Received: 5 December 2023 | Revised: 18 April 2024 | Accepted: 26 April 2024

https://doi.org/10.1016/j.rpth.2024.102433

ORIGINAL ARTICLE

A real-time prognostic model for venous thromboembolic events among hospitalized adults

Benjamin F. Tillman[1] | Henry J. Domenico[2] | Ryan P. Moore[2] | Daniel W. Byrne[2] | Colleen T. Morton[1] | Amanda S. Mixon[3,4,5] | Benjamin French[2]

[1]Division of Hematology and Oncology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA

[2]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee, USA

[3]Department of Medicine, Center for Quality Aging, Vanderbilt University Medical Center, Nashville, Tennessee, USA

[4]Geriatric Research, Education and Clinical Center, Department of Veterans Affairs, Tennessee Valley Healthcare System, Nashville, Tennessee, USA

[5]Division of General Internal Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Correspondence
Benjamin F. Tillman, Division of Hematology and Oncology, Department of Medicine, Vanderbilt University Medical Center, 1161 21st Ave S, Nashville, TN 37232, USA.
Email: benjamin.f.tillman@vumc.org

Handling Editor: Vania M. Morelli

Abstract

Background: Hospital-acquired venous thromboembolism (HA-VTE) is a leading cause of morbidity and mortality among hospitalized adults. Guidelines recommend use of a risk-prediction model to estimate HA-VTE risk for individual patients. Extant models do not perform well for broad patient populations and are not conducive to automation in clinical practice.

Objectives: To develop an automated, real-time prognostic model for venous thromboembolism during hospitalization among all adult inpatients using readily available data from the electronic health record.

Methods: The derivation cohort included inpatient hospitalizations ("encounters") for patients ≥16 years old at Vanderbilt University Medical Center between 2018 and 2020 (n = 132,330). HA-VTE events were identified using International Classification of Diseases, 10th Revision, codes. The prognostic model was developed using least absolute shrinkage and selection operator regression. Temporal external validation was performed in a validation cohort of encounters between 2021 and 2022 (n = 62,546). Prediction performance was assessed by discrimination accuracy (C statistic) and calibration (integrated calibration index).

Results: There were 1187 HA-VTEs in the derivation cohort (9.0 per 1000 encounters) and 864 in the validation cohort (13.8 per 1000 encounters). The prognostic model included 25 variables, with placement of a central line among the most important predictors. Prediction performance of the model was excellent (C statistic, 0.891; 95% CI, 0.882-0.900; integrated calibration index, 0.001). The model performed similarly well across subgroups of patients defined by age, sex, race, and type of admission.

Conclusion: This fully automated prognostic model uses readily available data from the electronic health record, exhibits superior prediction performance compared with existing models, and generates granular risk stratification in the form of a predicted probability of HA-VTE for each patient.

KEYWORDS
inpatients, prognosis, risk, safety, venous thromboembolism

Essentials

- We developed a prognostic model for hospital-acquired venous thromboembolism among adults.
- The model uses data readily available in the electronic health record.
- The model exhibited excellent prediction performance across a wide range of patient subgroups.
- The prognostic model can be automated for real-time risk assessment for venous thromboembolism.

1 | INTRODUCTION

Venous thromboembolism (VTE), including deep vein thrombosis and pulmonary embolism, is a leading cause of morbidity and mortality worldwide [1]. The development of VTE often arises as a complication of an acute medical or surgical admission, with most events experienced by currently or recently hospitalized patients [2]. These healthcare-associated VTEs are a significant cause of morbidity and mortality. For adult inpatients, hospital-acquired VTE (HA-VTE) is the leading cause of preventable in-hospital death [3]. The Institute of Medicine (IOM) considers HA-VTE to be a preventable event, and the Agency for Healthcare Research and Quality (AHRQ) cites it as one of the most important safety issues facing hospitalized patients [4].

Considerable efforts have been undertaken to reduce the incidence of HA-VTE. Current practice recommends the use of risk-prediction models to estimate an individual's risk. However, even with recognition of the problem and the adoption of clinical decision support tools, HA-VTE incidence is increasing [5]. While provider prompts increase adoption of guideline-recommended therapies [6], practice patterns typically return to baseline if the prompt is discontinued [7], and appropriate interventions are frequently underutilized [8]. The need for automated and personalized prevention tools for at-risk patients remains a top safety priority [4].

There are numerous published models of varying complexity for VTE among hospitalized adults; several of the more widely used models are described in Supplementary Table S1 [9]. Existing models have been developed in specific cohorts of hospitalized patients (eg, medical, surgical, or critical care). Unfortunately, there is no consensus on the best prediction tool, and the existing options fall short of the ideal characteristics as outlined by the AHRQ. These guiding principles include accurately identifying high-risk while excluding low-risk patients, using information that is readily available, being easily automated into the electronic health record (EHR) for use in daily practice, and offering actionable recommendations to providers [4].

Previous attempts to automate risk-prediction models into the EHR have underestimated the VTE risk compared with manual calculations [10]. Other highly complex models incorporate variables unavailable to the provider at the initial evaluation, are best performed prior to admission, and require too much time for use in routine clinical practice [11]. To meet the need for an individualized and automated assessment of HA-VTE risk, we sought to develop a prognostic model using data available in the EHR that updates in real time and provides nonbinary risk prediction for providers at the time of the evaluation. We focused on VTEs that develop during hospitalization when providers are able to implement preventative measures.

