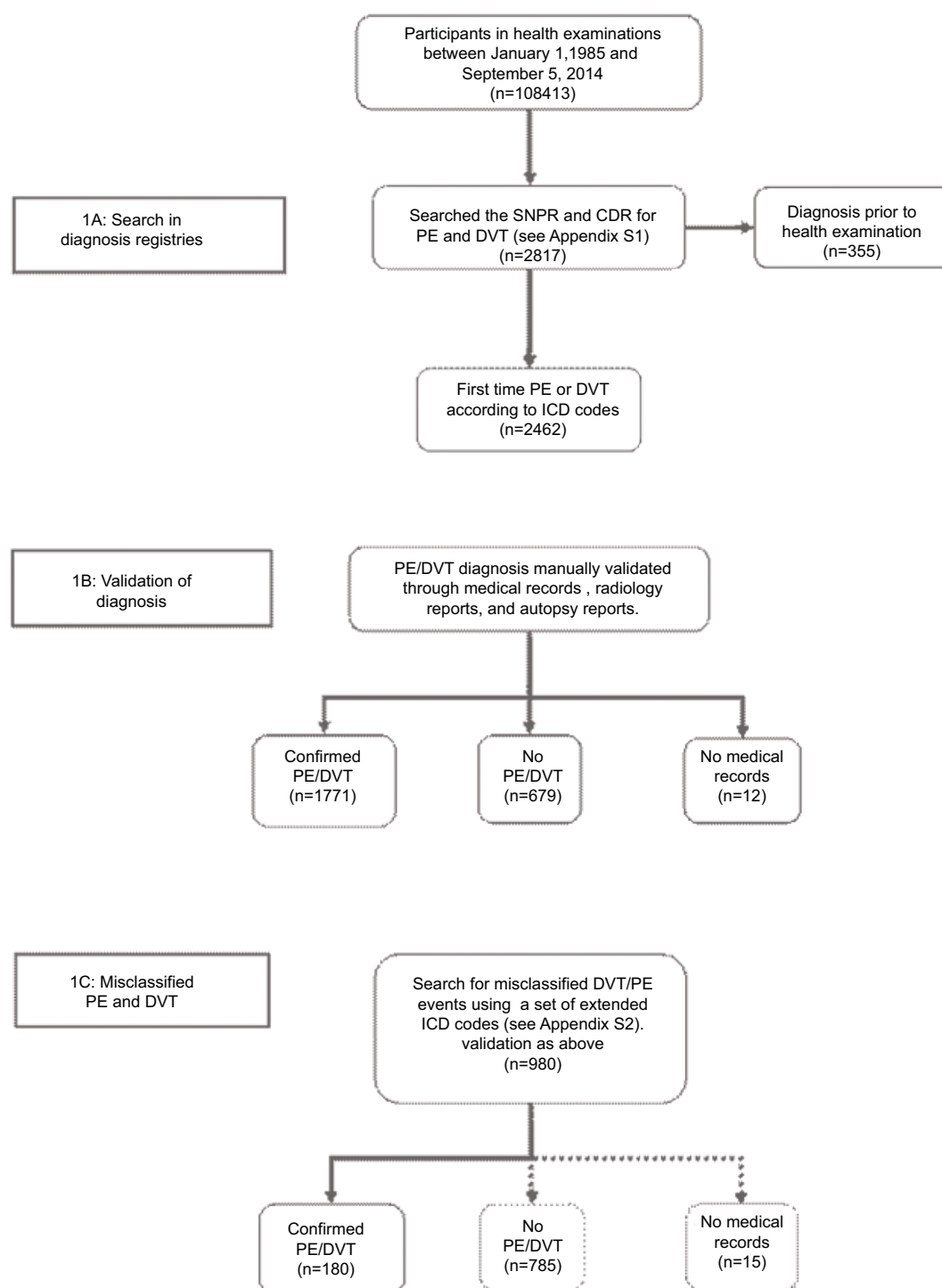
The extended ICD code search to identify potential misclassified PE or DVT events identified 980 additional participants whose medical records were reviewed (Table 2 and Figure 1C). The medical records could not be retrieved for 15 participants. Among the remaining 965 participants, a valid PE and/or DVT event could be confirmed for 180; 174 of these were misclassified DVT events and 10 were misclassified PE events. When restricting the observation period to 2009–2014, we identified 63 misclassified PE and/or DVT events (63 DVT and 4 PE events).

