Journal of the American Heart Association

ORIGINAL RESEARCH

Claims-Based Score for the Prediction of Bleeding in a Contemporary Cohort of Patients Receiving Oral Anticoagulation for Venous Thromboembolism

Alvaro Alonso, MD, PhD; Faye L. Norby, PhD, MPH; Richard F. MacLehose, PhD; Neil A. Zakai, MD, MSc; Rob F. Walker, MPH; Terrence J. Adam, MD, PhD; Pamela L. Lutsey, PhD

BACKGROUND: Current scores for bleeding risk assessment in patients with venous thromboembolism (VTE) undergoing oral anticoagulation have limited predictive capacity. We developed and internally validated a bleeding prediction model using healthcare claims data.

METHODS AND RESULTS: We selected patients with incident VTE initiating oral anticoagulation in the 2011 to 2017 MarketScan databases. Hospitalized bleeding events were identified using validated algorithms in the 180 days after VTE diagnosis. We evaluated demographic factors, comorbidities, and medication use before oral anticoagulation initiation as potential predictors of bleeding using stepwise selection of variables in Cox models run on 1000 bootstrap samples of the patient population. Variables included in >60% of all models were selected for the final analysis. We internally validated the model using bootstrapping and correcting for optimism. We included 165 434 patients with VTE and initiating oral anticoagulation, of whom 2294 had a bleeding event. After undergoing the variable selection process, the final model included 20 terms (15 main effects and 5 interactions). The c-statistic for the final model was 0.68 (95% CI, 0.67–0.69). The internally validated c-statistic corrected for optimism was 0.68 (95% CI, 0.67–0.69). For comparison, the c-statistic of the Hypertension, Abnormal Renal/Liver Function, Stroke, Bleeding History or Predisposition, Labile International Normalized Ratio, Elderly (>65 Years), Drugs/Alcohol Concomitantly (HAS-BLED) score in this population was 0.62 (95% CI, 0.61–0.63).

CONCLUSIONS: We have developed a novel model for bleeding prediction in VTE using large healthcare claims databases. Performance of the model was moderately good, highlighting the urgent need to identify better predictors of bleeding to inform treatment decisions.

Key Words: bleeding ■ MarketScan ■ oral anticoagulants ■ prediction ■ venous thromboembolism

One in 12 individuals will develop venous thromboembolism (VTE) during their lifetime.[1] Oral anticoagulation (OAC) is the cornerstone of treatment for patients with VTE, with current guidelines recommending that most of these patients receive at least 3 to 6 months of anticoagulation after their diagnosis.[2] Despite the potential risk of bleeding, the consequences of not treating acute VTE are severe enough that most individuals warrant anticoagulation for the primary treatment of VTE. Bleeding risk varies by choice of oral anticoagulant, with some of the newer oral agents having a lower major bleeding risk than warfarin.[3] Bleeding risk factors may also differ by anticoagulant choice. Therefore, accurately characterizing

Correspondence to: Alvaro Alonso, MD, PhD, Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Rd NE, CNR 3051, Atlanta, GA 30322. E-mail: alvaro.alonso@emory.edu

Preprint posted on medRxiv, February 3, 2021. doi: https://doi.org/10.1101/2021.02.01.21250924.

Supplementary Material for this article is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.121.021227

For Sources of Funding and Disclosures, see page 8.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- Using a claims database, we developed a novel model for bleeding prediction in venous thromboembolism that incorporates use of warfarin and direct oral anticoagulants as predictors.
- The model demonstrated similar predictive ability across subgroups of age, sex, and type of oral anticoagulant.

What Are the Clinical Implications?

- The model identified a high-risk group that included approximately half of all the bleeding events in this patient cohort.
- The overall predictive ability of the model, however, was not exceptional, despite the inclusion of a large number of predictors, indicating the need to identify more accurate predictors of bleeding risk.

Nonstandard Abbreviations and Acronyms

HAS-BLED	Hypertension, Abnormal Renal/Liver Function, Stroke, Bleeding History or Predisposition, Labile International Normalized Ratio, Elderly (>65 Years), Drugs/Alcohol Concomitantly
OAC	oral anticoagulation

individual patients' bleeding risk is key to tailoring individualized treatment choices for the management of acute VTE.[2] Over the years, several clinical prediction scores for major bleeding in patients with VTE have been developed to assist clinicians in this decision. These scores, however, were developed in cohorts with limited follow-up, did not compare bleeding risk across multiple oral anticoagulants, included small numbers of patients and bleeding events, and overall showed poor ability to discriminate risk.[4] The current guideline from the American College of Chest Physicians for VTE treatment does not specifically recommend the use of any of these scores. Instead, the guideline categorizes patients according to the number of risk factors for bleeding as low risk (no risk factors), moderate risk (1 risk factor), and high risk (≥2 risk factors).[2] Nonetheless, considerable variability in bleeding risk exists within each of these categories. Developing novel predictive models that quantify more accurately the risk of bleeding when receiving OAC is thus key to improve the care of people with acute VTE. To address this unmet need, we developed and internally validated a model for the prediction of bleeding in patients with VTE using a large healthcare claims database.