2 | METHODS

2.1 | Study population

We obtained retrospective data on all adults (≥16 years old at the time of admission) admitted to Vanderbilt University Medical Center (VUMC) between January 1, 2018, and June 30, 2022. VUMC is a not-for-profit, university-affiliated, tertiary referral hospital in Nashville, Tennessee, that also serves as the region's level 1 trauma center and burn center. The study period was chosen to provide a large sample size with consistent data collection following VUMC's implementation of the Epic EHR in late 2017. The study cohort included all patients admitted to VUMC as inpatients, including elective and urgent admissions, inter-hospital transfers, and through the emergency department. Patients admitted more than once during the study period contributed data from each unique encounter to the analysis (an "encounter" is defined as an inpatient hospitalization and spans from the time that an admission order is placed until the time that a discharge order is placed). Patients admitted to the hospital under observation status were not included unless they were subsequently changed to inpatient status. For development of the prognostic model, we defined a derivation cohort of inpatient encounters between January 1, 2018, and December 31, 2020 (n = 132,330). For temporal external validation of the model, we defined a validation cohort of inpatient encounters between January 1, 2021, and June 30, 2022 (n = 62,546). This study was approved by the institutional review board of VUMC.

2.2 | Data elements

Data were obtained from VUMC's Research Derivative—a database of clinical data derived from the electronic data warehouse and restructured for research. The coding and data extraction were performed by Research Derivative staff members and supported by a Vanderbilt Institute for Clinical and Translational Research pilot grant.

We used International Classification of Diseases, 10th Revision (ICD-10) codes to identify cases of HA-VTE that were diagnosed during each eligible encounter (Supplementary Table S2). We did not include VTE events that developed after discharge. VTEs identified as being present on arrival were not defined as cases. The classification of a VTE as "present on arrival" as well as the ICD-10 coding were performed independently of this research project by hospital coding and billing staff in the office of Clinical Documentation Integrity. Staff members review clinical records to determine if certain conditions were present on arrival, including VTE. If there is uncertainty about a specific diagnosis, a query is sent to the managing clinical team. A random sample of 120 cases was selected for manual review (B.F.T.) to assess the accuracy of the algorithm for International Classification of Diseases (ICD)–based identification of VTE diagnosis that was not present on arrival, all of which were confirmed to be the correct diagnosis. For the manual review, radiologic identification (computed tomography, venous ultrasound, and vascular duplex) of VTE was used as the standard for confirmation of the diagnosis.

We screened 82 potential risk factors for HA-VTE: demographic and clinical characteristics, diagnostic procedures, vital signs, and laboratory measurements included on complete blood counts and chemistry panels. We focused on risk factors routinely available in EHRs that could be captured in real time. Included components were selected based on availability as well as association with increased risk of VTE. These elements were collated from other published risk-prediction models and an earlier site-specific VTE prediction model. Clinical characteristics were inferred using the established coding method for identification of Elixhauser comorbidity measures [12,13]. The ICD-10 codes used to calculate the Elixhauser comorbidity measures were extracted from billing codes (Supplementary Table S3), which include known comorbidities and new diagnoses established during the encounter. We did not include inpatient or outpatient medications or procedure codes.

For model development, we used the earliest available data during the encounter. Categorical risk factors with unknown values included "unknown" as a separate category; unknown values for continuous risk factors (eg, vital signs and laboratory measurements) were imputed using the median value in the derivation cohort [14].

2.3 | Statistical analysis

Patient characteristics were summarized using standard descriptive statistics. To develop the prognostic model, we used logistic regression models, which provide granular risk stratification by calculating the predicted probability of HA-VTE. We maximized the use of available data by avoiding the arbitrary and inefficient categorization of continuous variables and calculation of point-based scores [15]. Based on the derivation cohort (n = 132,330) with 1187 HA-VTEs, a logistic regression model could expend as many as 80 degrees of freedom (ie, 1 per 15 events) without overfitting [16]. More realistically, a model with <40 degrees of freedom would attain an out-of-sample mean absolute prediction error of <0.16%, which is an acceptable error rate for prediction of HA-VTE [17]. A parsimonious model that balanced complexity with accuracy was identified via the least absolute shrinkage and selection operator (LASSO), for which multiple values of the shrinkage parameter were considered to identify the model with the lowest degrees of freedom that maintained high discrimination accuracy as measured by the cross-validated C statistic. Odds ratios with 95% CIs were estimated by refitting the logistic regression model (without LASSO) with the set of variables selected by the LASSO. For categorical variables with more than 2 categories (eg, type of admission), if the LASSO selected at least 1 category, then all categories (compared with the reference category) were included in the model. The importance of each variable was quantified by its Wald chi-squared statistic obtained from the logistic regression model. In the primary analysis, continuous variables were modeled with linear terms, and the model did not include any interaction terms; in sensitivity analyses, we considered restricted cubic splines for continuous variables and included interactions of all variables with age and, separately, with a history of metastatic cancer.

Prediction performance of the final prognostic model was assessed in the derivation and validation cohorts based on discrimination and calibration. Discrimination (ie, the ability of the model to discriminate between cases of HA-VTE and controls without HA-VTE) was quantified by the C statistic (ie, the area under the receiver operating characteristic curve for binary outcomes) with 95% CIs, for which a value of 1 denotes perfect discrimination. Calibration (ie, the concordance between the observed and predicted probability of HA-VTE) was quantified by the integrated calibration index (ICI); the ICI is based on the difference between observed and predicted probabilities, and a value of 0 denotes perfect calibration [18]. Prediction performance was assessed among all patients and among patient subgroups defined by age (<60 or ≥60 years), sex (female or male), race (Asian or Pacific Islander, Black, Native American including Alaskan Native, or White), ethnicity (Hispanic or not Hispanic), type of admission (elective admissions, through the emergency department, for trauma, interhospital transfer), and history of metastatic cancer (yes or no) to ensure accuracy and fairness of the prognostic model. Due to an expected decrement in calibration when performing external validation, the prognostic model was also recalibrated to the validation cohort by fitting a logistic regression model with HA-VTE (as the outcome) and the linear predictor obtained by applying the regression coefficients from the original model to the observed data (as the predictor) [19]. All analyses were performed in R version 4.2.3 (R Foundation for Statistical Computing), including the Hmisc, rms, and glmnet extension packages. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis reporting guidelines for prognostic studies were followed [20].

