Thrombosis Research 126 (2010) 61–67

Contents lists available at ScienceDirect

Thrombosis Research

journal homepage: www.elsevier.com/locate/thromres

Regular Article

# Evaluation of the predictive value of ICD-9-CM coded administrative data for venous thromboembolism in the United States☆

Richard H. White [a,*], Martina Garcia [a], Banafsheh Sadeghi [b], Daniel J. Tancredi [b], Patricia Zrelak [b], Joanne Cuny [c], Pradeep Sama [c], Harriet Gammon [d], Stephen Schmaltz [d], Patrick S. Romano [a]

[a] Division of General Medicine, University of California, Davis, USA
[b] Center for Healthcare Policy and Research, University of California, Davis, USA
[c] University HealthSystem Consortium, Chicago, Il, USA
[d] The Joint Commission, Chicago, IL, USA

ARTICLE INFO

*Article history:*
Received 11 December 2009
Received in revised form 15 March 2010
Accepted 17 March 2010
Available online 28 April 2010

*Keywords:*
Venous thromboembolism
ICD-9-CM
Coding
Predictive value
Administrative data
Deep-vein thrombosis
Pulmonary embolism

ABSTRACT

*Objective:* To determine the positive predictive value of International Classification of Disease, 9th Revision, Clinical Modification (ICD-9-CM) discharge codes for acute deep vein thrombosis or pulmonary embolism.
*Materials and Methods:* Retrospective review of 3456 cases hospitalized between 2005 and 2007 that had a discharge code for venous thromboembolism, using 3 sample populations: a single academic hospital, 33 University HealthSystem Consortium hospitals, and 35 community hospitals in a national Joint Commission study. Analysis was stratified by position of the code in the principal versus a secondary position.
*Results:* Among 1096 cases that had a thromboembolism code in the principal position the positive predictive value for *any* acute venous thrombosis was 95% (95%CI:93-97), whereas among 2360 cases that had a thromboembolism code in a secondary position the predictive value was lower, 75% (95%CI:71-80). The corresponding positive predictive values for *lower extremity deep-vein thrombosis or pulmonary embolism* were 91% (95%CI:86-95) and 50% (95%CI:41-58), respectively. More highly defined codes had higher predictive value. Among codes in a secondary position that were false positive, 22% (95%CI:16-27) had chronic/prior venous thrombosis, 15% (95%CI:10-19) had an upper extremity thrombosis, 6% (95%CI:4-8) had a superficial vein thrombosis, and 7% (95%CI:4-13) had no mention of any thrombosis.
*Conclusions:* ICD-9-CM codes for venous thromboembolism had high predictive value when present in the principal position, and lower predictive value when in a secondary position. New thromboembolism codes that were added in 2009 that specify chronic thrombosis, upper extremity thrombosis and superficial venous thrombosis should reduce the frequency of false-positive thromboembolism codes.


## Introduction

Increasingly, epidemiologic studies are using “administrative” hospital discharge data to identify patients with important vascular outcome events, such as deep-vein thrombosis or pulmonary embolism, which together comprise venous thromboembolism (VTE). Administrative data are computerized records that are gathered for some administrative purpose, but contain information that can be used for other purposes as well. In the United States (US), the Uniform Claim and Billing Form 04 (UB-04) requires International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) coding. The Joint Commission (TJC- the regulatory agency that oversees licensing of hospitals), the Agency for Healthcare Research and Quality (AHRQ - the US agency charged with improving the quality, safety, efficiency, and effectiveness of health care), and the Centers for Medicare & Medicaid Services (CMS) have launched quality measurement, quality improvement, and pay-for-performance initiatives that require identification of patients with acute VTE using a set of specific ICD-9-CM codes in hospital discharge records [1]. Researchers interested in vascular outcomes who use administrative data rely on ICD-9-CM codes to define presence or absence of acute VTE [2–6].

Prior to October 2009, there were twenty ICD-9-CM codes for VTE in non-pregnant patients, including 3 codes for pulmonary embolism, 10 codes for ‘thrombophlebitis’ (451 series) and 7 codes for ‘other venous thrombosis or embolism’ (453 series) [7]. In addition, there

*Abbreviations:* VTE, Venous thromboembolism; ICD-9-CM, International Classification of Diseases, 9th Revision , Clinical Modification; POA, Present on admission; UHC, University HealthSystem Consortium; TJC, The Joint Commission; UC, University of California; UCDMC, University of California, Davis, Medical Center; UB-04, Uniform Claim and Billing Form 04; CI, Confidence Interval.

☆ Presented at the International Society of Thrombosis and Haemostasis Meeting, July 16, 2009, Boston, MA.

* Corresponding author. Division of General Medicine, UCDMC Medical Center, Suite 2400 PSSB, 4150 V Street, Sacramento, CA, 95817. Tel.: +1 916 734 7005; fax: +1 916 734 2732.
*E-mail address:* rhwhite@ucdavis.edu (R.H. White).

0049-3848/$ – see front matter 
doi:10.1016/j.thromres.2010.03.009

are several codes for pregnancy-related VTE (671 series) [8]. ICD-9-CM coding rules are evolving, as evidenced by the creation in October 2004 of 3 new codes that specify deep-vein thrombosis in the leg, thigh, or calf, and the creation in October 2009 of 21 new codes that specify additional locations and the acuity of VTE [9]. In addition, an 'indicator' to specify if a condition was present-on-admission (POA) was introduced in New York and California in the 1990s [10], and since October 2007 is required by CMS. This indicator should theoretically aid in the identification of hospital-acquired acute VTE events (i.e., not present-on-admission) [3,4].