METHODS

Study Population

This study was conducted within the IBM MarketScan Commercial Claims and Encounters and Medicare Supplemental and Coordination of Benefits databases for the years 2011 through 2017. The MarketScan databases include individual-level Health Insurance Portability and Accountability Act–compliant healthcare claims information from employers, health plans, hospitals, and Medicare programs from across the United States.[5] Individual-level identifiers allow linkage across enrollment information and inpatient, outpatient, and pharmacy claims. The University of Minnesota Institutional Review Board deemed this research exempt from review and waived the need to obtain informed consent. Because of licensing restrictions, we cannot make available data and study materials to other investigators to reproduce results, but researchers may contact IBM Watson Health to obtain and license the data. This analysis included individuals aged ≥18 years with a diagnosis of VTE, at least one oral anticoagulant prescription within 1 month after VTE, no use of OAC before VTE diagnosis, and ≥90 days of continuous enrollment before their first oral anticoagulant prescription. We excluded dabigatran users because of small numbers (N=1141); there were no users of edoxaban. Patient follow-up was censored at 180 days after VTE

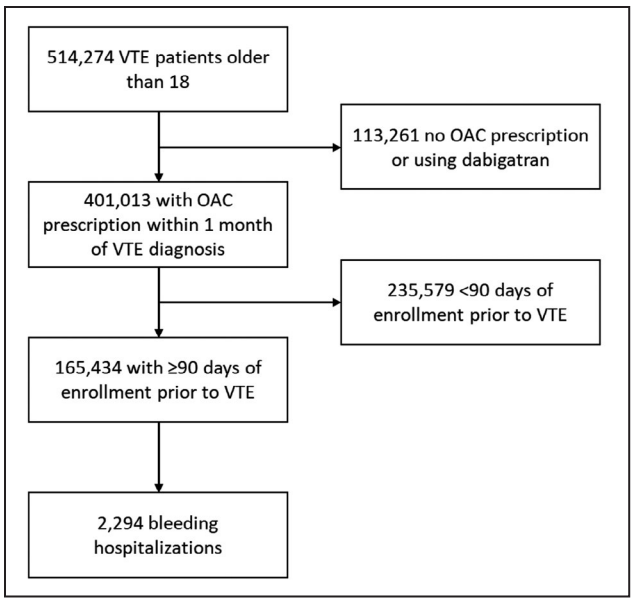

Figure 1. Flowchart of patient inclusion, MarketScan 2011 to 2017.
OAC indicates oral anticoagulation; and VTE, venous thromboembolism.

diagnosis. We defined VTE as having at least 1 inpatient claim or 2 outpatient claims 7 to 185 days apart, including any *International Classification of Diseases, Ninth* (*ICD-9*) or *Tenth revision* (*ICD-10*), code for VTE (Table S1) in any position. A validation study using a similar definition of VTE reported a 91% positive predictive value for this algorithm.[6]

Major Bleeding Events

The end point of interest was hospitalization for intracranial hemorrhage, gastrointestinal bleeding, or other major bleeding. Intracranial hemorrhage was defined as *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*), code 430.xx, 431.xx, or 432.xx or *International Classification of Diseases, Tenth Revision, Clinical Modification* (*ICD-10-CM*), code I60.xx, I61.xx, or I62.xx as the primary discharge diagnosis in an inpatient claim, with a positive predictive value of this definition estimated to be >90%.[7] Gastrointestinal bleeding and other major bleeding were defined using a previously described algorithm for identification of OAC-related bleeding that considers primary and secondary diagnosis in inpatient claims as well as the presence of transfusion codes.[8] Positive predictive value of this algorithm is close to 90%.[8]

Predictors

We identified potential predictors of bleeding from prior literature and existing risk scores. Predictors were defined according to validated algorithms when available using inpatient and outpatient *ICD* diagnosis codes and pharmacy claims.[9,10] Specifically, we considered the following 24 predictors: age, sex, hypertension, diabetes, chronic kidney disease, myocardial infarction, heart failure, ischemic stroke, transient ischemic attack, peripheral artery disease, chronic

Table 1. Characteristics of Patients With VTE by Anticoagulant Use, MarketScan 2011 to 2017

Characteristics	Overall	Warfarin	Rivaroxaban	Apixaban
Total No.	165 434	116 319	37 214	11 901
Age, y	58±16	59±16	56±15	60±16
Female sex	50	50	49	50
Hypertension	57	57	55	63
Diabetes	22	23	20	24
Alcohol abuse	0.9	0.7	1.2	2.2
Myocardial infarction	6.5	6.7	5.4	7.9
Heart failure	13	14	10	16
Ischemic stroke/TIA	11	12	9	8
Renal disease	10	11	7.0	13
Peripheral artery disease	13	13	11	15
Chronic pulmonary disease	27	27	26	27
Liver disease	8.8	8.6	9.4	9.6
Malignancy/metastatic cancer	18	18	16	17
Anemia	26	27	24	28
Thrombocytopenia	4.2	4.2	3.7	4.9
Peptic ulcer disease	0.7	0.7	0.7	0.7
Other previous bleeding	11	11	8.7	13
HAS-BLED score	1.7±1.3	1.7±1.3	1.6±1.3	1.8±1.3
Median (25th–75th percentile)	1 (1–2)	2 (1–3)	1 (1–2)	2 (1–3)
Warfarin	70	100	0	0
Rivaroxaban	22	0	100	0
Apixaban	7.1	0	0	100
Antiplatelets	6.2	6.6	4.9	6.6
NSAIDs	35	33	41	34
Gastroprotective drugs	29	29	29	31
SSRIs	28	28	28	28
Cytochrome P450 3A4 inhibitors	3.4	3.3	4.0	2.5

Values correspond to mean±SD or percentage, unless stated otherwise. HAS-BLED indicates Hypertension, Abnormal Renal/Liver Function, Stroke, Bleeding History or Predisposition, Labile International Normalized Ratio, Elderly (>65 Years), Drugs/Alcohol Concomitantly; SSRI, selective serotonin reuptake inhibitor; TIA, transient ischemic attack; and VTE, venous thromboembolism.