3 | RESULTS

The derivation cohort included 132,330 encounters from 89,311 hospitalized adults (2018-2020); 67,810 (75.9%), 12,756 (14.3%), and 8745 (9.8%) adults had 1, 2, and 3 or more encounters, respectively (Supplementary Figure S1). Among encounters in the derivation cohort,

TABLE 1 Characteristics of inpatient encounters in the derivation cohort.

Characteristic	All encounters (*n* = 132,330)	Encounters with HA-VTE (*n* = 1187)	Encounters without HA-VTE (*n* = 131,143)
Age (y), median (IQR)	55.8 (37.0-68.4)	56.8 (42.9-67.8)	55.8 (36.9-68.4)
Female sex, *n* (%)	69,249 (52.3)	457 (38.5)	68,792 (52.5)
Race, *n* (%)			
Asian or Pacific Islander	2330 (1.8)	11 (0.9)	2319 (1.8)
Black	22,115 (16.7)	239 (20.1)	21,876 (16.7)
Native American	360 (0.3)	3 (0.3)	357 (0.3)
White	103,750 (78.4)	902 (76.0)	102,848 (78.4)
Other or unknown	5469 (4.1)	46 (3.9)	5423 (4.1)
Ethnicity, *n* (%)			
Hispanic	5032 (3.8)	39 (3.3)	4993 (3.8)
Not Hispanic	125,489 (94.8)	1123 (94.6)	124,366 (94.8)
Unknown	1809 (1.4)	25 (2.1)	1784 (1.4)
Type of admission, *n* (%)			
Elective	31,550 (23.8)	149 (12.6)	31401 (23.9)
Emergency department	67,412 (50.9)	575 (48.4)	66,837 (51.0)
Trauma	2476 (1.9)	82 (6.9)	2394 (1.8)
Interhospital transfer	30,362 (22.9)	374 (31.5)	29,988 (22.9)
Unknown	530 (0.4)	7 (0.6)	523 (0.4)
Heart rate (bpm), median (IQR)[a]	85 (72-101)	80 (93-113)	85 (72-100)
BMI (kg/m^2), median (IQR)[b]	28.1 (23.9-33.3)	27.4 (23.6-32.9)	28.1 (23.9-33.3)
Comorbidities, *n* (%)[c]			
Acute kidney injury	27,219 (20.6)	656 (55.3)	26,563 (20.3)
Candidal stomatitis	1347 (1.0)	61 (5.1)	1286 (1.0)
Cardiac arrhythmia	49,872 (37.7)	995 (83.8)	48,877 (37.3)
Cerebrovascular disease	12,081 (9.1)	233 (19.6)	11,848 (9.0)
Chronic pulmonary disease	24,838 (18.8)	257 (21.7)	24,581 (18.7)
Coagulopathy	14,088 (10.6)	497 (41.9)	13,591 (10.4)
Fluid or electrolyte disorder	48,235 (36.5)	985 (83.0)	47,250 (36.0)
Hypoxemia	12,825 (9.7)	522 (44.0)	12,303 (9.4)
Metastatic cancer	9435 (7.1)	124 (10.4)	9311 (7.1)
Myocardial infarction	15,990 (12.1)	243 (20.5)	15,747 (12.0)
Other anemia	21,585 (16.3)	451 (38.0)	21,134 (16.1)
Other psychiatric disorders	13,448 (10.2)	220 (18.5)	13,228 (10.1)
Paralysis	4466 (3.4)	102 (8.6)	4364 (3.3)
Peptic ulcer disease	2110 (1.6)	60 (5.1)	2050 (1.6)
Pleural disease	6110 (4.6)	293 (24.7)	5817 (4.4)
Pneumonia	10,979 (8.3)	483 (40.7)	10,496 (8.0)
Respiratory symptoms	22,602 (17.1)	508 (42.8)	22,094 (16.8)
Thrombosis	7672 (5.8)	109 (9.2)	7563 (5.8)
Venous disease	569 (0.4)	8 (0.7)	561 (0.4)
Weight loss	8753 (6.6)	310 (26.1)	8443 (6.4)

(Continues)

TABLE 1 (Continued)

Characteristic	All encounters (*n* = 132,330)	Encounters with HA-VTE (*n* = 1187)	Encounters without HA-VTE (*n* = 131,143)
Central line placed, *n* (%)	3892 (2.9)	405 (34.1)	3487 (2.7)
Sodium (mEq/L), median (IQR)[d]	138 (135-140)	138 (135-140)	138 (135-140)
Chloride (mEq/L), median (IQR)[e]	105 (101-108)	105 (101-108)	105 (101-108)
BUN (mg/dL), median (IQR)[f]	16 (11-24)	19 (13-30)	16 (11-24)
CRP (mg/dL), median (IQR)[g]	39.0 (10.2-115.9)	106.0 (32.0-202.2)	38.5 (10.1-113.8)

BMI, body mass index; BUN, blood urea nitrogen; CRP, C-reactive protein; HA-VTE, hospital-acquired venous thromboembolism.
[a]Unknown for 64,155 (48.5%) patients.
[b]Unknown for 9160 (6.9%) patients.
[c]Identified from International Classification of Diseases, 10th Revision, Clinical Modification codes using Elixhauser comorbidity measures.
[d]Unknown for 21,402 (16.2%) patients.
[e]Unknown for 22,382 (16.9%) patients.
[f]Unknown for 22,389 (16.9%) patients.
[g]Unknown for 113,367 (85.7%) patients.

there were 1187 HA-VTE events (9.0 per 1000 encounters). For encounters with vs without an HA-VTE, the in-hospital mortality rate was 179.4 vs 29.3 per 1000 encounters, respectively. Patients who developed HA-VTE were more likely to be male (Table 1). Among clinical characteristics, certain findings were more likely to be found among the group that developed HA-VTE, including the concomitant presence of a cardiac arrhythmia (83.3%) and central line placement (34.1%).