Only a few studies have retrospectively analyzed the positive predictive value of VTE codes. An early study showed that the most commonly used codes, when present in the principal position (specific for the condition that occasioned hospital admission), had very high positive predictive value for acute VTE (i.e., approximately 95%) [11]. However, other studies have reported that VTE codes in any position (principal or secondary) have much lower predictive value for acute VTE, in the range of 70-75%. In a multisite prospective cohort study, Cushman et al reported that the predictive value of any VTE code was 68% (95%CI: 63-74) [12]. More recently, Arnason et al [13] at a single hospital, and Heckbert et al [14], from the Women's Health Initiative, reported positive predictive values of 74% (95%CI: 64-82) and 78% (95%CI: 70-85), respectively. In another study using data from CMS, the predictive value of coding for deep vein thrombosis was 72% [15]. Finally, Zhan and coworkers analyzed the predictive value of ICD-9-CM coding for post-operative VTE using Medicare data and found a predictive value of only 29% for VTE [16]. These published studies have generally analyzed the predictive value of a group of VTE codes for *any* acute VTE event, not strictly lower extremity deep-vein thrombosis OR pulmonary embolism.

The aim of this study was to analyze a large sample of records from a wide array of hospitals throughout the United States to determine the predictive value of individual ICD-9-CM codes located in the principal position versus a secondary position for either *any* acute VTE or *acute lower extremity deep-vein thrombosis or pulmonary embolism*. Our objectives were to provide the ICD-9-CM Coordination and Maintenance Committee with the information necessary to enable them to restructure VTE codes in a more logical can comprehensive fashion, and to inform researchers and quality improvement professionals who use VTE codes for surveillance purposes about the predictive value of individual ICD-9-CM VTE codes [17].

## Methods

The present study represents the compilation of three independent chart abstraction efforts by the University of California, Davis Medical Center (UCDMC), the University HealthSystem Consortium (UHC), and The Joint Commission (TJC). Because the research goals and abstraction methods were similar across projects, the first and last authors, who were involved in all three projects, decided to pool the data in order to enhance reliability. The UHC project also aimed to estimate the false negative rate, or sensitivity, of VTE coding, to ensure that ICD-9-CM based surveillance definitions are not compromised by underreporting of VTE events.

### *Case identification*

#### *UC Davis*

The records of all patients hospitalized at UCDMC between July 1, 2005 and June 30, 2006, who were at least 18 years of age at admission and who had at least one of the pre-specified ICD-9-CM VTE codes in either the principal (diagnosis occasioning admission to the hospital) or a secondary position (all other medical diagnoses) were abstracted.

#### *The Joint Commission*

TJC identified cases from 35 community and regional hospitals across the country that volunteered to pilot test eight VTE performance measures as part of the National Voluntary Consensus Standards for the Prevention and Care of VTE Project [18]. Five of the measures specifically required identifying patients with acute VTE. Eligible cases had one of the pre-specified ICD-9-CM VTE codes in either the principal or secondary position, were discharged from the hospital between October, 2006 and March, 2007, and were over the age of 18 years. Hospitals were instructed to obtain a random sample of approximately 10 charts per measure.

#### *University HealthSystem Consortium*

Thirty-three major teaching hospitals in 21 states volunteered to participate in a UHC VTE performance improvement project that randomly selected records of approximately 1000 medical and 1000 surgical patients hospitalized between January 1, 2006 and March 31, 2007 who were coded with (50%) or without (50%) an ICD-9-CM code for acute VTE *during* the hospitalization, specifically excluding all cases that were admitted specifically for treatment of acute VTE (i.e., VTE was not the principal diagnosis). In accord with the definition of AHRQ's Patient Safety Indicator #12 (postoperative pulmonary embolism or deep vein thrombosis), cases were ineligible if they were pregnant, if age was less than 18 years, or if an inferior vena cava filter was placed before or on the same day as major surgery (or was the only procedure). None of the UHC hospitals participated in the TJC project. UCDMC did participated in the UHC study, but based on the calendar time that samples were collected, no specific cases were included in both studies.

At each hospital in the UHC study, 100 eligible cases were randomly identified: 50 surgery cases that underwent a valid operating room procedure [19] ,and 50 medical cases that had a length of stay of two or more days with a severity of illness score higher than 'minor'[20]. In both the medical and surgical cohorts, 25 of the sampled cases had a secondary diagnosis code for VTE and 25 did not. Using these lists, the first 15 of these randomly selected cases were abstracted.

### *Venous thromboembolism case definition*

In each of the 3 sub-studies, potential VTE cases were identified if the discharge record had one or more of the 13 ICD-9-CM codes (Table 2) that can be used to define deep venous thrombosis or pulmonary embolism [9]. Any case that had more than one of these VTE codes was assigned the code with the highest specificity for acute VTE, using a hierarchy that was defined *a priori*, from the highest, pulmonary embolism, to 'venous thrombosis of deep veins of the lower extremity,' 'venous thrombosis of an other specified vein,' 'phlebitis or thrombophlebitis' involving a deep vein in the lower extremity,' and finally to the remaining 'other' venous thrombosis codes. Upper extremity phlebitis codes were not eligible (e.g. 451.0, 451.82-451.89).

### *Abstraction process*

At UCDMC, two abstractors (RHW, MG) reviewed all records and entered data into a preconfigured Excel spreadsheet. UHC abstractors were trained by study staff via teleconferences and/or web-based sessions that focused on how to complete the data collection tools and how to apply the accompanying detailed guidelines. Data were entered directly into a web-based application. TJC also developed an electronic abstraction tool to facilitate data collection efforts. Each of the hospitals working with UHC and TJC used registered nurses with varying levels of experience and backgrounds to abstract the sampled records following standardized instructions and guidelines.