Table 2. Predictors of Bleeding Considered in Cox Regression Models, MarketScan 2011 to 2017

Predictor	No. of samples	β Coefficient	HR (95% CI)
Age, per year	1000	0.011	1.01 (1.008–1.014)
Malignancy/metastatic cancer	1000	0.355	1.43 (1.30–1.57)
Anemia	1000	0.500	1.65 (1.51–1.81)
Rivaroxaban (vs warfarin)	1000	−0.155	0.86 (0.77–0.95)
Apixaban (vs warfarin)	1000	−0.635	0.53 (0.43–0.65)
Antiplatelets	998	0.375	1.46 (1.27–1.66)
Liver disease	996	0.319	1.38 (1.22–1.55)
Diabetes	991	0.223	1.25 (1.14–1.37)
Other previous bleeding	986	0.265	1.30 (1.17–1.46)
Chronic pulmonary disease	930	0.182	1.20 (1.10–1.31)
Renal disease	896	0.213	1.24 (1.11–1.39)
Alcohol abuse	857	0.547	1.73 (1.26–2.36)
Female sex	818	0.130	1.14 (1.05–1.24)
Ischemic stroke/TIA	740	0.163	1.18 (1.05–1.32)
Thrombocytopenia	607	0.194	1.21 (1.03–1.43)
NSAIDs	552		
Gastroprotective drugs	520		
Heart failure	462		
Peptic ulcer disease	422		
SSRIs	397		
Hypertension	222		
Myocardial infarction	139		
Peripheral artery disease	88		
Cytochrome P450 3A4 inhibitors	42		

Number of samples indicates the times that a variable was included in any of the 1000 bootstrap samples. The β coefficient and HR (95% CI) are for the final model, including all covariates selected in >60% of the models. HR indicates hazard ratio; SSRI, selective serotonin reuptake inhibitor; and TIA, transient ischemic attack.

obstructive pulmonary disease, liver disease, cancer, previous bleeding, anemia, excessive alcohol consumption, thrombocytopenia, and peptic ulcer disease. We also considered the following medications: OAC type (warfarin, rivaroxaban, or apixaban), antiplatelets, nonsteroidal anti-inflammatory drugs, gastroprotective drugs (H2 receptor blockers, proton pump inhibitors, or others), selective serotonin reuptake inhibitors, and cytochrome p450 3A4 inhibitors (atazanavir, clarithromycin, indinavir, itraconazole, ketoconazole, nefazodone, ritonavir, saquinavir, buprenorphine, or telithromycin). We calculated the Hypertension, Abnormal Renal/Liver Function, Stroke, Bleeding History or Predisposition, Labile International Normalized Ratio, Elderly (>65 Years), Drugs/Alcohol Concomitantly (HAS-BLED) score based on claims-derived diagnoses, with the exception of labile international normalized ratio attributable to unavailability of this information.[11] Similarly, we calculated the VTE-BLEED score also using information from the claims data (including cancer, male patient with hypertension, anemia, history of bleeding, renal dysfunction, and age ≥60 years).[12] Table S2 provides a list of *ICD-9-CM* and *ICD-10-CM* codes used to define these covariates.

Statistical Analysis

We followed up patients who initiated OAC after a VTE diagnosis from the time of OAC initiation to first occurrence of major bleeding hospitalization, day 180 post-VTE diagnosis, or December 31, 2017, whichever occurred earlier.

To select predictors of bleeding risk, we ran a Cox proportional hazards model, including all the potential predictors listed above, with stepwise backward selection of variables using $P<0.05$ as the inclusion threshold. This process was repeated in 1000 bootstrap samples of the study population, and predictors included in >60% of the samples were selected for the final model.[13] Once the initial list of predictors for the final models was selected via this process, we examined interactions between age, sex, OAC type, and each one of the selected predictors. Individual interactions that were significant at $P<0.05$ were simultaneously added to the final model, and

those remaining statistically significant were kept. We evaluated the discriminatory value of the model by calculating the c-statistic and model calibration by comparing observed versus predicted probabilities by deciles of predicted risk. Model-based individual 180-day bleeding risk was calculated using the Breslow estimator, which is based on the empirical cumulative hazard function.[14]

Because we did not have access to an external data set, we performed an internal validation as recommended in existing guidelines for reporting of predictive models.[15] Internal validation was done by creating 500 bootstrap samples of the study population and calculating the c-statistic in each sample using the model derived in the previous step.[16] Because the model was derived and validated in the same data set, we corrected the c-statistic for optimism.[17]

To facilitate comparison of the discriminative ability of the new model with that of predictive models commonly used by clinicians, we calculated the c-statistic using the HAS-BLED score and the VTE-BLEED score.

RESULTS

The initial sample included 514 274 patients with VTE who were aged >18 years. After restricting to OAC users, the sample was composed of 401 013 patients. Requiring >90 days of enrollment before the first OAC prescription and excluding dabigatran users led to a final sample size of 165 434 patients with VTE. Follow-up was censored at 180 days after VTE diagnosis, which was attained by 76% of patients. During a mean (SD) follow-up time of 158 (46) days, we identified 2294 bleeding events (3.2 events per 100 person-years). Of these events, 207 were intracranial hemorrhages, 1371 were gastrointestinal bleeds, and 716 were other types of bleeding. Figure 1 provides a flowchart of patient inclusion in the analysis.

Table 1 shows descriptive characteristics of study patients overall and by type of OAC. Mean age (SD) of patients was 58 (16) years, and 50% were women. The mean (SD) HAS-BLED score was 1.7 (1.3). Patient characteristics across type of OAC were similar, except a slightly younger age and lower HAS-BLED score in rivaroxaban users than warfarin or apixaban users.