## Discussion

Over the whole study period, the PPV was 81% for PE and 59% for DVT. Restricting the analysis to the most recent 5 years resulted in a higher PPV of 86% for PE and a lower PPV of 54% for DVT. The search for misclassified PE or DVT events revealed that approximately 9% of all events were incorrectly coded, with more than 90% being DVT events.

Two large studies validating patient registries were performed between 1994 and 2006, both using the Danish National Patient Registry. The Diet, Cancer and Health study estimated a PPV of 67% for PE and 55% for DVT.<sup>5</sup> The second study investigated pediatric thrombosis diagnoses and found a PPV of 48% for PE and 51% for DVT.<sup>6</sup> Both studies showed substantial improvements in the PPV when restricting the diagnoses to those coded at wards. Similar results of poor predictive values for outpatient encounters between 2004 and 2010 were reported by Fang et al.<sup>14</sup> In our study, the overall PPV for PE (81%) was clearly higher and the PPV for DVT (59%) marginally higher than the PPVs reported in the two Danish studies. One possible explanation for the high PPV for PE is that patients with diagnosed PE were mostly admitted to the hospital for observation and initial treatment. Consequently, in the wards, a diagnosis of PE was made more often than a diagnosis of DVT.

A recent Danish study of a random sample of 100 patients between 2010 and 2012 reported a PPV of 90% for PE and 86% for DVT, and that inpatient diagnosis or primary diagnosis increased the PPV.<sup>7</sup> When restricting the time period to 2009–2014, we found that the PPV for PE improved from 81% to 86%, but the PPV for DVT decreased. These results imply that PE diagnoses in registries improved during the last decade, possibly due to improvement in both the diagnostic possibilities and the diagnostic coding. In 2006, hospitals introduced quality control for inpatient ICD diagnoses performed by specifically trained personnel. This may have improved the validity of the hospital registry and



**Figure 1** Overview of the study population in the ICD code search, diagnosis validation, and search for misclassified events.

**Abbreviations:** SNPR, Swedish National Patient Register; CDR, the Cause of Death Register; PE, pulmonary embolism; DVT, deep vein thrombosis.

subsequently improved the accuracy of PE diagnoses, but could only have had a marginal impact on the validity of DVT diagnoses, as they are mainly made in an outpatient setting. Improvements in radiological imaging for PE, such as the

introduction and more frequent use of CT angiography, may also have led to more conclusive reports and decreased the number of probable or possible events that were subsequently considered to be invalid in the present study. Our method for

**Table 1** Positive predictive value (PPV) of diagnoses for PE and DVT in the Swedish National Patient Registry and Cause of Death Register validated by a review of medical records, radiology reports, or autopsy

Time period	1985–2014		Valid/Diagnoses	2009–2014		Valid/Diagnoses
	PPV%	95% CI		PPV%	95% CI	
DVT or PE	72.3	70.3–74.1	1771/2450	71.1	68.6–73.6	896/1261
PE	80.7	78.4–82.9	934/1158	85.8	83.1–88.5	544/634
DVT	59.2	56.7–61.7	885/1495	54.1	50.5–57.7	389/719

**Notes:** Valid/Diagnoses, number of confirmed events/number of ICD diagnoses.

**Abbreviations:** PE, pulmonary embolism; DVT, deep vein thrombosis.

**Table 2** Proportion of misclassified events of PE and DVT in the Swedish National Patient Registry and Cause of Death Register

Time period	1985–2014		N/N	2009–2014		N/N
	% Misclassified	95% CI		% Misclassified	95% CI	
DVT or PE	9.2	7.9–10.5	180/1951	6.6	5.0–8.1	63/959
PE	1.1	0.4–1.7	10/944	0.9	0.0–1.7	4/458
DVT	16.4	14.2–18.7	174/1059	13.2	10.0–16.3	59/448

**Notes:** N/N, number of misclassified events/total number of valid events.

**Abbreviations:** PE, pulmonary embolism; DVT, deep vein thrombosis.

validating DVT was based entirely on radiology reports and did not allow possible DVT with clinical signs and symptoms but inconclusive radiology to be confirmed.

