Received: 8 June 2023 | Revised: 9 November 2023 | Accepted: 11 November 2023

https://doi.org/10.1016/j.rpth.2023.102285

ORIGINAL ARTICLE

Risk scores for major bleeding from direct oral anticoagulants: comparing predictive performance in patients with atrial fibrillation

Alessandra M. Campos-Staffico[1] | Juliet P. Jacoby[1] | Michael P. Dorsch[1] | Nita A. Limdi[2] | Geoffrey D. Barnes[3] | Jasmine A. Luzum[1]

[1]Department of Clinical Pharmacy, College of Pharmacy, University of Michigan, Ann Arbor, Michigan, USA

[2]Department of Neurology, School of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA

[3]Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA

Correspondence
Jasmine A. Luzum, Department of Clinical Pharmacy, College of Pharmacy, University of Michigan, 428 Church St, Ann Arbor, MI 48109, USA.
Email: jluzum@med.umich.edu

Handling Editor: Dr Kristen Sanfilippo

Abstract

Background: Despite direct oral anticoagulants (DOACs) being safer than warfarin for stroke prevention in atrial fibrillation (AF), major bleeding concerns persist. Most bleeding risk scores predate DOAC approval.

Objectives: This study aimed to compare the Age, history of Bleeding, and non-bleeding related Hospitalisation [ABH] score's performance—derived for DOAC-treated patients—with those of 5 other scores (Anticoagulation and Risk Factors in Atrial Fibrillation [ATRIA], Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly [>65 years], Drugs/alcohol concomitantly [HAS-BLED], Hepatic, Hepatic or Renal Disease, Ethanol Abuse, Malignancy, Older Age, Reduced Platelet Count or Function, Re-Bleeding, Hypertension, Anemia, Genetic Factors, Excessive Fall Risk and Stroke [HEMORR$_2$HAGES], Outcomes Registry for Better Informed Treatment of Atrial Fibrillation [ORBIT-AF], and Congestive heart failure, Hypertension, Age ≥75 [doubled], Diabetes, Stroke [doubled]-Vascular disease, Age 65–74, Sex category [CHA$_2$DS$_2$-VASc]) in predicting DOAC-related major bleeding in patients with AF.

Methods: In this retrospective study of 2364 patients with nonvalvular AF on rivaroxaban or apixaban (median age, 68.3 years; 32.1% women), International Society on Thrombosis and Haemostasis–defined major bleeding (incidence, 4.1%; n = 97) was analyzed. C-statistics from time-dependent receiver operating characteristic (ROC) curves for continuous risk scores were the primary comparison metric, but other metrics, such as decision curves, were also compared.

Results: At 100 days, C-statistics were highest for ORBIT-AF and ATRIA (0.62 and 0.61, respectively, with other scores having an area under the ROC curve of <0.60); some significant differences favored ORBIT-AF. At 1100 days, C-statistics remained highest for ORBIT-AF and ATRIA (0.62 and 0.61, respectively, with other scores having an area under the ROC curve of <0.60 again), and ORBIT-AF had significantly higher C-statistics than those for all other risk scores (P < .05), except for ATRIA. At 2100 days, all C-statistics were <0.60 with no significant differences. Decision curves showed

the greatest net benefit for ORBIT-AF and ATRIA at 100 days and for ATRIA at 1100 days, with no discernible net benefit for any of the scores at 2100 days.

Conclusion: ORBIT-AF and ATRIA provided the best bleeding risk prediction within the first 1100 days. None of the 6 bleeding risk scores provided predictive benefit over 2100 days of DOAC treatment.

KEYWORDS
anticoagulants, atrial fibrillation, blood coagulation, factor Xa inhibitors, hemorrhage

Essentials

- Direct oral anticoagulants (DOACs) are safer than warfarin, but concerns about bleeding persist.
- Comparative performance of 6 scores in predicting major bleeding from DOACs was assessed.
- Outcomes Registry for Better Informed Treatment of Atrial Fibrilation [ORBIT-AF] and Anticoagulation and Risk Factors in Atrial Fibrillation [ATRIA] showed the highest C-statistics in the initial 1100 days of DOAC treatment.
- Outcomes Registry for Better Informed Treatment of Atrial Fibrillation [ORBIT-AF] and Anticoagulation and Risk Factors in Atrial Fibrillation [ATRIA] outperformed 4 other risk scores in identifying major bleeding from DOACs.