After running a stepwise Cox regression model in 1000 bootstrap samples, 15 variables were selected in >60% of the samples. Age, cancer history, anemia, and type of OAC were selected in all predictive models. Antiplatelets, liver disease, diabetes, previous bleeding, and chronic pulmonary disease were selected in 90% to 99% of the models, whereas renal disease, alcohol abuse, female sex, prior ischemic stroke/transient ischemic attack, and thrombocytopenia were selected in 60% to 89% of the models (Table 2).

Testing for interactions between age, sex, OAC class, and the covariates selected in the final model identified 10 interactions with $P<0.05$ (Table S3), most of them between age and comorbidities. After including these interactions in the final model, 5 of them remained significant. Table 3 shows the β coefficients and *P* values for all the significant predictors and their interactions in the final model. We have developed an Excel calculator that allows calculation of the predicted bleeding risk based on the patient characteristics (Table S4).

The c-statistic for the final model, including main effects and interactions, was 0.68 (95% CI, 0.67–0.69). Calibration of the model, assessed by

Table 3. β Coefficients, SEs, and *P* Values for Bleeding Predictors Selected in Final Model, MarketScan 2011 to 2017

Predictor	β Coefficient	SE	*P* value
Age, per year	0.021	0.002	<0.001
Female sex	0.211	0.051	<0.001
Diabetes	0.216	0.047	<0.001
Alcohol abuse	0.528	0.160	0.001
Ischemic stroke/TIA	0.182	0.057	0.001
Renal disease	0.233	0.058	<0.001
Chronic pulmonary disease	0.184	0.045	<0.001
Liver disease	0.294	0.062	<0.001
Malignancy/metastatic cancer	1.318	0.234	<0.001
Anemia	1.269	0.185	<0.001
Thrombocytopenia	0.180	0.083	0.03
Other previous bleeding	1.192	0.232	<0.001
Rivaroxaban (vs warfarin)	−0.182	0.059	0.002
Apixaban (vs warfarin)	−0.763	0.126	<0.001
Antiplatelets	0.379	0.068	<0.001
Age*cancer	−0.012	0.003	<0.001
Age*anemia	−0.012	0.003	<0.001
Age*previous bleed	−0.016	0.004	<0.001
Female sex*cancer	−0.347	0.093	<0.001
Rivaroxaban*previous bleed	0.212	0.141	0.13
Apixaban*previous bleed	0.577	0.238	0.02

The 1-year risk of bleeding can be calculated as follows: 1−(0.98768)^Exp[0.021*(age−58.2)+0.211*(female sex−0.499)+0.216*(diabetes−0.221)+0.528*(alcohol abuse−0.009)+0.182*(ischemic stroke/TIA−0.111)+0.233*(renal disease−0.101)+0.184*(chronic pulmonary disease−0.266)+0.294*(liver disease−0.088)+1.318*(cancer−0.177)+1.269*(anemia−0.264)−0.180*(thrombocytopenia−0.041)+1.192*(other previous bleeding−0.108)−0.182*(rivaroxaban−0.225)−0.763*(apixaban−0.072)+0.379*(antiplatelets−0.062)−0.012*(age*cancer−11.5)−0.012*(age*anemia−16.3)−0.016*(age*previous bleed−6.57)−0.347*(female sex*cancer−0.088)+0.212 (rivaroxaban*previous bleed−0.020)+0.577*(apixaban*previous bleed−0.009)]. TIA indicates transient ischemic attack.