This study did not examine sensitivity, as no administrative sources or quality registries are available to identify false-negative events. However, we did attempt to estimate the number of misclassified PE and DVT events in our study cohort. We identified 180 (9%) participants with a confirmed PE or DVT event but without a diagnosis of PE or DVT. Of these, >90% were misclassified DVT events. Another Swedish study reported that PE and DVT diagnoses were coded incorrectly in 10% to 20% of registry cases.<sup>15</sup> A French study estimating the sensitivity of PE and DVT discharge diagnoses found a significantly higher sensitivity for PE (89%) than DVT (59%).<sup>16</sup> Our findings of a higher rate of misclassification for DVT than PE are in line with these results.

The major strength of the present study is the large population-based cohort covering approximately two-thirds of the population over 30 years of age in the area. Objective criteria were used to validate diagnoses. Only first time PE or DVT were included, and a valid diagnosis demanded a conclusive radiology or autopsy report. Probable or possible PE or DVT events based on reported signs and symptoms, initiated anticoagulant treatment, or physician diagnosis of PE or DVT in medical records were not considered valid events. These criteria for confirmed PE or DVT can lead to a lower PPV compared to studies in which possible or probable events were included.<sup>5</sup> On the other hand, an asymptomatic

PE or DVT accidentally found when radiology is performed as part of an ongoing investigation of another condition (eg, thorax/abdomen CT in the setting of a suspected malignancy) may fulfill the criteria for a valid event in the present study.

The relatively long study period, from 1985–2014, is a limitation, as there were substantial changes during this nearly 30-year period. Three different ICD coding systems (ICD-8, ICD-9, and ICD-10) have been used, with a transition period for each coding system. Several new diagnostic tools and treatment with low molecular heparins have been introduced, allowing treatment without admission to hospital wards. Taken together, these factors may affect the accuracy of the diagnosis. We did not require the date of the event to match the date of the diagnosis. A previous Danish study on diagnoses of VTE during pregnancy and puerperium showed that several of the confirmed events did not occur in relation to the pregnancy, reducing the overall PPV by 8% when restricting the time period.<sup>3</sup> However, we found in several cases that, when the initial ICD diagnosis was incorrect, a correct ICD diagnosis for the event was given at a subsequent follow-up visit or new hospitalization.

## Conclusion

Our results suggest that registry data on PE, especially from the most recent time period, is of acceptable quality and can be considered for use in registry-based studies. For DVT, we found poor quality data, in regards to both PPV and misclassification, and it should not be used without validation.