1 | INTRODUCTION

Atrial fibrillation (AF) is a persistent cardiac arrhythmia that increases the risk of ischemic strokes [1], whose risk can be reduced by approximately 66% with appropriate use of anticoagulants [2]. Currently, direct oral anticoagulants (DOACs) are the guideline-recommended first-line therapy for AF as they are as effective as warfarin in preventing cerebral ischemia [3] with lower rates of bleeding overall [4]. Despite broad efforts to raise awareness about the benefits of long-term anticoagulation in patients with AF, a recent study of more than 1.2 million Americans diagnosed with AF at increased risk of stroke revealed that merely half of the patients were prescribed oral anticoagulants [5]. This underutilization of oral anticoagulants was primarily attributed to prior major bleeding episodes [5].

Bleeding can be a life-threatening condition that often arises due to the use of anticoagulants, even in patients who are unsuspecting for the less obvious risk factors such as genetics, drug interactions, and nutrition [6]. Between 2011 and 2019, there was a significant increase in the number of visits to the emergency department for anticoagulant-related bleeding [7,8]. This upward trend mirrors the sharp rise in the number of anticoagulant prescriptions in the United States during this timeframe [9]. Although DOACs have a lower overall risk of bleeding [4], there is still a residual annual risk of major bleeding ranging from 2.1% to 4.9% [10–12] that contributed to a significant increase in emergency department visits for DOAC-related bleeding from 2.3% to 37.9% between 2011 and 2017 [13]. Furthermore, experiencing bleeding increases the likelihood of patients with AF discontinuing anticoagulants by 20% [14], ultimately increasing their risk of thromboembolic events [15]. Therefore, early identification of individuals at high risk of major bleeding is crucial in tailoring anticoagulation and preventing the potentially catastrophic consequences that may unfold with major bleeding.

To improve the accuracy of predicting the risk of anticoagulant-related bleeding, various risk prediction scores have been derived by integrating clinical and genetic characteristics. One guideline-recommended score for assessing bleeding risk in patients to be prescribed anticoagulant drugs is the Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly (>65 years), Drugs/alcohol concomitantly [HAS-BLED] score [3]. However, most of the risk stratification scores for predicting major bleeding in patients with AF were derived before the DOAC era (eg, Outcomes Registry for Better Informed Treatment of Atrial Fibrillation [ORBIT-AF], Anticoagulation and Risk Factors in Atrial Fibrillation [ATRIA], Hepatic or Renal Disease, Ethanol Abuse, Malignancy, Older Age, Reduced Platelet Count or Function, Re-Bleeding, Hypertension, Anemia, Genetic Factors, Excessive Fall Risk and Stroke [HEMORR$_2$HAGES], and HAS-BLED) [16–19]. Therefore, these scores may compute variables related to the risk of bleeding from vitamin K antagonists (VKAs) that are not applicable to DOACs, such as labile international normalized ratio in HAS-BLED score and *CYP2C9* genetic variant in HEMORR$_2$HAGES score [17,18]. DOACs and VKAs only share the common feature of oral anticoagulants, as their pharmacokinetic and pharmacodynamic properties differ enormously [20]. Considering that these pharmacological intricacies can influence the susceptibility to bleeding, employing risk scores derived from patient cohorts undergoing DOAC treatment would likely enhance the performance of predicting bleeding risks associated with these medications. Although there are several scoring systems available for predicting the risk of DOAC-related bleeding, the Age, history of Bleeding, and non-bleeding related Hospitalisation (ABH) score stands out as the unique score that has been derived exclusively from a cohort consisting of patients on DOAC therapy, with its primary aim focused on assessing the risk of major bleeding events as well as clinically relevant non-major bleeding from these drugs in the patient population with AF [21]. However, there is a lack of evidence comparing the performance of

the ABH score in the prediction of DOAC-related bleeding to those of other risk prediction scores. Hence, this study aimed to address this knowledge gap by comparing the performance of 5 DOAC-unspecific bleeding risk scores (ATRIA, HAS-BLED, HEMORR$_2$HAGES, ORBIT-AF, and Congestive heart failure, Hypertension, Age ≥75 [doubled], Diabetes, Stroke [doubled]-Vascular disease, Age 65–74, Sex category [CHA$_2$DS$_2$-VASc]) to the DOAC-specific ABH score in predicting the risk of major bleeding in a real-world cohort of patients with AF treated with either rivaroxaban or apixaban.