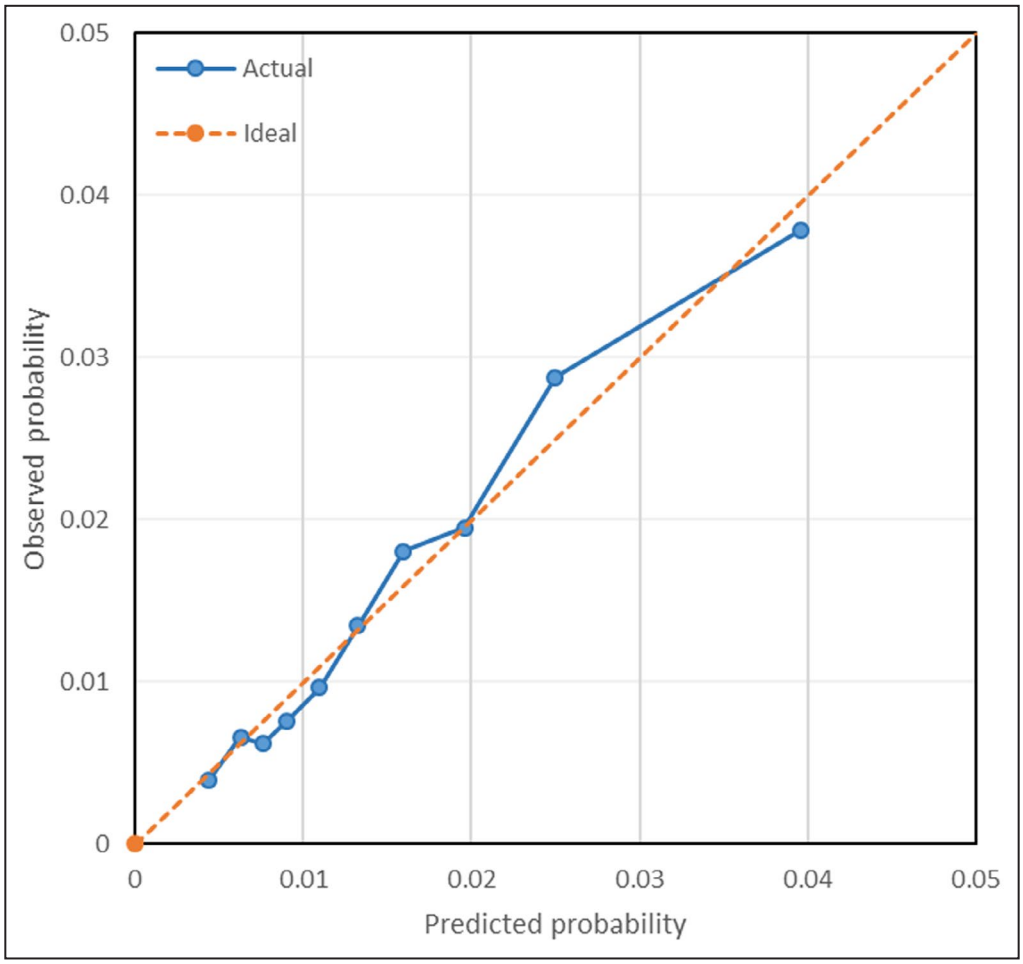

Figure 2. Calibration of predictive model, MarketScan 2011 to 2017.
The plot shows the predicted vs observed probabilities by deciles of predicted risk (blue circles). Perfect calibration corresponds to the orange dashed line.

comparing observed and predicted probabilities across deciles of predicted probabilities, was adequate (Figure 2). Patients in the top 2 deciles of predicted risk were at particularly high risk of bleeding (>2% over 180 days). Figure 3 shows the cumulative risk of bleeding by categories of predicted risk (low or <1%, moderate or 1%–<2%, and high or ≥2%). Patients in the high-risk category accounted for 24% of the sample and 48% of all the bleeding events. Corresponding figures were 36% and 35% for the moderate-risk group and 40% and 17% for the low-risk group, respectively. Correcting the c-statistic for optimism using 500 bootstrap samples resulted in essentially the same discrimination (c-statistic, 0.68; 95% CI, 0.67–0.69). The c-statistic was similar when the model was applied to prediction of events during the first 90 days of follow-up (n=1609 events; c-statistic, 0.67; 95% CI, 0.65–0.68).

Discrimination of the model was similar in men and women, slightly better in younger patients, and slightly better in the direct oral anticoagulants apixaban and rivaroxaban users compared with warfarin users (Table 4). The model showed better ability to predict intracranial hemorrhages and gastrointestinal bleeds than other types of bleeding (Table 4). Calibration was adequate across all subgroups of age category, sex, and type of OAC, and for the different types of bleeding.

The c-statistic for the HAS-BLED score (minus labile international normalized ratio) was 0.62 (95% CI, 0.61–0.63), whereas the c-statistic for the VTE-BLEED score was 0.65 (95% CI, 0.64–0.66). Dichotomizing the VTE-BLEED score as ≥2 (high risk) and <2 (low risk) resulted in a lower c-statistic (0.61; 95% CI, 0.60–0.62). Both the HAS-BLED and VTE-BLEED scores performed slightly better in direct OAC users than in warfarin users (Table S5).

DISCUSSION

We have developed and internally validated a model for the prediction of bleeding in patients with VTE based on information available in healthcare claims.

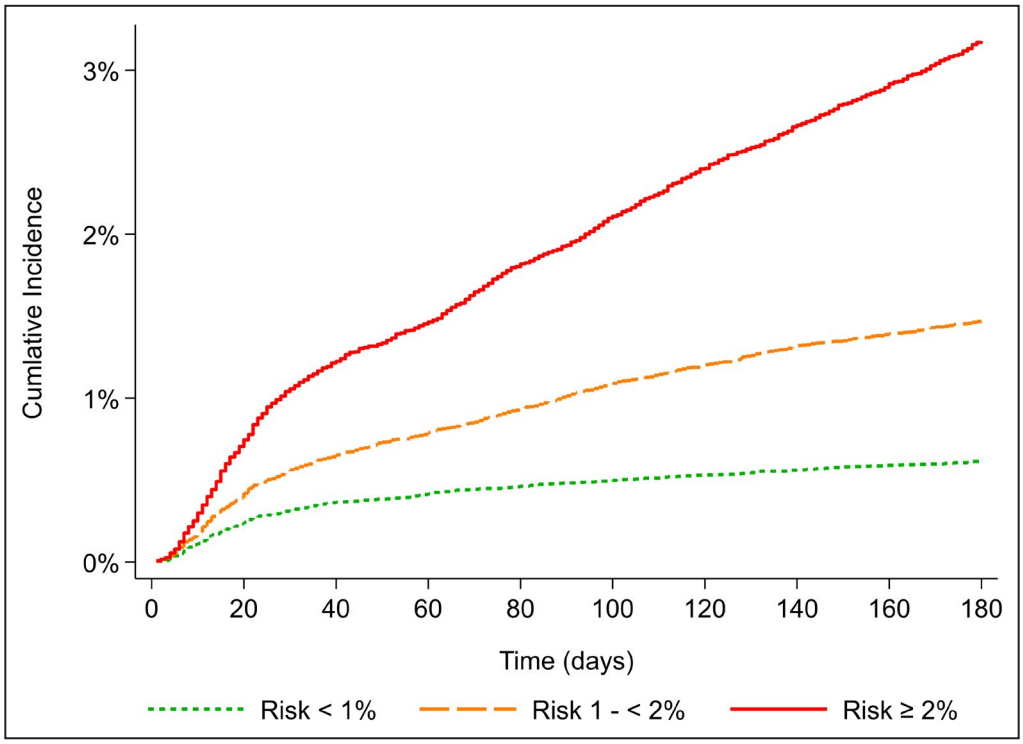

Figure 3. Cumulative incidence of hospitalized major bleeding by categories of 180-day predicted risk (<1%, 1%–<2%, and ≥2%), MarketScan 2011 to 2017.