### *Outcomes identified by abstraction*

Within each sub-study the abstraction tools were identical at each site. In all three sub-studies abstractors gathered the following information (unless otherwise noted): 1) presence or absence of an acute deep-vein thrombosis or pulmonary embolism diagnosed using an objectively confirmed test; 2) the location of the acute VTE (name or site of veins involved), which was *not* required in the TJC substudy; 3) use or non-use of full-dose intravenous heparin, low molecular weight heparin or fondaparinux to treat the acute VTE; 4) use of warfarin and start date; and 5) presence or absence of a previous history of VTE prior to the index hospitalization.

Among the cases identified at UCDMC and at UHC hospitals, if there was more than one site of acute thrombosis, the assigned site for each case was the highest based on the following descending hierarchy: acute pulmonary embolism, deep-vein thrombosis in the lower extremity or vena cava, upper extremity-thoracic venous thrombosis, acute thrombosis in a superficial vein in the leg or arm, and finally, location not specified.

*Abstraction outcomes*

*Objective confirmation of VTE*

For all three sub-studies, objectively confirmed VTE required a report stating that acute VTE was found in the pulmonary arteries, the deep-veins of the lower extremity, the inferior or superior vena cava, a thoracic vein, or an upper extremity/neck vein, using compression/duplex ultrasound, chest computerized tomographic angiogram, pulmonary arteriogram, pulmonary scan (high probability) or venography. Superficial venous thrombosis could be documented by compression ultrasound or by clinical diagnosis alone.

*Reference Standards Used To Define Acute VTE*

At UCDMC and the UHC hospitals, each case was classified as true-positive or false-positive for acute VTE based on two reference standards: *Criterion 1*: acute VTE in *any* location that was symptomatic and objectively documented (upper or lower extremity, superficial or deep), and *Criterion 2*: *acute lower extremity deep-vein thrombosis or pulmonary embolism* that was symptomatic and objectively documented. Thus, if a case had an acute superficial venous thrombosis in a leg, or an upper extremity deep-vein thrombosis, and was coded as 453.8 ('venous embolism and thrombosis of other specified vein'), then it was classified as true positive using *Criterion 1* but false positive using *Criterion 2*. The TJC sub-study only used *Criterion 1* because the abstraction tool used did not require documentation of the venous segment(s) affected. Overall, 5 cases did not have objectively documented pulmonary embolism because a diagnostic study could not be performed, but were classified as positive for pulmonary embolism because they were treated for acute pulmonary embolism using therapeutic doses of heparin followed by warfarin.

*Data management and analysis*

We estimated that a sample size of N = 93 would be sufficient to estimate the predictive value of any individual ICD-9-CM code, or combination of codes, with a confidence interval width of 20% (i.e., 40%-60%). Given the fact that 6 of the VTE codes are used commonly and 14 are used less commonly, we estimated we would need to collect a cohort of over 3000 cases to get sufficiently narrow confidence limits to make any comparisons between commonly used codes in a principal and secondary position.

Predictive values were estimated using the survey data analysis procedure SURVEYFREQ in SAS version 9.1 to adjust estimates for features of the sampling design, including the clustering of observations within hospitals, unequal sampling probabilities and, for the UHC sample, the stratification of cases as medical or surgical. For small sample sizes less than 9, exact 95% confidence limits were calculated.

Because UCDMC and TJC identified only cases that had a VTE code, only positive predictive values were calculated. Because UHC assembled separate probability samples of at least 15 VTE code-positive and 15 VTE code-negative cases from each hospital for both the medical and surgical groups, these data were used to estimate sensitivity, specificity, and positive and negative predictive values. UHC cases that did not have a VTE code were sampled with much lower probability than cases with a VTE code, resulting in what is termed a 'verification-biased sample' [21]. To correct for this, UHC estimates were weighted by the inverse of the sampling fraction, which differed across patients and across hospitals. Sensitivity and specificity were estimated using ratio estimators, to properly account for random error in the identification of true cases. If a point estimate was 100%, which vitiated the use of the asymptotic variance estimators used in procedure SURVEYFREQ, exact binomial 95% confidence intervals were reported. Interval estimates for sensitivity were estimated by first using Bayes' theorem to express sensitivity as a function of indicator prevalence and then substituting the endpoints of the 95% CI estimates for positive predictive value and negative predictive value into the expression to describe a range of plausible values.

This study was approved by the Human Subjects Committee at UCDMC. Research use of the de-identified UHC and TJC data was also approved by the UCDMC Human Subjects Committee and by the respective organizations.

## Results

The demographic characteristics of the samples from UCDMC, TJC and the two UHC subgroups (medical and surgical) that had a VTE code are shown in Table 1**.** The Joint Commission (N = 2052) sample had a median age that was 10 years older than both the UCDMC (N = 413) and UHC samples (N = 991), and the percentage of women in the TJC sample was also higher.

*ICD-9-CM VTE Codes in the Principal Position*

Table 2 shows the positive predictive value of each of the ICD-9-CM codes in the principal position, estimated from the two sub-studies that included such cases. The overall positive predictive value for *any* acute VTE (*Criterion* 1) was 95% (95%CI:93-97), with values of 98% (95%CI: 94-99) for the UDCMC sample and 95% (95%CI: 93-97) for the TJC sample. The positive predictive value for objectively confirmed *lower extremity deep-vein thrombosis or pulmonary embolism* (*Criterion* 2), which was estimated only at UCDMC, was 91% (95%CI: 86-95). For the combined samples, more highly defined codes for lower extremity deep-vein thrombosis, including 453.40 - 453.42 and 453.8, had predictive values for any acute *VTE* that were greater than 90%, whereas the 451 series codes, which specify thrombophlebitis, and 453.9, which is 'thrombosis of unspecified site', had somewhat lower predictive values.