## Acknowledgments

Financial support was provided by a regional agreement between Umeå University and the Västerbotten County Council for cooperation in the fields of Medicine, Odontology, and Health, and by the Foundation for Medical Research in Skellefteå.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Norberg J, Bäckström S, Jansson JH, Johansson L. Estimating the prevalence of atrial fibrillation in a general population using validated electronic health data. *Clin Epidemiol*. 2013;5:475–481.
2. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health*. 2011;11:450.
3. Larsen TB, Johnsen SP, Möller CI, Larsen H, Sørensen HT. A review of medical records and discharge summary data found moderate to high predictive values of discharge diagnoses of venous thromboembolism during pregnancy and postpartum. *J Clin Epidemiol*. 2005;58(3):316–319.
4. White RH, Brickner LA, Scannell KA. ICD-9-CM codes poorly identified venous thromboembolism during pregnancy. *J Clin Epidemiol*. 2004;57(9):985–988.
5. Severinsen MT, Kristensen SR, Overvad K, Dethlefsen C, Tjønneland A, Johnsen SP. Venous thromboembolism discharge diagnoses in the Danish National Patient Registry should be used with caution. *J Clin Epidemiol*. 2010;63(2):223–228.
6. Tuckuviene R, Kristensen SR, Helgestad J, Christensen AL, Johnsen SP. Predictive value of pediatric thrombosis diagnoses in the Danish National Patient Registry. *Clin Epidemiol*. 2010;2:107–122.
7. Sundbøll J, Adelborg K, Munch T, et al. Positive predictive value of cardiovascular diagnoses in the Danish National Patient Registry: a validation study. *BMJ Open*. 2016;6(11):e012832.
8. Johansson M, Johansson L, Lind M. Incidence of venous thromboembolism in northern Sweden (VEINS): a population-based study. *Thromb J*. 2014;12(1):6.
9. Johansson M, Lind M, Jansson JH, Fährm E, Johansson L. Fasting plasma glucose, oral glucose tolerance test, and the risk of first-time venous thromboembolism. A report from the VEINS cohort study. *Thromb Res*. 2018;165:86–94.
10. Norberg M, Blomstedt Y, Lönnberg G, et al. Community participation and sustainability—evidence over 25 years in the Västerbotten Intervention Programme. *Glob Health Action*. 2012;5:1–9.
11. Norberg M, Wall S, Boman K, Weinehall L. The Västerbotten Intervention Programme: background, design and implications. *Glob Health Action*. 2010;3:10.3402/gha.v3i0.4643.
12. Socialstyrelsen. In English – the National Patient Register; 2018. Available from: <http://www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish>.
13. Brooke HL, Talbäck M, Hörnblad J, et al. The Swedish cause of death register. *Eur J Epidemiol*. 2017;32(9):765–773.
14. 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(12):e137–e143.
15. Schulman S, Lindmarker P. Incidence of cancer after prophylaxis with warfarin against recurrent venous thromboembolism. Duration of Anticoagulation Trial. *N Engl J Med*. 2000;342(26):1953–1958.
16. Casez P, Labarère J, Sevestre MA, et al. ICD-10 hospital discharge diagnosis codes were sensitive for identifying pulmonary embolism but not deep vein thrombosis. *J Clin Epidemiol*. 2010;63(7):790–797.



## Supplementary materials

### Appendix S1

ICD codes used for the searching the Swedish National Patient Registry (SNPR) and Cause of Death Register (CDR).

#### Pulmonary embolism:

ICD10: I26.0, I26.9, O88.2

ICD9: 415B, 673C

ICD8: 450.00, 450.01, 450.03, 450.09, 673.98

#### Deep Vein Thrombosis:

ICD10: I80.1, I80.2, I80.3, I80.8, I80.9, O22.3, O87.1

ICD9: 451B, 451C, 451W, 451X, 671D, 671E

ICD8: 451.98, 451.99, 671.01, 671.02

### Appendix S2

ICD codes for possible venous thromboembolism (VTE) events used when searching for misclassified events.

ICD 8: 321.00; 321.09; 426.02; 426.08; 426.09; 438.00; 438.99; 440.20; 450.02; 451.00; 452.99; 453.00; 453.99; 631.00; 631.10; 631.11; 631.20; 631.30; 634.50; 634.99; 642.00; 642.20; 643.00; 643.20; 644.00; 644.20; 671.00; 671.08; 671.09; 673.00; 673.10; 673.99; 674.99; 677.98; 677.99

ICD 9: 325X; 415A; 416A; 416B; 416W; 416X; 437G; 444C; 451A; 452X; 453A; 453B; 453C; 453D; 453W; 453X; 557A; 634G; 634 H; 634W; 635G; 635 H; 635W; 636G; 636 H; 636W; 637G; 637 H; 637W; 638G; 638 H; 638W; 639G; 639W; 639X; 671C; 671F; 671W; 671X; 673A; 673B; 673D; 673W; 674A

ICD 10: I27.8; I27.9; I67.6; I74.3; I80.0; I81.9; I82.0; I82.1; I82.2; I82.3; I82.8; I82.9; K55.0; O08.2; O08.7; O22.2; O22.5; O22.8; O22.9; O87.0; O87.2; O87.3; O87.8; O87.9; O88.0; O88.1; O88.3; O88.8

#### Clinical Epidemiology

#### Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

systematic reviews, risk and safety of medical interventions, epidemiology and biostatistical methods, and evaluation of guidelines, translational medicine, health policies and economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Dovepress