The model identified a high-risk group that included approximately half of all the bleeding events in this patient cohort. In addition, the model performed similarly across different subgroups and had better discrimination than the established HAS-BLED score, in this data set. The overall predictive ability of the model, however, was not exceptional, despite the inclusion of a large number of predictors.

Identifying patients with VTE at high risk of bleeding complications from OAC is an unmet clinical need. Even though OAC for the primary treatment of VTE will outweigh almost any bleeding risk, clinicians still need to make decisions about the length of primary treatment (which could be affected by that bleeding risk), patients may demand objective information about the risks of complications, and patient characteristics may interact with the type of OAC to increase bleeding risk. However, existing predictive models and scores, such as the HAS-BLED and VTE-BLEED scores, have consistently shown mediocre performance when assessed by measures like the c-statistic.[18] The model developed in this analysis performed slightly better than the HAS-BLED and VTE-BLEED scores, but not well enough to warrant extensive application. The limited ability of this newly developed algorithm and previous scores to predict major bleeding can be attributable, in part, to heterogeneity in the outcome, with different bleeding types having specific risk factors. Nonetheless, the information in the model has clinical relevance. First, it confirms that use of direct oral anticoagulants, particularly apixaban, instead of warfarin could result in overall lower bleeding risk in this patient group, as demonstrated in randomized trials and real-world effectiveness studies.[3,19,20] Second, it identifies several comorbidities linked with increased bleeding risk. Whether better management of these comorbidities (eg, anemia, diabetes, or alcohol abuse) results in lower bleeding risk merits further study. Third, our algorithm identified a significant interaction between prior bleeding and type of OAC, whereby the protective associations of apixaban and rivaroxaban are negated in patients with prior bleeding. This is a group underrepresented in clinical trials and, therefore, deserving of further study. Although a substudy of the ARISTOTLE (Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation) trial did not find differences in the risk of bleeding of apixaban versus warfarin by prior bleeding history, this analysis was underpowered because of the small number of patients with such history.[21] Finally, being derived from claims data, this predictive model could be easily implemented and automatically calculated in electronic health record systems, making it easier to inform clinicians' decisions.

Table 4. Discrimination and Calibration of the Prediction Model Overall and Across Subgroups

Group	C-statistic (95% CI)	Calibration χ^2 (*P* value)
Full sample	0.682 (0.671–0.692)	27.6 (0.001)
Women	0.666 (0.651–0.681)	9.1 (0.43)
Men	0.699 (0.683–0.714)	14.4 (0.11)
Aged ≤58 y	0.688 (0.670–0.707)	16.3 (0.06)
Aged >58 y	0.652 (0.639–0.666)	16.7 (0.05)
Warfarin users	0.669 (0.656–0.681)	15.9 (0.07)
Rivaroxaban users	0.705 (0.679–0.730)	15.8 (0.07)
Apixaban users	0.709 (0.657–0.760)	15.4 (0.08)
Intracranial hemorrhage	0.720 (0.687–0.752)	6.5 (0.69)
Gastrointestinal bleed	0.716 (0.703–0.729)	20.9 (0.01)
Other bleeding	0.662 (0.641–0.681)	12.1 (0.21)

In addition to the HAS-BLED score, there are several other scores for the prediction of bleeding in patients with VTE receiving OACs.[4] These scores generally include few variables and are weighted using a point system to facilitate clinical application. Our model, in contrast, includes a larger number of predictors than other scores and does not assign points to each variable. Rather, it uses precise information from the model coefficients to calculate risk in a predictive equation. This approach has the advantage of not discarding information and providing more accurate predictions. We have successfully used this approach in the past to develop models for prediction of bleeding and stroke in patients with atrial fibrillation.[10,22]

The present analysis has some notable strengths. The model has been developed in a large population, including a sizable number of bleeding events, one order of magnitude larger than many other published scores. It also takes advantage of the extensive information on comorbidities and medication use available through claims data. Finally, it incorporates direct OACs as predictors, making it more applicable in the contemporary clinical setting. However, this analysis also has weaknesses, most importantly the lack of an external validation sample and the limitations in the validity of claims diagnoses to define bleeding and comorbidities, despite the use of validated algorithms. Although our end point definition had a high positive predictive value, potential for false negatives attributable to reduced sensitivity exists, resulting in underestimates of the predicted risk. Comparisons with the HAS-BLED score are likely biased in favor of our algorithm because we developed and tested the algorithm in the same data set and HAS-BLED was developed in a separate data set. Other shortcomings are the inadequate sample size to properly evaluate model performance in subgroups, unavailability of information on laboratory values and clinical measurements, absence of information on over-the-counter drug use, and the potential lack of generalizability to uninsured individuals.

In conclusion, we developed a novel model to predict bleeding in patients with VTE receiving OAC using healthcare claims information. This model predicted slightly better than a modified HAS-BLED score, but overall predictive performance was wanting. Additional work is needed to evaluate whether additional types of information, such as biomarkers, genetic factors, or drug therapy problem data, and alternative approaches for variable selection and statistical modeling could improve predictive ability of this or other models. Future research should also aim to develop predictive models for bleeding risk in patients with VTE needing long-term OAC treatment and to combine prediction of VTE recurrence and bleeding to facilitate decisions to clinicians and patients.

ARTICLE INFORMATION

Received February 8, 2021; accepted July 26, 2021.