The validation cohort included 62,546 encounters from 46,968 hospitalized adults (2021-2022); 37,948 (80.8%), 5755 (12.3%), and 3265 (7.0%) adults had 1, 2, and 3 or more encounters, respectively (Supplementary Figure S1). Among encounters in the validation cohort, there were 864 HA-VTE events (13.8 per 1000 encounters). For encounters with vs without an HA-VTE, the in-hospital mortality rate was 224.5 vs 32.2 per 1000 encounters, respectively. Patients who developed HA-VTE were more likely to be older, male, and Black (Table 2). As in the derivation cohort, there were certain comorbidities such as oral candidiasis and weight loss that were more likely to be observed among the HA-VTE group in the validation cohort.

The characteristics of the derivation and validation cohorts were similar, with the same trends noted in both cohorts with respect to those with and without HA-VTE. Patient characteristics summarized according to the availability of vital signs and laboratory measurements are provided in Supplementary Table S4. Patients for whom heart rate (51.9%), metabolic panel measurements (82.8%), and C-reactive protein levels (14.7%) were known were older and had a greater burden of comorbidities and higher risk for HA-VTE; there were no meaningful differences by race, ethnicity, or calendar year.

There were 10,560 ICD-10 codes that identified the 2051 encounters with an HA-VTE across the derivation and validation cohorts (Supplementary Table S2); an individual patient with an HA-VTE at a particular encounter could have more than 1 documented diagnosis code at that encounter. Pulmonary embolism (2602, 24.6%) and acute embolism and thrombosis of deep veins of the lower extremity (2784, 26.4%) accounted for the majority of the identified codes, while almost 1 out of 3 ICD-10 codes identified an upper extremity event (3024, 28.6%).

The prognostic model identified via LASSO included 25 variables (Figure 1; Supplementary Figure S2). The strongest predictor of HA-VTE was placement of a central line (odds ratio, 3.13; 95% CI, 2.69-3.65; chi-squared statistic, 216.8). Other important predictors were type of admission, comorbidities (cardiac arrhythmia, fluid or electrolyte disorder, pneumonia, hypoxemia, recent weight loss, coagulopathy, pleural disease, other anemia, acute kidney injury, other psychiatric disorders, cerebrovascular disease, respiratory symptoms, peptic ulcer disease, candidal stomatitis, and paralysis), vital signs (heart rate), and laboratory measurements (C-reactive protein, blood urea nitrogen, sodium, and chloride). The estimated regression coefficients for calculating the predicted probability of HA-VTE are provided in Supplementary Table S5.

The model exhibited excellent discrimination accuracy (*C* statistic, 0.891; 95% CI, 0.882-0.900) and calibration (ICI, 0.001) in the derivation cohort (Table 3; Supplementary Figure S3). Modeling continuous variables using restricted cubic splines vs linear terms did not improve the model's prediction performance (*C* statistic, 0.892). Performance was also not improved by inclusion of interactions of all variables with age (*C* statistic, 0.894) and a history of metastatic cancer (*C* statistic, 0.895). We observed similar prediction performance across patient subgroups defined by age, sex, race, ethnicity, type of admission, and history of metastatic cancer (Figure 2). The model also exhibited excellent prediction performance in the validation cohort (*C* statistic, 0.897; 95% CI, 0.887-0.905; ICI without recalibration, 0.003; ICI with recalibration, 0.002; Supplementary Figure S4). Although the model exhibited suboptimal calibration at higher values for the predicted probability of HA-VTE (ie, ≥0.20), there were very few encounters in this higher-risk group: 699 of 132,330 (0.5%) in the derivation cohort and 441 of 62,546 (0.7%) in the validation cohort.

4 | DISCUSSION

There are numerous prediction models for evaluating a hospitalized patient's risk for developing HA-VTE, but the current models are

TABLE 2 Characteristics of inpatient encounters in the validation cohort.