2 | METHODS

2.1 | Study design and cohort identification

This single-center retrospective cohort study was conducted with 2364 outpatients diagnosed with nonvalvular AF who were treated with either rivaroxaban or apixaban. The study cohort consisted of outpatients with AF who received care at Michigan Medicine—a large academic healthcare system affiliated with the University of Michigan. Individuals were identified automatically based on AF diagnosis in the electronic health record, followed by the first prescription of rivaroxaban or apixaban at Michigan Medicine. Individuals were evaluated to determine if they met the criteria for inclusion or exclusion in the study. The flowchart detailing the selection process of the patients can be found in Supplementary Figure S1. Since this is a substudy of a previously published pharmacogenetic study [22], patients aged ≥18 years who self-identified as White, and were genotyped and treated for nonvalvular AF either with rivaroxaban or with apixaban between October 1, 2012, and August 31, 2022, were considered eligible for this cohort. Patients were excluded if they (i) were diagnosed with moderate-to-severe mitral stenosis; (ii) had a history of mechanical valve replacement; (iii) were in stage 5 of chronic kidney disease, defined as a creatinine clearance of less than 15 mL/min estimated by the Cockcroft–Gault equation [23]; (iv) required renal replacement therapy; (v) were diagnosed with severe liver disease, including those with nonalcoholic fatty liver disease, cirrhosis, total bilirubin twice as high as the upper limit of normal with aspartate transaminase, alanine transaminase or alkaline phosphatase 3 times as high as the upper limit of normal, or other severe liver impairment as noted by the physician; (vi) were not routinely followed-up by Michigan Medicine; and (vii) did not have genotype data available through the Michigan Genomics Initiative biobank [24]. The study was conducted in accordance with principles laid out in the Declaration of Helsinki and was approved by the local institutional review board with a waiver of informed consent.

2.2 | Clinical and biochemical evaluation

The eligible patients' clinical and biochemical data were retrieved from the University of Michigan data warehouse by trained investigators, as previously explained elsewhere [25]. The first day of DOAC therapy was established as the index date for the study. The patients' active problems list was used to identify comorbidities using the code list specified in Supplementary Table S1 or keyword note search at the baseline period, which was defined as 1 year before the index date. The duration of DOAC treatment was determined by reviewing the thorough prescription history and conducting a keyword search using a natural language processing software, Electronic Medical Record Search Engine (EMERSE, manufactured by the Regents of the University of Michigan) [26]. A standardized data collection method was used, and study data were gathered using a customized form in Research Electronic Data Capture (REDCap) [27,28]. All electronic data capture tools are hosted at the University of Michigan.

2.3 | Drug–drug interaction

Drug–drug interactions with DOACs were defined as the concomitant and systemic use of either cytochrome P450 (CYP)/p-glycoprotein (p-gp) inhibitors or inducers with good or excellent documentation of evidence on Micromedex (Truven Health Analytics) [29]. Twelve drugs were considered as CYP/p-gp inhibitors: (i) amiodarone, (ii) clarithromycin, (iii) conivaptan, (iv) cyclosporine, (v) diltiazem, (vi) dronedarone, (vii) fluconazole, (viii) itraconazole, (ix) ritonavir, (x) erythromycin, (xi) ketoconazole, and (xii) verapamil. Seven drugs were considered as CYP/p-gp inducers for drug interactions with DOACs: (i) apalutamide, (ii) carbamazepine, (iii) fosphenytoin, (iv) nevirapine, (v) oxcarbazepine, (vi) phenytoin, and (vii) rifampin.