Affiliations

Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA (A.A.); Center for Cardiac Arrest Prevention, Department of Cardiology, Cedars-Sinai Smidt Heart Institute, Los Angeles, CA (F.L.N.); Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN (F.L.N., R.F.M., R.F.W., P.L.L.); Division of Hematology/Oncology, Department of Medicine and Department of Pathology and Laboratory Medicine, Larner College of Medicine at the University of Vermont, Burlington, VT (N.A.Z.); and Department of Pharmaceutical Care and Health Systems, College of Pharmacy, University of Minnesota, Minneapolis, MN (T.J.A.).

Sources of Funding

Research reported in this publication was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under award numbers R01HL1311579, R01HL122200, and K24HL148521. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosures

None.

Supplementary Material

Tables S1-S3, S5
Table S4

REFERENCES

1. Bell EJ, Lutsey PL, Basu S, Cushman M, Heckbert SR, Lloyd-Jones DM, Folsom AR. Lifetime risk of venous thromboembolism in two cohort

studies. *Am J Med*. 2016;129:339.e319–339.e326. doi: 10.1016/j.amjmed.2015.10.014
2. Kearon C, Akl EA, Ornelas J, Blaivas A, Jimenez D, Bounameaux H, Huisman M, King CS, Morris TA, Sood N, et al. Antithrombotic therapy for VTE disease: CHEST Guideline and Expert Panel Report. *Chest*. 2016;149:315–352. doi: 10.1016/j.chest.2015.11.026
3. Lutsey PL, Zakai NA, MacLehose RF, Norby FL, Walker RF, Roetker NS, Adam TJ, Alonso A. Risk of hospitalised bleeding in comparisons of oral anticoagulant options for the primary treatment of venous thromboembolism. *Br J Haematol*. 2019;185:903–911. doi: 10.1111/bjh.15857
4. van Es N, Wells PS, Carrier M. Bleeding risk in patients with unprovoked venous thromboembolism: a critical appraisal of clinical prediction scores. *Thromb Res*. 2017;152:52–60. doi: 10.1016/j.thromres.2017.02.016
5. IBM Watson Health. *IBM MarketScan Research Databases for Health Services Research-White Paper*. Somers; 2018.
6. Sanfilippo KM, Wang T-F, Gage BF, Liu W, Carson KR. Improving accuracy of international classification of diseases codes for venous thromboembolism in administrative data. *Thromb Res*. 2015;135:616–620. doi: 10.1016/j.thromres.2015.01.012
7. Andrade SE, Harrold LR, Tjia J, Cutrona SL, Saczynski JS, Dodd KS, Goldberg RJ, Gurwitz JH. A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. *Pharmacoepidemiol Drug Saf*. 2012;21(suppl 1):100–128.
8. Cunningham A, Stein CM, Chung CP, Daugherty JR, Smalley WE, Ray WA. An automated database case definition for serious bleeding related to oral anticoagulant use. *Pharmacoepidemiol Drug Saf*. 2011;20:560–566. doi: 10.1002/pds.2109
9. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, Saunders LD, Beck CA, Feasby TE, Ghali WA. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43:1130–1139. doi: 10.1097/01.mlr.0000182534.19832.83
10. Claxton JNS, MacLehose RF, Lutsey PL, Norby FL, Chen LY, O'Neal WT, Chamberlain AM, Bengtson LGS, Alonso A. A new model to predict major bleeding in patients with atrial fibrillation using warfarin or direct oral anticoagulants. *PLoS One*. 2018;13:e0203599. doi: 10.1371/journal.pone.0203599
11. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJGM, Lip GYH. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. *Chest*. 2010;138:1093–1100. doi: 10.1378/chest.10-0134
12. Klok FA, Hösel V, Clemens A, Yollo WD, Tilke C, Schulman S, Lankeit M, Konstantinides SV. Prediction of bleeding events in patients with venous thromboembolism on stable anticoagulation treatment. *Eur Respir J*. 2016;48:1369–1376. doi: 10.1183/13993003.00280-2016
13. Austin PC, Tu JV. Bootstrap methods for developing predictive models. *Am Stat*. 2004;58:131–137. doi: 10.1198/0003130043277
14. Breslow NE. Discussion of Professor Cox's paper. *J R Stat Soc B*. 1972;34:216–217.
15. 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. doi: 10.1136/bmj.g7594
16. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med*. 2004;23:2109–2123. doi: 10.1002/sim.1802
17. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361–387. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361:AID-SIM168>3.0.CO;2-4
18. Vedovati MC, Mancuso A, Pierpaoli L, Paliani U, Conti S, Ascani A, Galeotti G, Di Filippo F, Caponi C, Agnelli G, et al. Prediction of major bleeding in patients receiving DOACs for venous thromboembolism: a prospective cohort study. *Int J Cardiol*. 2020;301:167–172. doi: 10.1016/j.ijcard.2019.11.105
19. The EINSTEIN-PE Investigators. Oral rivaroxaban for the treatment of symptomatic pulmonary embolism. *N Engl J Med*. 2012;366:1287–1297. doi: 10.1056/NEJMoa1113572
20. Agnelli G, Buller HR, Cohen A, Curto M, Gallus AS, Johnson M, Masiukiewicz U, Pak R, Thompson J, Raskob GE, et al. Oral apixaban for the treatment of acute venous thromboembolism. *N Engl J Med*. 2013;369:799–808. doi: 10.1056/NEJMoa1302507
21. Garcia DA, Fisher DA, Mulder H, Wruck L, De Caterina R, Halvorsen S, Granger CB, Held C, Wallentin L, Alexander JH, et al. Gastrointestinal bleeding in patients with atrial fibrillation treated with apixaban or warfarin: insights from the Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation (ARISTOTLE) trial. *Am Heart J*. 2020;221:1–8. doi: 10.1016/j.ahj.2019.10.013
22. Claxton JS, MacLehose RF, Lutsey PL, Norby FL, Chen LY, O'Neal WT, Chamberlain AM, Bengtson LGS, Alonso A. A new model to predict ischemic stroke in patients with atrial fibrillation using warfarin or direct oral anticoagulants. *Heart Rhythm*. 2019;16:820–826. doi: 10.1016/j.hrthm.2018.12.005

SUPPLEMENTAL MATERIAL

Table S1. International Classification of Disease (ICD) Clinical Modification (CM) diagnosis codes for venous thromboembolism (VTE).