Characteristic	All encounters (*n* = 62,546)	Encounters with HA-VTE (*n* = 864)	Encounters without HA-VTE (*n* = 61,682)
Age (y), median (IQR)	55.5 (36.5-68.8)	58.0 (44.5-67.7)	55.5 (36.4-68.8)
Female sex, *n* (%)	32,923 (52.6)	361 (41.8)	32,562 (52.8)
Race, *n* (%)			
Asian or Pacific Islander	972 (1.6)	12 (1.4)	960 (1.6)
Black	10,334 (16.5)	174 (20.1)	10,160 (16.5)
Native American	212 (0.3)	3 (0.3)	209 (0.3)
White	48,376 (77.3)	647 (74.9)	47,729 (77.4)
Other or unknown	3379 (5.4)	37 (4.3)	3342 (5.4)
Ethnicity, *n* (%)			
Hispanic	2682 (4.3)	20 (2.3)	2662 (4.3)
Not Hispanic	58,755 (93.9)	819 (94.8)	57,936 (93.9)
Unknown	1109 (1.8)	25 (2.9)	1084 (1.8)
Type of admission, *n* (%)			
Elective	14,995 (24.0)	113 (13.1)	14,882 (24.1)
Emergency department	28,596 (45.7)	348 (40.3)	28,248 (45.8)
Trauma	4690 (7.5)	147 (17.0)	4543 (7.4)
Interhospital transfer	14,198 (22.7)	256 (29.6)	13,942 (22.6)
Unknown	67 (0.1)	0 (0)	67 (0.1)
Heart rate (bpm), median (IQR)[a]	86 (72-101)	94 (78-110)	85 (72-101)
BMI (kg/m^2), median (IQR)[b]	28.2 (24.0-33.5)	27.9 (23.7-33.3)	28.2 (24.0-33.5)
Comorbidities, *n* (%)[c]			
Acute kidney injury	13,659 (21.8)	475 (55.0)	13,184 (21.4)
Candidal stomatitis	571 (0.9)	53 (6.1)	518 (0.8)
Cardiac arrhythmia	24,595 (39.3)	748 (86.6)	23,847 (38.7)
Cerebrovascular disease	5901 (9.4)	194 (22.5)	5707 (9.3)
Chronic pulmonary disease	12,167 (19.5)	218 (25.2)	11,949 (19.4)
Coagulopathy	7890 (12.6)	401 (46.4)	7489 (12.1)
Fluid or electrolyte disorder	24,422 (39.0)	738 (85.4)	23,684 (38.4)
Hypoxemia	6631 (10.6)	374 (43.3)	6257 (10.1)
Metastatic cancer	4398 (7.0)	77 (8.9)	4321 (7.0)
Myocardial infarction	7884 (12.6)	196 (22.7)	7688 (12.5)
Other anemia	11,550 (18.5)	326 (37.7)	11,224 (18.2)
Other psychiatric disorders	7563 (12.1)	170 (19.7)	7393 (12.0)
Paralysis	2267 (3.6)	101 (11.7)	2166 (3.5)
Peptic ulcer disease	948 (1.5)	40 (4.6)	908 (1.5)
Pleural disease	2965 (4.7)	187 (21.6)	2778 (4.5)
Pneumonia	5580 (8.9)	398 (46.1)	5182 (8.4)
Respiratory symptoms	10,340 (16.5)	392 (45.4)	9948 (16.1)
Thrombosis	3679 (5.9)	68 (7.9)	3611 (5.9)
Venous disease	320 (0.5)	3 (0.3)	317 (0.5)
Weight loss	3949 (6.3)	195 (22.6)	3754 (6.1)

(Continues)

TABLE 2 (Continued)

Characteristic	All encounters (*n* = 62,546)	Encounters with HA-VTE (*n* = 864)	Encounters without HA-VTE (*n* = 61,682)
Central line placed, *n* (%)	2120 (3.4)	281 (32.5)	1839 (3.0)
Sodium (mEq/L), median (IQR)[d]	138 (135-140)	138 (135-140)	138 (135-140)
Chloride (mEq/L), median (IQR)[e]	104 (101-107)	104 (101-107)	104 (101-107)
BUN (mg/dL), median (IQR)[f]	16 (11-25)	18 (12-30)	16 (11-24)
CRP (mg/dL), median (IQR)[g]	50.6 (13.5-124.4)	112.4 (41.8-183.3)	48.8 (13.0-121.0)

BMI, body mass index; BUN, blood urea nitrogen; CRP, C-reactive protein; HA-VTE, hospital-acquired venous thromboembolism.
[a]Unknown for 29,677 (47.4%) patients.
[b]Unknown for 4655 (7.4%) patients.
[c]Identified from International Classification of Diseases, 10th Revision, Clinical Modification codes using Elixhauser comorbidity measures.
[d]Unknown for 13,157 (21.0%) patients.
[e]Unknown for 13,379 (21.4%) patients.
[f]Unknown for 13,388 (21.4%) patients.
[g]Unknown for 52,957 (84.7%) patients.