2.4 | Study endpoints and adjudication of events

The primary endpoint of this study was major bleeding according to the International Society on Thrombosis and Haemostasis definition [30]. Any hemorrhagic event during the period of DOAC medication (established within the start and stop dates of prescriptions obtained by the medical chart reviews) was considered DOAC-related bleeding. Briefly, major bleeding was defined as clinically overt nonsurgical bleeding with the symptomatic presentation and (i) fatal outcome and/or (ii) involvement of critical anatomical area or site such as intracranial, spinal, intraocular followed by vision changes, pericardial, articular, retroperitoneal, intramuscular with compartment syndrome, and/or (iii) hemoglobin fall of 2 g/dL or more, or leading to transfusion of 2 or more units of whole blood or red cells. As the International Society on Thrombosis and Haemostasis criteria are purely clinical and there is no validated algorithm to help identify bleeding cases electronically to date, trained clinicians examined the medical records of all patients using the EMERSE [26] to adjudicate and classify bleeding events. The patients' follow-up time was calculated as the difference between censoring and index dates. Patients were censored at the time of their first major bleeding event; the stop date of the DOAC; or, in the case of no stop date of the DOAC, the date when medical notes were last reviewed.

2.5 | Bleeding scores and risk classification

Six different prediction risk scores were used to estimate the risk of bleeding in patients with AF who were taking rivaroxaban or apixaban. These scores were ascertained as described in their original articles and based on clinical and biochemical characteristics at baseline, as outlined in Supplementary Table S2. While most of the scores were derived to predict bleeding risk in patients with AF on anticoagulation, 1 score, CHA_2DS_2-VASc, was originally intended to predict stroke risk but has also been used to predict major bleeds in patients with AF [31].

2.6 | Incidence rate of major bleeding

The continuous risk scores were categorized as low, intermediate, or high risk based on the predetermined cutoffs, as shown in Supplementary Table S3. Then, the incidence rate of major bleeding per 100 person-years for each risk category was determined so that comparisons could be made between our findings and the findings of previous studies. Specifically, the incidence rate per 100 person-years was calculated by dividing the total number of major bleeding events by the number of person-years for each risk category and multiplying by 100. Additionally, the cumulative incidence of major bleeding was defined as the number of bleeding events divided by the total number of individuals in each risk category.

2.7 | Statistical analysis

Patients were stratified into 2 subgroups based on the presence or absence of DOAC-related major bleeding. Baseline characteristics were outlined for both the entire patient cohort and these subgroups. Categorical variables were expressed as counts and percentages, while continuous variables were expressed as mean ± SD (if normally distributed) or as median and IQR (if not normally distributed). Subgroup differences were assessed using the χ^2 or Fisher's exact test for categorical variables, and Student's *t*-test or Mann–Whitney U-test for continuous variables. To assess the normality of distribution, the Kolmogorov–Smirnov test was employed.

The primary metric to compare predictive performance was C-statistics from time-dependent receiver operating characteristic (ROC) analyses of the risk scores as continuous variables both over the entire follow-up duration and at specified time points. Temporal fluctuations in the area under the ROC curve (AUROC) alongside their corresponding 95% CIs for each predictive risk score were plotted against follow-up time. To further assess the predictive performance of the 6 risk scores, an exploratory analysis was conducted using continuous risk scores to estimate and compare the AUROC at 3 time points: 100, 1100, and 2100 days. Subsequently, true positive rates (sensitivity) were plotted against the false positive rates for major bleeding (1 - specificity) to obtain ROC curves. The comparison of the AUROC was assessed through the Hanley–McNeil test [32], and an AUROC of <0.7 was considered to represent poor risk discrimination [33].

Several other secondary metrics were also used to compare different dimensions of risk score performance. With the aim of enabling the comparative assessment of incidence rates across risk categories, continuous risk scores were condensed into predefined risk categories in accordance with the cutoffs outlined in their original publications (Supplementary Table S3). The incidence rate of major bleeding per 100 person-years was estimated for each risk category. Additionally, calibration plots were employed to compare our incidence rates with prior incidence rate findings across risk categories.

Diagnostic efficiency measures (ie, sensitivity, specificity, negative predictive value, positive predictive value, Youden's index, and Matthew's correlation coefficient [MCC]) were computed for patients stratified into high-risk categories for major bleeding based on risk prediction scores. These measures were contrasted with the reference group composed of a composite of low- and intermediate-risk categories.

The individual probability of experiencing major bleeding was predicted through Cox proportional hazards regression models for each of the 6 risk scores at 100, 1100, and 2100 days of follow-up. Subsequently, the net benefit of predicting major bleeding risk was assessed through decision curve analysis [34].