**Table 1**
Demographic characteristics of cases from samples.

| | UCDMC, N | The Joint Commission, N | University HealthSystem Consortium Medical, N | University HealthSystem Consortium Surgical, N |
|---|---|---|---|---|
| Number | 413 | 2052 | 501 | 490 |
| Mean age ± SD | 53.8 ± 18.5 | 63.8 ± 18 | 56 (54-58) | 56.1 (54.2-58) |
| Median | 56 | 66 | 57 | 56 |
| Gender age | | | | |
| Male, N (%) | 217 (53%) | 955 (47%) | 260 (52%) | 268 (55%) |
| Female, N (%) | 196 (47%) | 1097 (53%) | 241 (48%) | 222 (45%) |
| Race | | | | |
| Caucasian | 213 (52%) | N/A | 298 (59%) | 313 (64%) |
| African-American | 55 (13%) | N/A | 143 (29%) | 119 (24%) |
| Hispanic | 45 (11%) | N/A | 33 (6.6%) | 36 (7.3%) |
| Asian/Pacific Islander | 19 (4.6%) | N/A | 11 (2.2%) | 9 (1.8%) |
| Other/Mixed | 71 (17%) | N/A | 16 (3.2%) | 17 (3.4%) |

**Table 2**
Predictive value of ICD-9-CM codes in the principal position for two different reference standard definitions of venous thromboembolism.

| ICD-9-CM Code | Reference standard measure: any acute venous thromboembolism* | | | Reference standard measure: acute pulmonary embolism or lower extremity deep vein thrombosis# | | |
|---|---|---|---|---|---|---|
| | Cases | Events | PPV ,%, (95%CI) | Cases | Events | PPV ,% (95%CI) |
| All VTE Codes, N (%) | 1096 | 1043 | 95% (93-97) | 164 | 150 | 91% (86-95) |
| *415 Codes: pulmonary embolism* | | | | | | |
| 415.11 PE and infarction, 'iatrogenic' | 52 | 50 | 96% (86-99) | 9 | 8 | 89% (68-100) |
| 415.19 PE and infarction, other | 632 | 608 | 96% (94-98) | 94 | 93 | 99% (97-100) |
| All pulmonary embolism codes | 684 | 658 | 96% (94-98) | 103 | 101 | 98% (93-99) |
| *451 Codes: thrombophlebitis* | | | | | | |
| 451.11 Phlebitis, femoral vein | 4 | 4 | 100% (4-100) | - | - | - |
| 451.19 Phlebitis, other deep vein | 7 | 6 | 86% (57-100) | - | - | - |
| 451.2 Phlebitis, leg, unspecified | 2 | 1 | 50% (0-100) | - | - | - |
| 451.9 Phlebitis, unspecified site | 1 | 1 | 100% (3-100) | - | - | - |
| All thrombophlebitis codes | 14 | 12 | 86% (66-100) | - | - | - |
| *453 Codes: venous thrombosis* | | | | | | |
| 453.1 Thrombophlebitis migrans | 1 | 1 | 100% (3-100) | 61 | 1 | 100% (3-100) |
| 453.2 Vena cava | 12 | 8 | 67% (39-90) | 1 | 0 | 0% (0-97) |
| 453.40 Lower extremity DVT, not specified | 57 | 53 | 93% (86-100) | 1 | 1 | 100% (3-100) |
| 453.41 Lower extremity DVT, proximal | 179 | 170 | 95% (92-99) | 19 | 17 | 89% (66-99) |
| 453.42 Lower extremity DVT, distal | 78 | 74 | 95% (86-99) | 9 | 7 | 78% (38-99) |
| 453.8 Thrombosis , other specified vein | 70 | 66 | 94% (85-99) | 30 | 23 | 77% (56-90) |
| 453.9 Thrombosis of unspecified site | 1 | 1 | 100% (3-100) | - | - | - |
| Prox or Distal DVT: 453.41 or 453.42 | 257 | 245 | 95% (90-98) | 28 | 24 | 86% (67-96) |
| Other DVT codes: 453.1, 453.2, 453.9 | 14 | 9 | 64% (33-88) | - | - | - |
| All venous thrombosis codes | 398 | 373 | 94% (90-97) | 61 | 49 | 80% (68-89) |

ICD-9-CM = International Classification of Disease VTE = venous thromboembolism; DVT = lower extremity deep-vein thrombosis; PE = pulmonary embolism, PPV = positive predictive value.
* combined UCDMC and The Joint Commission sample.
# only the UCDMC sample.
CI = 95% Confidence limits rounded to nearest integer adjusted for clustering in The Joint Commission sample.

### *ICD-9-CM VTE Codes in a Secondary Position*

The findings are summarized in Table 3. Coding groups 415.11 or 415.19 ('pulmonary embolism'), 453.40-453.42 ('deep-vein thrombosis in the leg') and 453.8 ('venous thrombosis of other specified vein') made up 96% of all secondary codes. Thrombophlebitis codes (451 series codes) made up only 1.4% of all the cases. The overall predictive value of a VTE code in a secondary position for *any* acute VTE was 75% (95%CI: 71-80). The predictive value of individual ICD-9-CM codes for *any* acute VTE varied moderately, with more highly defined codes having higher predictive values. For example, cases coded as 415.11, 'iatrogenic' (hospital associated) pulmonary embolism had a positive predictive value of 93% (95%CI: 88-98) compared to 79% (95%CI: 74-84) for 415.19, 'other' pulmonary embolism. Similarly 453.41 and 453.42, 'proximal and distal lower extremity deep-vein thrombosis', respectively, had a combined predictive value of 85% (95%CI: 82-88) for *any* acute VTE, which was higher than the predictive value of 61% (95%CI: 51-71) for 453.40 ('lower extremity deep-vein thrombosis not otherwise specified'). The commonly used but less specific code 453.8 ('venous thrombosis of other specified vein') had a predictive value of 70% (95%CI: 62-78) for any acute VTE. Non-specific codes, including the 451 thrombophlebitis codes and 453.9, ('thrombosis of unspecified site') had predictive values for *any* acute VTE that were lower than 70%.