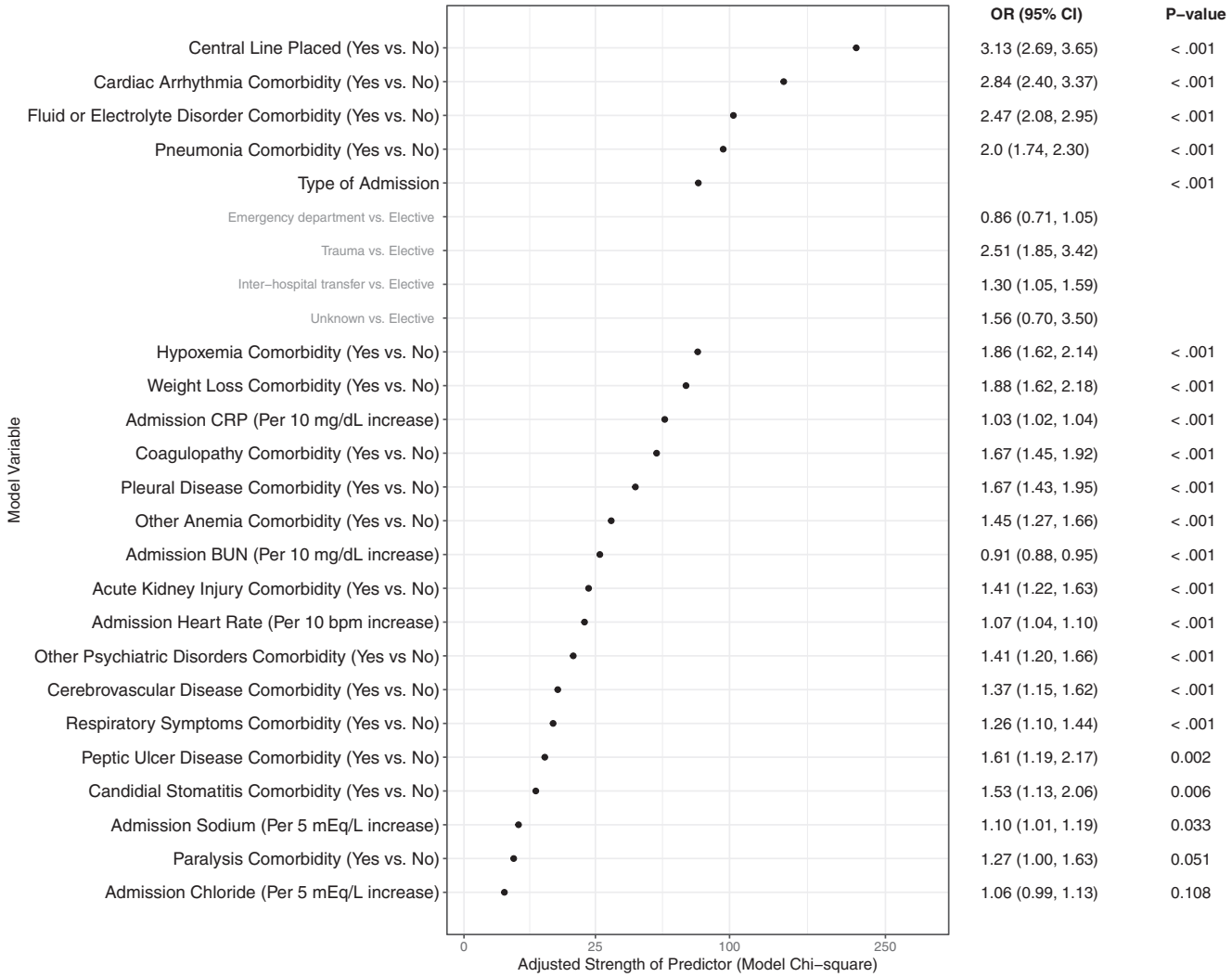

FIGURE 1 Relative importance of each variable in the final prognostic model, as quantified by its Wald chi-squared statistic along with odds ratios (ORs) and 95% CIs. BUN, blood urea nitrogen; CRP, C-reactive protein.

TABLE 3 Prediction performance of the final prognostic model.

Cohort	C statistic (95% CI)	ICI
Derivation cohort	0.891 (0.882-0.900)	0.001
Validation cohort		
Without recalibration	0.897 (0.887-0.905)	0.003
With recalibration	0.897 (0.887-0.905)	0.002

ICI, integrated calibration index.

limited by the need for manual provider input and dichotomous risk stratification of the outcome [9]. Our prognostic model presented here, which we named VTE among adult inpatients (VTE-AI), delivers a real-time risk assessment for individual patients using readily available data from the EHR to allow providers to make personalized decisions about risk and necessary preventative measures for each patient encounter. The incidence of inpatient development of an acute VTE remains clinically significant, occurring in between 0.9% and 1.4% of inpatient admissions in our study. The VTE-AI model has high discrimination accuracy with excellent prediction performance. The design and development of the VTE-AI model facilitate rapid identification of those at highest risk for HA-VTE.

How does the VTE-AI model improve on the existing landscape with multiple widely used prediction tools available? The model can be integrated into the EHR, providing an automated evaluation that updates in a dynamic fashion as the clinical course evolves. Prior attempts to automate the Padua score resulted in underestimation of patient risk compared with manual provider input [10]. Use of the Caprini score is cumbersome, requiring consideration of up to 36 variables, and does not lend itself to automation. Other widely used models such as the IMPROVE-VTE have not been validated in surgical or intensive care patient groups.

Similar to other published prediction tools, well-recognized high-risk factors for VTE such as the presence of a central line were strongly associated with risk for HA-VTE in our study; however, novel risk factors are identified here, such as the presence of a cardiac arrhythmia, weight loss, or oral candidiasis, through the EHR using a readily available comorbidity index algorithm. Our model is specific for the likelihood of the inpatient development of HA-VTE during the current hospitalization. This targeted outcome differentiates our model from others that include an estimate of VTE risk for up to several months after hospital discharge. There is a risk of development of VTE after discharge, which we have not included in this evaluation.