The assessment encompassing time-dependent AUROC and decision curve analyses were performed using the "timeROC" and "dcurves" packages, respectively. For statistical analyses, both RStudio (version 4.0.2, Posit, PBC) and STATA (version 15.0, StataCorp LLC) for Mac were employed. A 2-sided *P* value of <.05 was considered statistically significant.

3 | RESULTS

3.1 | Patient sample and baseline characteristics

A cohort of 2364 outpatients diagnosed with nonvalvular AF and treated with either rivaroxaban or apixaban was identified within Michigan Medicine's records. Of these, 97 patients (4.1%) experienced major bleeding incidents over an average follow-up period of 2.27 years. The baseline characteristics of these patients stratified by the occurrence of major bleeding over the entire follow-up time are presented in Table 1. Comparative analysis unveiled that patients who experienced major bleeding were significantly older, less frequently prescribed apixaban, exhibited a higher frequency of switching between different DOACs, and had more frequent drug–drug interactions involving DOACs and CYP/p-gp inhibitors. They also displayed a significantly higher prevalence of conditions such as hypertension and chronic obstructive pulmonary disease, transcatheter aortic valve replacement, and a history of smoking habit compared to their counterparts who did not experience major bleeding. Furthermore, patients who encountered major bleeding had significantly lower diastolic blood pressure, glomerular filtration rate, and hemoglobin levels than those who did not experience bleeding (as outlined in Table 1). All the assessed risk prediction scores had significantly higher values for patients with major bleeding than their counterparts. All other baseline variables displayed no significant differences across the study subgroups.

TABLE 1 Clinical and biochemical baseline characteristics of the patients overall and whether or not they had major bleeding over the entire follow-up time.

Characteristics	Overall	Major bleeding		*P* value
		No	Yes	
Study patients, *n* (%)	2364 (100.0)	2267 (95.9)	97 (4.1)	_
Age (y)	68.3 (13.6)	68.1 (13.5)	74.1 (9.1)	**<.001**
Sex, *n* (%)				.125
Female	758 (32.1)	720 (31.8)	38 (39.2)	
Male	1606 (67.9)	1547 (68.2)	59 (60.8)	
Ethnicity, *n* (%)				*1.000*
Non-Hispanic or Latino	2312 (97.8)	2216 (97.8)	96 (99.0)	
Hispanic or Latino	10 (0.4)	10 (0.4)	0 (0.0)	
Unknown	42 (1.8)	41 (1.8)	1 (1.0)	
DOAC, *n* (%)				**.006**
Rivaroxaban	802 (33.9)	772 (34.1)	30 (30.9)	
Apixaban	1324 (56.0)	1276 (56.3)	48 (49.5)	
Both DOACs not simultaneously	238 (10.1)	219 (9.7)	19 (19.6)	
Daily dose of DOAC (mg), mean ± SD				
Rivaroxaban	19.3 ± 3.5	19.3 ± 3.5	18.7 ± 2.7	.308
Apixaban	7.8 ± 4.1	7.7 ± 4.1	8.7 ± 3.6	.056
Cumulative dose of DOAC (g), mean ± SD				
Rivaroxaban	16.5 ± 15.5	16.6 ± 15.5	16.0 ± 14.7	.816
Apixaban	7.2 ± 6.5	7.2 ± 6.5	7.1 ± 6.3	.424
Drug–drug interactions with DOACs, *n* (%)				
CYP/p-gp inhibitors	1046 (44.5)	983 (43.4)	63 (65.0)	**<.001**
CYP/p-gp inducers	33 (1.4)	31 (1.4)	2 (2.1)	.568
Hypertension, *n* (%)	933 (39.5)	885 (39.0)	48 (49.5)	**.039**
Diabetes mellitus, *n* (%)	421 (17.8)	398 (17.6)	23 (23.7)	.121
Heart failure, *n* (%)	321 (13.6)	303 (13.4)	18 (18.6)	.144
Stroke, *n* (%)	50 (2.1)	47 (2.1)	3 (3.1)	.494
Transient ischemic attack, *n* (%)	37 (1.6)	35 (1.5)	2 (2.1)	.687
Thromboembolism, *n* (%)	90 (3.8)	85 (3.8)	5 (5.2)	.479
Previous bleeding, *n* (%)	491 (20.8)	465 (20.5)	26 (26.8)	.135
COPD, *n* (%)	157 (6.6)	145 (6.4)	12 (12.4)	**.021**
Surgery or trauma, *n* (%)	1449 (61.3)	1381 (60.9)	68 (70.1)	.069
Transcatheter aortic valve replacement, *n* (%)	22 (0.9)	17 (0.8)	5 (5.2)	**<.001**
Elixhauser comorbidities score, mean ± SD	11.0 ± 11.3	10.9 ± 11.2	13.9 ± 12.0	**<.001**
ABH score, mean ± SD	3.0 ± 1.8	3.0 ± 1.8	3.9 ± 2.0	**<.001**
ATRIA score, mean ± SD	1.9 ± 2.0	1.9 ± 2.0	3.0 ± 1.9	**<.001**
HAS-BLED score, mean ± SD	1.2 ± 1.0	1.2 ± 1.0	1.5 ± 1.0	**<.001**
$HEMORR_2HAGES$ score, mean ± SD	1.9 ± 1.7	1.9 ± 1.7	2.6 ± 1.8	**<.001**
ORBIT-AF score, mean ± SD	1.7 ± 1.6	1.6 ± 1.6	2.5 ± 1.7	**<.001**
CHA_2DS_2-VASc score, mean ± SD	2.4 ± 1.5	2.4 ± 1.5	3.0 ± 1.7	**<.001**