### *Predictive Value of ICD-9-CM VTE Codes in a Secondary Position for Lower Extremity Deep Vein Thrombosis or Pulmonary Embolism*

Reference standard *Criterion* 2, which required documented *deep-vein thrombosis in a leg or pulmonary embolism*, could be analyzed using 1240 cases from UCDMC and UHC. Overall, the predictive value of secondary VTE codes was 50% (95%CI: 41-58). For the pulmonary embolism codes 415.11 and 415.19, the predictive value was 75% (95%CI: 68-83), whereas for 451 and 453 series codes combined, the predictive value was only 39% (95%CI: 29-49). Again, the predictive value of more highly defined VTE codes was higher for lower extremity venous thrombosis than poorly specified codes: 73% (95%CI:65-81) for 453.41 ('proximal deep-vein thrombosis'), 65% (95%CI:52-78) for 453.42 ('distal deep-vein thrombosis'), 38% (95%CI: 27-49) for 453.40 ('lower extremity deep-vein thrombosis, not otherwise specified'), and 26% (95%CI:6-46) for 453.8 ('thrombosis of other specified vein').

Among the 1240 cases with a secondary code for VTE that had the site of thrombosis documented, 876 (71%; 95%CI: 63-78) had an acute VTE event, 269 (22%, 95%CI: 16-27) had only a prior or chronic VTE, and 95 (7%, 95%CI:4-13) had no mention of either acute VTE or a recent or chronic VTE. In this same group, 616 (49.8%, 95%CI: 41-58) specifically had an *acute lower extremity deep-vein thrombosis or pulmonary embolism*, 182 (15%, 95%CI:10-19) had an acute upper extremity deep-vein thrombosis, 74 (6%, 95%CI: 4-8) had an acute thrombosis in a superficial vein, and the location was not recorded in 4 (0.5%, CI: 0.1-1.2) cases.

Table 4 shows the effect of selecting progressively less highly-specified secondary position ICD-9-CM VTE codes on the corresponding positive predictive value. The tradeoff between positive predictive value and sensitivity (approximated by the percent of all cases identified) can be readily seen. For example, including Group 4 (the single code 453.8, 'thrombosis of other specified vein') in a surveillance definition would increase the sensitivity for identifying lower extremity deep-vein thrombosis or pulmonary embolism substantially, from 58.5% to 95.5%, but use of this code would decrease the positive predictive value, from 67.2% to 51.3%.

### *Present-on-admission (POA) Indicator*

Use of the POA indicator for VTE codes in a secondary position could only be evaluated in the UCDMC sub-study, because this information was not available in the data from UHC or TJC. Among the 249 cases with a secondary VTE code, 132 (51%) had an indicator of *not* being present on admission (hospital acquired, or POA = N), 67 (21%)

**Table 3**
Predictive value of ICD-9-CM codes in a secondary position for two different reference standard definitions of venous thromboembolism.

| ICD-9-CM Code | Reference standard measure: any acute venous thromboembolism* | | | Reference standard measure: acute pulmonary embolism or lower extremity deep vein thrombosis# | | |
|---|---|---|---|---|---|---|
| | Cases | Events | PPV ,% (95%CI) | Cases | Events | PPV ,% (95%CI) |
| All VTE Codes, N (%) | 2360 | 1779 | 75% (71-80) | 1240 | 617 | 50% (41-58) |
| *415 Codes: pulmonary embolism* | | | | | | |
| 415.11 PE and infarction, 'iatrogenic' | 126 | 117 | 93% (88-98) | 57 | 52 | 91% (83-99) |
| 415.19 PE and infarction, other | 644 | 509 | 79% (74-84) | 311 | 225 | 72% (63-81) |
| Any PE Code: 415.11 or 415.19 | 770 | 626 | 81% (77-860 | 368 | 277 | 75% (68-83) |
| *451 Codes: thrombophlebitis* | | | | | | |
| 451.11 Phlebitis, femoral vein | 3 | 1 | 33% (8-90) | 2 | 0 | 0% (0-84) |
| 451.19 Phlebitis, other deep vein | 9 | 6 | 67% (29-92) | 4 | 2 | 50% (6-93) |
| 451.2 Phlebitis, leg, unspecified | 5 | 3 | 60% (15-94) | 2 | 0 | 0% (0-84) |
| 451.9 Phlebitis, unspecified site | 17 | 7 | 41% (17-67) | 14 | 2 | 14% (0-31) |
| Any thrombophlebitis code | 34 | 17 | 50% (31-69) | 22 | 4 | 18%(2-34) |
| *453 Codes: venous thrombosis* | | | | | | |
| 453.1 Thrombophlebitis migrans | 1 | 1 | 100% (3-100) | - | - | - |
| 453.2 Vena cava | 29 | 18 | 62% (42-83) | 7 | 3 | 43% (17-69) |
| 453.40 Lower extremity DVT, NOS | 236 | 144 | 61% (51-71) | 126 | 48 | 38% (27-49) |
| 453.41 Lower extremity DVT, proximal | 355 | 306 | 86% (82-91) | 158 | 115 | 73% (65-81) |
| 453.42 Lower extremity DVT, distal | 184 | 152 | 83% (74-91) | 74 | 48 | 65% (52-78) |
| 453.8 Thrombosis,,other specified vein | 709 | 495 | 70% (62-78) | 458 | 119 | 26% (6-49) |
| 453.9 Thrombosis of unspecified vein | 42 | 20 | 48% (27-68) | 27 | 3 | 11% (1-30) |
| Prox or Distal DVT: 453.41 or 453.42 | 539 | 458 | 85% (80-89) | 232 | 163 | 70% (62-78) |
| Other DVT Codes: 453.1, 453.2, 453.9 | 72 | 39 | 54% (31-69) | 22 | 4 | 18% (2-34) |
| All venous thrombosis codes | 1556 | 1136 | 74% (69-76) | 850 | 336 | 40% (35-44) |