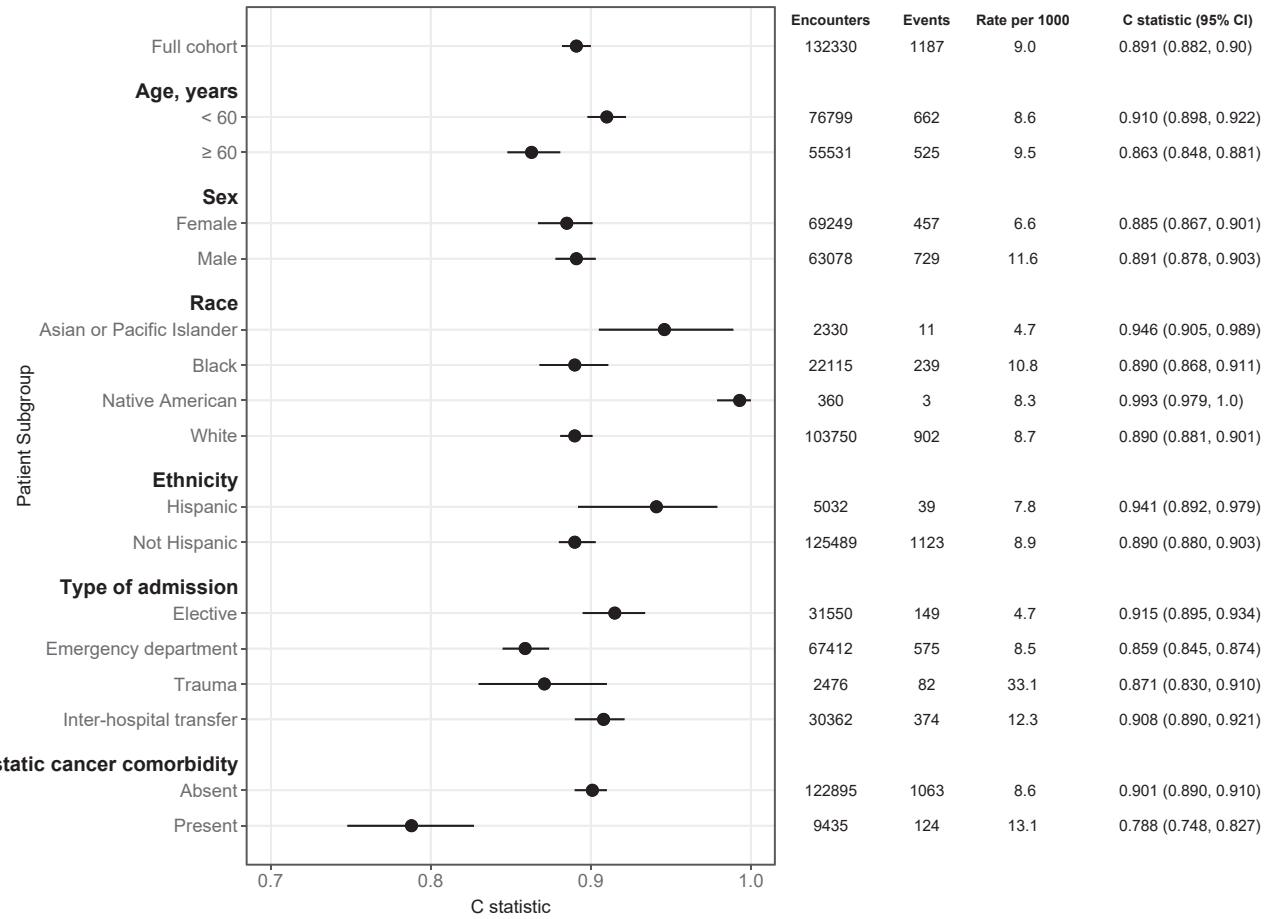

FIGURE 2 Prediction performance of the final prognostic model among patient subgroups in the derivation cohort, presented as *C* statistics with 95% CIs. Note: "Encounters" provides the number of encounters, "Events" provides the number of hospital-acquired venous thromboembolic events, and "Rate per 1000" provides the event rate per 1000 encounters.

Our model is intended to identify those patients at risk for the event during the hospital stay. The use of thromboprophylaxis after discharge remains unclear and is limited to select patient populations. Identification of individuals at risk for VTE during the hospital period helps reach the unmet need established by the IOM and AHRQ to prevent this in-hospital complication. Existing options are limited to specific patient populations (medical vs surgical) and are unable to be integrated into the EHR as a fully automated tool [21–23]. We have provided a prognostic model that works well across a broad and diverse inpatient population.

We acknowledge the following limitations. First, definitions for several of the variables—specifically comorbidity indices—are dependent on the use of ICD codes. The use of ICD codes to define a clinical phenotype has limitations. While these limitations are not unique to our healthcare system, these do impose challenges in extrapolation to other centers due to differences in the coding practices between institutions. To minimize this limitation of our model, we utilized the well-established Elixhauser comorbidity index, which encompasses a large number of ICD-9 and ICD-10 diagnostic codes to define a clinical phenotype of interest. Second, development of the prognostic model did not consider pharmacologic data for the current admission or outpatient prescription records. The use of prophylactic or systemic anticoagulant agents or antiplatelet agents is not available for our cohort. However, as those patients believed to be at higher risk by the estimation of the treating provider were more likely to receive pharmacologic thromboprophylaxis, there is a risk that the incorporation of preventative measures may be paradoxically associated with a higher event rate. The omission of pharmacologic data also excluded the use of medications with an increased VTE risk (eg, hormone therapy, chemotherapy, and certain immunomodulatory agents).

The VTE-AI model accurately identifies patients at risk for developing a VTE while in the hospital, which is identified by the IOM as one of the greatest risks for hospitalized patients. As a result, we did not identify those patients who might develop a thrombosis after discharge. The group of patients at risk for hospital-associated (distinct from hospital-acquired) VTE represents yet another subgroup of patients at great risk for VTE. While we recognize that this is an important subset of patients, our model is not designed to identify those who would benefit from intervention upon leaving the hospital.

Notable strengths of the VTE-AI model are the broad application across highly variable patients in a tertiary hospital system, granular stratification of the real-time risk, and ability for integration and automation into the existing EHR. Readily available items such as vital signs, laboratory measurements, and type of admission (transfer, emergent, urgent, etc.) are incorporated for all patients at the initiation of the counter. Missing variables are accounted for in the model development, and it even performs well without relying on traditionally heavily weighted variables such as presence of malignancy. While we did not include other predictive tool scores in the data extraction and development of this model for comparison, the *C* statistic of our model is similar to or higher than all existing models. We have implemented the VTE-AI model as a Shiny application at https://cqs.app.vumc.org/shiny/AdultVTEPrediction/.

In summary, the prediction performance of the VTE-AI model is excellent with strong discriminatory capabilities. It outperforms the existing risk-assessment tools without the need for manual provider input of variables or calculations. Prevention of HA-VTE highlights a major patient safety concern by the IOM and AHRQ. Our model focuses on HA-VTE and excels in this area across a diverse cohort of hospitalized patients in a tertiary care center. In keeping with the ideal VTE prevention tool as outlined by the AHRQ, our fully automated model accurately identifies patients at risk for development of VTE and reliably excludes low-risk patients. Automation that adapts to EHR integration suits this model for increased adoption and improved rates of implementation with meaningful provider prompts and risk prediction for the individual patient. The busy clinical provider needs a seamlessly incorporated prognostic tool to guide bedside decisions about prophylaxis and other interventions. The next step in the evaluation of this model is external validation in other populations and implementation in a prospective pragmatic evaluation.