(Continues)

TABLE 1 (Continued)

Characteristics	Overall	Major bleeding		*P* value
		No	Yes	
Smoking habit, *n* (%)				**.003**
Never	1052 (44.5)	1022 (45.1)	30 (30.9)	
Current	67 (2.8)	67 (3.0)	0 (0.0)	
Former	1222 (51.7)	1155 (51.0)	67 (69.1)	
Unknown	23 (1.0)	23 (1.0)	0 (0.0)	
Body mass index (kg/m^2)	29.9 (8.4)	29.9 (8.4)	29.5 (8.2)	.618
Systolic blood pressure (mmHg)	127.5 (21.5)	127.5 (22.0)	127.5 (17.5)	.905
Diastolic blood pressure (mmHg)	71.0 (11.5)	71.0 (11.5)	67.5 (11.0)	**<.001**
Glomerular filtration rate (mL/min)	96.5 (50.4)	92.6 (49.2)	75.5 (42.5)	**<.001**
Platelet count (10^9/L)	209.8 (74.0)	209.5 (73.0)	216.0 (95.0)	.193
Hemoglobin (g/dL)	13.5 (2.5)	13.5 (2.4)	12.4 (3.4)	**<.001**
Follow-up time (d), mean ± SD	828.2 ± 739.8	830.0 ± 740.0	787.8 ± 735.6	.195

Bolded *P* values are <.05 and considered statistically significants.

The italicized *P* value for categorical variables represents the output of the Fisher's exact test, while the nonitalicized *P* values correspond to the output of χ^2 test.

ABH, Age, history of Bleeding, and non-bleeding related Hospitalisation; ATRIA, Anticoagulation and Risk Factors in Atrial Fibrillation; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age ≥75 (doubled), Diabetes, Stroke (doubled)-Vascular disease, Age 65–74, Sex category; COPD, chronic obstructive pulmonary disease; CYP, cytochrome P450; DOAC, direct oral anticoagulants; HAS-BLED, Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly (>65 years), Drugs/alcohol concomitantly; $HEMORR_2HAGES$, Hepatic or Renal Disease, Ethanol Abuse, Malignancy, Older Age, Reduced Platelet Count or Function, Re-Bleeding, Hypertension, Anemia, Genetic Factors, Excessive Fall Risk and Stroke; ORBIT-AF, Outcomes Registry for Better Informed Treatment of Atrial Fibrillation; p-gp, p-glycoprotein.

3.2 | Primary comparison metric: time-dependent ROC analysis

Temporal fluctuations in AUROC for each risk score as a continuous variable were plotted against follow-up time, and findings are depicted in Figure 1 and Supplementary Figure S2.