ICD-9-CM = International Classification of Disease, 9th Revision; VTE = venous thromboembolism; DVT = deep-vein thrombosis; PE = pulmonary embolism, PPV = positive predictive value.
* All samples combined, adjusted for clustering.
# UCDMC and UHC surgery and medical samples were adjusted for clustering.
95% confidence limits rounded to nearest integer.

had an indicator of being present on admission (POA = Y) and 50 (38%) had an indicator that was missing or unknown). Among the 132 POA = N cases, 128 (97%, 95%CI: 92-99) had an objectively documented acute VTE. However, 29 of these (22%, 95%CI: 15-30) had an acute upper extremity venous thrombosis. Among the 67 cases that were indexed as having a VTE event that was present-on-admission (POA = Y), only 43 (64%, 95%CI: 52-76) had an objectively documented acute VTE. Among the 50 cases that had a "missing" or "unknown" indicator, 38 (76%, 95% CI: 62-87) had an acute VTE.

*Sensitivity and Specificity*

The fully weighted sensitivity and specificity of a secondary diagnosis code of acute lower extremity deep-vein thrombosis or pulmonary embolism was calculated using the UHC medical and surgical samples. In the UHC medical sample, 2 of 501 patients without a diagnosis code for VTE had documented acute deep-vein thrombosis, leading to a weighted sensitivity of 77% (95%CI: 4-100) and a weighted specificity of 99.0% (95%CI: 98.7-99.2). In the UHC surgical sample, the weighted sensitivity was 100% (95%CI: 48-100), given there were zero false negative cases, and the weighted specificity was 99.0% (95%CI: 98.8-99.2).

## Discussion

All hospitals in the United States collect coded patient discharge data, which are used for a multitude of purposes [22] and are submitted to state health organizations and to various vendors, including The Joint Commission. The American Health Information Management Association

**Table 4**
Effect of Selecting Specific Secondary Position ICD-9-CM Venous Thromboembolism Codes on the Proportion of Cases Identified, and the Positive Predictive Value.

| ICD-9-CM code groups | VTE codes in a secondary position | | | |
|---|---|---|---|---|
| | Percent cases | PPV for: any VTE event | Percent cases | PPV for pulmonary embolism or lower extremity deep vein thrombosis |
| Group 1 only | 32.6% | 81.2% | 29.7% | 75.3% |
| Group 1 & 2 | 52.5% | 82.8% | 48.4% | 73.3% |
| Groups 1-3 | 65.4% | 79.5% | 58.5% | 67.2% |
| Groups 1-4 | 95.5% | 76.4% | 95.5% | 51.3% |
| Groups 1-5 | 98.6% | 75.7% | 98.2% | 50.3% |
| All Groups: 1-6 | 100% | 75.4% | 100% | 49.7% |

ICD-9-CM; International Classification of Disease, 9th Revision, PPV = Positive predictive value.
Group 1: Pulmonary Embolism Codes: 415.11, 415.19.
Group 2: Specific lower extremity DVT codes: 453.41, 453.42.
Group 3 : Nonspecific lower extremity DVT code: 453.40.
Group 4: Thrombosis of other specified vein: 453.8.
Group 5: Other venous thrombosis codes: 453.1, 453.2, 453.9.
Group 6. Thrombophlebitis codes: 451.11, 451.19, 451.2, 451.9.

recognizes that "the collection of accurate and complete coded data is critical to healthcare delivery, research, public reporting, reimbursement, and policy-making. The integrity of coded data and the ability to turn it into functional information requires all users to consistently apply the same official coding rules, conventions, guidelines, and definitions (the basis of coding standards)." For researchers who study the epidemiology of VTE, use of administrative data is the primary means of identifying patients with this disorder [23], and hospitals use these data to assess quality of care [24].

Unfortunately, only a few studies have been conducted that have determined the accuracy of ICD-9-CM coding for acute VTE [3,11–15], and few of these studies have determined the predictive value of specific VTE codes in either the principal or a secondary position [8], which is essential to guide researchers in designing surveillance definitions. The principal diagnosis is defined in the Uniform Hospital Discharge Data Set (UHDDS) as "that condition established after study to be chiefly responsible for occasioning the admission of the patient to the hospital for care." [9]. In contrast secondary diagnoses are "all conditions that coexist at the time of admission, that develop subsequently, or that affect the treatment received and/or the length of stay." Diagnoses that relate to an earlier episode which have no bearing on the current hospital stay are not included. Coders have little or no role to play in the decision regarding the selection of the principal diagnosis but much greater discretion in the selection of secondary codes.

An important finding of the present study was the confirmation that ICD-9-CM codes for VTE that are in the principal position have a positive predictive value for *any* acute VTE event of over 95% [11]. In fact, based on abstraction of 164 records at UCDMC, 91% of the cases with a principal diagnosis VTE had documented acute lower extremity deep-vein thrombosis or pulmonary embolism. Equally important, the positive predictive value of a VTE codes in a secondary position for *any* acute VTE was much lower, approximately 75%, which is consistent with prior reports [12,14,15]. This finding was not surprising because a code in a secondary position may reflect either an acute medical problem or a pre-existing condition. Indeed, the major reason that VTE codes in a secondary position had lower predictive value was that a sizeable proportion of these patients had a prior or chronic VTE. In the combined UCDMC and UHC samples, 71% of the patients with a secondary VTE code had a validated acute VTE, but 22% had a prior or chronic VTE, and the remaining 7% had no mention of either a recent acute or a prior VTE.