FUNDING

This work was supported by the National Heart, Lung, and Blood Institute (award number R01-HL164482) and the Vanderbilt Institute for Clinical and Translational Research, which is funded by the National Center for Advancing Translational Sciences Clinical Translational Science Award Program (award number UL1-TR002243).

ETHICS STATEMENT

This study was approved by the institutional review board of Vanderbilt University Medical Center with a waiver of informed consent.

AUTHOR CONTRIBUTIONS

B.F.T. conceived the research with assistance from D.W.B., oversaw the electronic health record data extraction, and wrote the first draft of the manuscript with assistance from B.F.; H.J.D. and R.P.M. performed the statistical analysis under the direction of D.W.B. and B.F.; C.T.M. and A.S.M. provided input on model development and implementation. A.S.M. and B.F. obtained funding. All authors took part in reviewing and editing the manuscript and approved the final version of the manuscript.

RELATIONSHIP DISCLOSURE

There are no competing interests to disclose.

ORCID

Benjamin French https://orcid.org/0000-0001-9265-5378

REFERENCES

[1] Heit JA, Spencer FA, White RH. The epidemiology of venous thromboembolism. *J Thromb Thrombolysis*. 2016;41:3–14.

[2] Heit JA. Epidemiology of venous thromboembolism. *Nat Rev Cardiol*. 2015;12:464–74.

[3] Rathbun S. Cardiology patient pages. The Surgeon General's call to action to prevent deep vein thrombosis and pulmonary embolism. *Circulation*. 2009;119:e480–2.

[4] Maynard G. *Preventing hospital-associated venous thromboembolism: a guide for effective quality improvement*. 2nd ed. Rockville, MD: Agency for Healthcare Research and Quality; 2015. AHRQ Publication No. 16-0001-EF.
[5] Heit JA, Ashrani A, Crusan DJ, McBane RD, Petterson TM, Bailey KR. Reasons for the persistent incidence of venous thromboembolism. *Thromb Haemost*. 2017;117:390–400.
[6] Kucher N, Koo S, Quiroz R, Cooper JM, Paterno MD, Soukonnikov B, et al. Electronic alerts to prevent venous thromboembolism among hospitalized patients. *N Engl J Med*. 2005;352:969–77.
[7] Wheeler AP, Jaquiss RD, Newman JH. Physician practices in the treatment of pulmonary embolism and deep venous thrombosis. *Arch Intern Med*. 1988;148:1321–5.
[8] Bratzler DW, Raskob GE, Murray CK, Bumpus LJ, Piatt DS. Underuse of venous thromboembolism prophylaxis for general surgery patients: physician practices in the community hospital setting. *Arch Intern Med*. 1998;158:1909–12.
[9] Stuck AK, Spirk D, Schaudt J, Kucher N. Risk assessment models for venous thromboembolism in acutely ill medical patients. A systematic review. *Thromb Haemost*. 2017;117:801–8.
[10] Pavon JM, Sloane RJ, Pieper CF, Colón-Emeric CS, Cohen HJ, Gallagher D, et al. Automated versus manual data extraction of the Padua prediction score for venous thromboembolism risk in hospitalized older adults. *Appl Clin Inform*. 2018;9:743–51.
[11] Cronin M, Dengler N, Krauss ES, Segal A, Wei N, Daly M, et al. Completion of the updated Caprini risk assessment model (2013 version). *Clin Appl Thromb Hemost*. 2019;25:1076029619838052. https://doi.org/10.1177/1076029619838052
[12] Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36:8–27.
[13] Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43:1130–9.
[14] Janssen KJM, Vergouwe Y, Donders ART, Harrell Jr FE, Chen Q, Grobbee DE, et al. Dealing with missing predictor values when applying clinical prediction models. *Clin Chem*. 2009;55:994–1001.
[15] French B, Saha-Chaudhuri P, Ky B, Cappola TP, Heagerty PJ. Development and evaluation of multi-marker risk scores for clinical prognosis. *Stat Methods Med Res*. 2016;25:255–71.
[16] Riley RD, Snell KI, Ensor J, Burke DL, Harrell Jr FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: part II - binary and time-to-event outcomes. *Stat Med*. 2019;38:1276–96.
[17] van Smeden M, Moons KGM, de Groot JA, Collins GS, Altman DG, Eijkemans MJ, et al. Sample size for binary logistic prediction models: beyond events per variable criteria. *Stat Methods Med Res*. 2019;28:2455–74.
[18] Austin PC, Steyerberg EW. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. *Stat Med*. 2019;38:4051–65.
[19] Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17:230. https://doi.org/10.1186/s12916-019-1466-7
[20] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594. https://doi.org/10.1136/bmj.g7594
[21] Spyropoulos AC, Anderson Jr FA, FitzGerald G, Decousus H, Pini M, Chong BH, et al. Predictive and associative models to identify hospitalized medical patients at risk for VTE. *Chest*. 2011;140:706–14.
[22] Caprini JA, Arcelus JI, Reyna JJ. Effective risk stratification of surgical and nonsurgical patients for venous thromboembolic disease. *Semin Hematol*. 2001;38(Suppl 5):12–9.
[23] Chopard P, Spirk D, Bounameaux H. Identifying acutely ill medical patients requiring thromboprophylaxis. *J Thromb Haemost*. 2006;4:915–6.

SUPPLEMENTARY MATERIAL

The online version contains supplementary material available at https://doi.org/10.1016/j.rpth.2024.102433