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 © 2010 Elsevier Ltd. All rights reserved.
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