Not surprisingly, VTE codes that specified a deep-vein thrombosis in the thigh or lower leg (453.41, 453.42), or a pulmonary embolism (415.1x), had the highest positive predictive value, whereas less highly defined codes, such as 453.40 ('lower extremity deep-vein thrombosis, not otherwise specified') and 453.8 ('other specified deep-vein thrombosis'), had lower positive predictive values. Nevertheless, even among the patients who had a highly specified VTE code, a substantial percentage had a prior or chronic VTE, not an acute VTE. We were not able determine why coders selected one particular VTE code instead of another, but the principal factor that constrains and directs coders is physician documentation [7]. If a physician only documents that a lower extremity venous thrombosis is present, the coder cannot be certain of the location (thigh versus calf, or superficial versus deep), particularly if there is no ultrasound report for review or the physician does not reply to a query.

Using the presence of a lower extremity deep-vein thrombosis or pulmonary embolism as the criterion standard for VTE, the predictive value of a VTE code in a secondary position was much lower, approximately 50%. The explanation for this finding is that in addition to including many patients with chronic VTE, 15% of the patients who had a VTE code in a secondary position had an acute upper extremity deep-vein thrombosis and 6% had an acute superficial venous thrombosis. Interestingly, this observed prevalence of upper extremity VTE was similar to the 15% prevalence reported among hospitalized patients in the Worcester, Massachusetts area [23].

False negative VTE cases were rare, and the estimated sensitivity of ICD-9-CM coding for lower extremity deep-vein thrombosis or pulmonary embolism was 100% in the UHC surgery sample and 77% in the UHC medical sample. However, given the relatively small number of VTE negative cases that were sampled and abstracted, the confidence limits for the estimated sensitivity were very wide. Nevertheless, it is clear that some non-surgical patients with acute VTE are not coded as having VTE.

Beginning October 1, 2007 the Center for Medicare & Medicaid Services required that all diagnoses include an indicator code to identify if the condition was present on admission (P), not present on admission (N), missing (M) or unknown (U) [25]. Conditions that develop in the hospital have a POA = N indicator. Although we did not specifically validate if VTE events with a POA = N indicator were or were not present on admission, simply the presence of the POA = N indicator dramatically improved the predictive value of these codes for acute VTE, very likely because the patients who had a prior or chronic VTE were assigned a POA = Y indicator. Overall, 97% of patients whose VTE code had an indicator showing that the condition was not present on admission had an objectively confirmed acute VTE event compared to a positive predictive value of 69% for events with codes with a POA = Y, N or M. Nevertheless, 22% of the patients whose VTE code had a POA = N indicator had an upper extremity thrombosis, indicating that the POA indicator improved the identification of acute VTE events but not the location. Similar to our findings, a prior study also reported that hospital acquired VTE events had a very high positive predictive value [4], but that this came at the expense of a higher false negative rate.

The major limitation of this study was that three separate groups of cases were pooled, and each used somewhat different inclusion criteria and slightly different definitions of outcomes. The outcome for TJC cases was *any* objectively confirmed venous thrombotic event at any site, whereas for UCDMC and UHC the outcome was the exact site of the objectively confirmed VTE event. Although a total of 68 hospitals participated in the UHC and TJC studies, the findings may not generalize to all hospitals because of differences in local coding practices and the underlying prevalence of VTE. Although the positive predictive value of VTE codes is theoretically sensitive to the prevalence of disease, as well as the sensitivity and specificity of reporting, epidemiologic studies have found relatively little geographic variation in the age-adjusted prevalence of VTE in developed countries [23,26–28]. We analyzed the ICD-9-CM codes for VTE that were available between October 2004 and October 2009. The findings are therefore most applicable to data gathered during this time period.

Four major conclusions can be made. First, the positive predictive value of ICD-9-CM VTE codes for *any* acute VTE was very high (approximately 95%) if they were in the principal position, and lower (approximately 75%) if the codes were in a secondary position. Second, more rigorously defined codes (e.g. deep-vein thrombosis in proximal leg) had higher predictive value than less highly specified codes (e.g., 'venous thrombosis, unspecified site'). Third, the low predictive value of VTE codes located in a secondary position was largely due to use of these codes in patients who had a chronic deep-vein thrombosis, an upper extremity deep-vein thrombosis, or a superficial venous thrombosis. Fourth, there was a tradeoff between sensitivity and positive predictive value associated with the including poorly specified venous thrombosis codes (e.g., 453.1, 453.9) and thrombophlebitis (all 451 series codes) codes to identify cases with acute VTE.

In October 2009, partially in response to the presentation and consideration of a preliminary version of these research findings, the ICD9-CM Coordination and Maintenance Committee added new codes for chronic VTE, upper extremity deep-vein thrombosis and superficial thrombosis. Availability of theses codes should greatly improve the positive predictive value of VTE coding and facilitate future use of hospital administrative data for VTE surveillance and for the epidemiologic studies of VTE risk factors and outcomes.

## Conflict of interest statement

The authors have no actual or potential conflict of interest specifically they have no financial, personal or other relationships with other people or organizations within three years of beginning the work submitted that could inappropriately influence their work.

## Acknowledgment

This study was supported, in part, by a contract (290-04-0020, Support for Quality Indicators) with the Agency for Healthcare Research and Quality and, in part, by the Hibbard E Williams Endowment for General Medicine at UCDMC.

## References

[1] The Joint Commission. Venous Thromboembolism (VTE) Core Measure Set - Last Updated 4/2009; 2009 [updated 2009 April, 2009; cited June 17, 2009]; Available from: http://www.jointcommission.org/PerformanceMeasurement/Performance-Measurement/VTE.htm.

[2] Henderson K, Recktenwald AJ, Reichley RM, Bailey TC, Waterman BM, Diekemper RL, et al. Clinical Validation of the AHRQ Postoperative Venous Thromboembolism Patient Safety Indicator. Jt Comm J Qual Patient Saf 2009;35:370–6.