Within the initial 100 days of follow-up, there was a notable decline in predictive performance across all risk scores during the initial 30 days, which was subsequently followed by a recovery and consistent maintenance of predictive accuracy (Figure 1A). During the entire follow-up period, ATRIA and ORBIT-AF scores consistently displayed superior predictive performance (Figure 1B). They exhibited notably narrower 95% CIs, with their lower limits remaining below 0.5 for a brief duration of the follow-up (approximately at the 2500th day). In contrast, the remaining 4 scores showcased broader CIs and/or AUROC values that dipped below 0.5 (Supplementary Figure S2).

To comprehensively assess the predictive performance of the 6 risk scores, an exploratory analysis was conducted using continuous risk scores to compute and compare the AUROC values at 3 specific time points: 100, 1100, and 2100 days (as illustrated in Figure 2). Of the total 97 bleeding events, 22 (22.7%) occurred within the first 100 days; 48 (49.5%) between 101 and 1100 days; 20 (20.6%) between 1101 and 2100 days; and 7 (7.2%) after 2101 days.

On the 100th day of the follow-up period, the ORBIT-AF score exhibited a significantly higher predictive performance in comparison to both the ABH and $HEMORR_2HAGES$ scores (Figure 2A). Extending the assessment to the 1100th day mark, the ORBIT-AF score consistently demonstrated significantly higher predictive performance in predicting DOAC-related major bleeding if compared to all the risk scores, except for ATRIA. In contrast, ATRIA showed significantly higher performance in predicting DOAC-related major bleeding if compared to all the risk scores, except for ORBIT-AF and CHA_2DS_2-VASc scores (Figure 2B). However, as the follow-up period extended to 2100 days, no statistically significant differences in the predictive performance of DOAC-related major bleeding among the 6 scores were observed (Figure 2C).

The performance of the 6 scores to discriminate the risk of major bleeding in individuals using rivaroxaban and apixaban was individually assessed and compared through a time-dependent exploratory analysis of AUROC at 3 distinct time points. Findings of this evaluation are illustrated in Supplementary Figure S3. On the 100th day of the follow-up period, the ATRIA score exhibited a significantly better performance in predicting major bleeding associated with apixaban when contrasted with rivaroxaban. A similar trend was observed with the ABH score. Conversely, no statistically significant differences were observed among the remaining scores (Supplementary Figure S3A). In assessments conducted at the 1100th and 2100th days of follow-up, all 6 risk scores showed no significant differences in predicting major bleeding risk from rivaroxaban compared to apixaban (comparative *P* values >.05, Supplementary Figures S3B, C).