Version	VTE codes
ICD-9-CM	415.1x, 451.1x, 453.2, 453.4x, 453.82, 453.83, 453.84, 453.85, 453.86, 453.87, 453.89, 453.9
ICD-10-CM	I26.0x, I26.9x, I80.1x, I80.20x, I82.210, I80.22x, I80.23x, I80.29x, I82.40x, I82.41x, I82.42x, I82.43x, I82.44x, I82.49x, I82.4Yx, I82.4Zx, I82.60x, I82.62x, I82.890, I82.A1x, I82.B1x, I82.C1x

Table S2. International Classification of Disease (ICD) Clinical Modification (CM) diagnosis codes for comorbidities considered as potential predictors.

Condition	ICD-9-CM	ICD-10-CM
Alcohol abuse	265.2, 291.1, 291.2, 291.3, 291.5, 291.6, 291.7, 291.8, 291.9, 303.0, 303.9, 305.0, 357.5, 425.5, 535.3, 571.0, 571.1, 571.2, 571.3, 980, V11.3	F10, E52, G62.1, I42.6, K29.2, K70.0, K70.3, K70.9, T51.x, Z50.2, Z71.4, Z72.1
Anemia	280.x-284.x, 285.1, 285.2, 285.3, 285.8, 285.9	D50.x-D53.x, D55.x-D64.x
Chronic pulmonary disease	416.8, 416.9, 490.x-505.x, 506.4, 508.1, 508.8	I27.8, I27.9, J40.x-J47.x, J60.x-J67.x, J68.4, J70.1, J70.3
Diabetes	250.x	E10.0-E10.9, E11.0-E11.9, E12.0-E12.9, E13.0-E13.9, E14.0-E14.9
Heart failure	398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 425.4-425.9, 428.x	I09.9, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5-I42.9, I43.x, I50.x, P29.0
Hypertension	401.x, 402.x, 403.x, 404.x, 405.x	I10.x, I11.x-I13.x, I15.x
Ischemic stroke / TIA	362.34, 430.x-438.x	G45.x, G46.x, H34.0, I60.x-I69.x
Liver disease	070.22, 070.23, 070.32, 070.33, 070.44, 070.54, 070.6, 070.9, 456.0, 456.1, 456.2, 570, 571, 572.2, 572.3, 572.4, 572.5, 572.6, 572.7, 572.8, 573.3, 573.4, 573.8, 573.9, V42.7	B18.x, K70.0-K70.3, K70.9, K71.3-K71.5, K71.7, K73.x, K74.x, K76.0, K76.2-K76.4, K76.8, K76.9, Z94.4, I85.0, I85.9, I86.4, I98.2, K70.4, K71.1, K72.1, K72.9, K76.5, K76.6, K76.7
Malignancy / metastatic cancer	140.x-172.x, 174.x-195.8, 196.x-199.x, 200.x-208.x, 238.6	C00.x-C26.x, C30.x-C34.x, C37.x-C41.x, C43.x, C45.x-C58.x, C60.x-C76.x, C77.x-C80.x, C81.x-C85.x, C88.x, C90.x-C97.x
Myocardial infarction	410.x, 412.x	I21.x, I22.x, I25.2
Peptic ulcer disease	533.x	K27.0-K27.7, K27.9

Peripheral artery disease	093.0, 437.3, 440.x, 441.x, 443.1-443.9, 47.1, 557.1, 557.9, V43.4	I70.x, I71.x, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9
Renal disease	403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93, 582, 583.0, 583.1, 583.2, 583.3, 583.4, 583.5, 583.6, 583.7, 585, 586, 588.0, V42.0, V45.1, V56	I12.0, I13.1, N03.2-N03.7, N05.2-N05.7, N18.x, N19.x, N25.0, Z49.0-Z49.2, Z94.0, Z99.2
Thrombocytopenia	287.1, 287.3, 287.4, 287.5	D69.3, D69.6

Table S3. Significant interactions identified in the development of the prediction model.

Interactions	P for interaction
Age*ischemic stroke	0.002
Age*renal disease	0.001
Age*liver disease	0.002
Age*cancer	<0.0001
Age*anemia	<0.001
Age*thrombocytopenia	0.04
Age*previous bleed	<0.001
Sex*liver disease	0.04
Sex*cancer	<0.001
OAC class*previous bleed	0.03

All interactions of age, sex, and OAC type with the variables included in the model were tested. Only those with p-value <0.05 were retained for further testing.

Table S4. Bleeding risk calculator (see separate Excel file).

Table S5. Discrimination of the HAS-BLED and VTE-BLEED scores by type of oral anticoagulant.

	HAS-BLED	VTE-BLEED
	C-statistic (95% CI)	
Warfarin	0.612 (0.599-0.624)	0.638 (0.626-0.651)
Rivaroxaban	0.642 (0.616-0.668)	0.668 (0.642-.694)
Apixaban	0.647 (0.587-0.708)	0.678 (0.625-0.731)