[3] Houchens RL, Elixhauser A, Romano PS. How often are potential patient safety events present on admission? Jt Comm J Qual Patient Saf Mar 2008;34:154–63.

[4] Leibson CL, Needleman J, Buerhaus P, Heit JA, Melton III LJ, Naessens JM, et al. Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology Project VTE cohort. Med Care Feb 2008;46:127–32.

[5] Romano PS, Mull HJ, Rivard PE, Zhao S, Henderson WG, Loveland S, et al. Validity of selected AHRQ Patient Safety Indicators based on VA National Surgical Quality Improvement Program Data. Health Serv Res 2009;44(1):182–204.

[6] White RH, Sadeghi B, Tancredi DJ, Zrelak P, Cuny J, Sama P, et al. How Valid is the ICD-9-CM Based AHRQ Patient Safety Indicator for Postoperative Venous Thromboembolism? Med Care Sep 25 2009.

[7] Center for Disease Control. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM), Sixth Edition. Hyattsville, MD 2008 [updated September 04, 2008; cited 2010 Jan 31]; Available from: http://www.cdc.gov/nchs/products/elec_prods/subject/icd96ed.htm.

[8] White R, Brickner L, Scannell K. ICD-9-CM codes poorly indentified venous thromboembolism during pregnancy. J Clin Epidemiol 2004;57:985–8.

[9] Center for Disease Control. ICD-9-CM Official Guidelines for Coding and Reporting. Journal [serial on the Internet]. 2009 Date [cited 2009 October 1]: Available from: http://www.cdc.gov/nchs/datawh/ftpserv/ftpicd9/ftpicd9.htm.

[10] Naessens JM, Campbell CR, Berg B, Williams AR, Culbertson R. Impact of diagnosis-timing indicators on measures of safety, comorbidity, and case mix groupings from administrative data sources. Med Care Aug 2007;45:781–8.

[11] White RH, Romano PS, Zhou H, Rodrigo J, Bargar W. Incidence and time course of thromboembolic outcomes following total hip or knee arthroplasty. Arch Intern Med 1998;158:1525–31.

[12] Cushman M, Tsai AW, White RH, Heckbert SR, Rosamond WD, Enright P, et al. Deep vein thrombosis and pulmonary embolism in two cohorts: the longitudinal investigation of thromboembolism etiology. Am J Med Jul 1 2004;117:19–25.

[13] Arnason T, Wells PS, van Walraven C, Forster AJ. Accuracy of coding for possible warfarin complications in hospital discharge abstracts. Thromb Res 2006;118:253–62.

[14] Heckbert SR, Kooperberg C, Safford MM, Psaty BM, Hsia J, McTiernan A, et al. Comparison of self-report, hospital discharge codes, and adjudication of cardiovascular events in the Women's Health Initiative. Am J Epidemiol Dec 15 2004;160:1152–8.

[15] Birman-Deych E, Waterman AD, Yan Y, Nilasena DS, Radford MJ, Gage BF. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. Med Care May 2005;43:480–5.

[16] Zhan C, Battles J, Chiang YP, Hunt D. The validity of ICD-9-CM codes in identifying postoperative deep vein thrombosis and pulmonary embolism. Jt Comm J Qual Patient Saf Jun 2007;33:326–31.

[17] Center for Disease Control. ICD-9-CM Coordination and Maintenance Committee; 2010. p. 11–4.

[18] Forum NQ, cartographer^cartographers. National Voluntary Consensus Standards for the Prevention and Care of VTE Project. Washington DC: National Quality Forum; 2008.

[19] Procedure Classes 2009 [database on the Internet]. 2009 [cited March 16, 2009]. Available from: http://www.hcup-us.ahrq.gov/toolssoftware/procedure/procedure.jsp.

[20] Avrill RF, Goldfield N, Hughes JS, Bonazelli J, McCullough EC, Steinbeck BA, et al. All Patient Refined Diagnosis Related Groups (APR-DRGs) Version 20.0: Methodology Overview. Wallingford, CT: 3M Health Information Systems; 2003.

[21] Alonzo TA, Pepe MS, Lumley T. Estimating disease prevalence in two-phase studies. Biostatistics Apr 2003;4:313–26.

[22] Schoenman JA, Sutton JP, Elixhauser A, Love D. Understanding and Enhancing the Value of Hospital Discharge Data. Med Care Res Rev August 1 2007;64:449–68.

[23] Spencer FA, Emery C, Lessard D, Anderson F, Emani S, Aragam J, et al. The Worcester Venous Thromboembolism study: a population-based study of the clinical epidemiology of venous thromboembolism. J Gen Intern Med Jul 2006;21:722–7.

[24] AHRQ. Patient Safety Indicators; 2006 [updated 2006 2006]; cited 2008 July 6; Available from: http://qualityindicators.ahrq.gov/downloads/psi/2006-Feb-PatientSafetyIndicators.pdf.

[25] Center for Medicare & Medicaid Services. Hospital-Acquired Conditions (Present on Admission Indicator); 2007 [updated 2007; cited 2010 Jan 31, 2010]; Available from: http://www.cms.hhs.gov/HospitalAcqCond/.

[26] Heit JA, Silverstein MD, Mohr DN, Petterson TM, Lohse CM, O'Fallon WM, et al. The epidemiology of venous thromboembolism in the community. Thromb Haemost Jul 2001;86:452–63.

[27] Huerta C, Johansson S, Wallander MA, Garcia Rodriguez LA. Risk factors and short-term mortality of venous thromboembolism diagnosed in the primary care setting in the United Kingdom. Arch Intern Med May 14 2007;167:935–43.

[28] White RH. The epidemiology of venous thromboembolism. Circulation Jun 17 2003;107:I4–8.