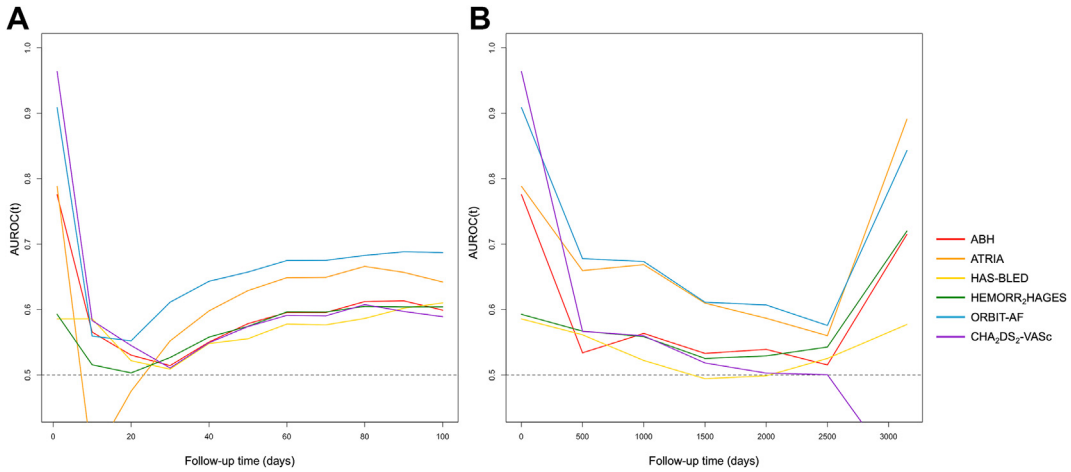


FIGURE 1 Temporal variation in area under the receiver operating characteristic curves (AUROCs) for 6 predictive risk scores (A) during the first 100 days and (B) over the entire follow-up period. ABH, Age, history of Bleeding, and non-bleeding related Hospitalisation; ATRIA, Anticoagulation and Risk Factors in Atrial Fibrillation; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age ≥75 (doubled), Diabetes, Stroke (doubled)-Vascular disease, Age 65–74, Sex category; HAS-BLED, Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly (>65 years), Drugs/alcohol concomitantly; $HEMORR_2HAGES$, Hepatic or Renal Disease, Ethanol Abuse, Malignancy, Older Age, Reduced Platelet Count or Function, Re-Bleeding, Hypertension, Anemia, Genetic Factors, Excessive Fall Risk and Stroke; ORBIT-AF, Outcomes Registry for Better Informed Treatment of Atrial Fibrillation.

3.3 | Other comparison metrics

3.3.1 | Incidence rates of major bleeding

A total of 97 cases of major bleeding were observed over an average follow-up period of 2.27 years, resulting in 1.81 major bleeds per 100 person-years. Details regarding the incidence rates of major bleeding per 100 person-years, along with their corresponding 95% CIs across risk categories, are comprehensively presented in Supplementary Tables S3 and S4.

The only statistically significant pairwise difference in major bleeding incidence rates between risk score categories was observed with the intermediate-risk patients (referred as I) of ATRIA and HAS-BLED scores. The ATRIA intermediate group exhibited a significantly higher major bleeding incidence rate at 2.55 major bleeds per 100 person-years compared to 1.49 major bleeds per 100 person-years for the HAS-BLED intermediate group (Supplementary Table S3).

Using these risk scores, the frequency distribution of patients and the corresponding incidence rates per 100 person-years were calculated across both risk categories and discrete scores. These comprehensive findings have been compiled and are presented in Supplementary Tables S4 and S5.

3.3.2 | Cumulative incidence of major bleeding and diagnostic efficiency measures

We evaluated major bleeding incidence across risk categories using cumulative incidence calculations. Diagnostic efficiency measures were assessed for high-risk major bleeding patients versus those at low-to-intermediate risk over the entire follow-up period. Findings are summarized in Table 2.

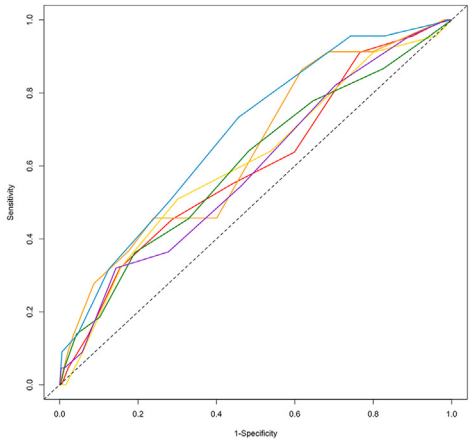
ATRIA, HAS-BLED, and ORBIT-AF were the scores that disrupted the incremental rise in major bleeding incidences as risk severity progressed. Negative and positive predictive values showed consistent values of 96.1% to 96.9% and 4.3% to 5.1%, respectively, across the 6 risk scores. However, notable variations were observed in sensitivity, specificity, Youden's index, and MCC. CHA_2DS_2-VASc score upheld the highest sensitivity (89.7%) in DOAC-related major bleeding screening, while ORBIT-AF followed by ABH scores exhibited the highest specificity values (71.2% and 71.1%, respectively). Since Youden's index represents the maximization of the equilibrium between sensitivity and specificity through equal weight, this metric has been used for determining the cutoff point in diagnostic tests [35]. Thus, we compared the Youden's index of our cohort's results with the pre-established high-risk cutoffs for each predictive score. Notably, only the ABH score demonstrated a Youden's index that matched the predefined high-risk cutoff of 7 (Table 2).

On the other hand, MCC stands out as a statistical metric that yields a high score exclusively when all 4 diagnostic efficiency measures exhibit heightened values [36]. Since MCC offers greater informativeness and reliability, we computed this metric for all 6 risk scores. The results varied across the 6 risk scores, with HAS-BLED and ORBIT-AF having the same and highest MCC values (MCC = 0.032; Table 2), followed by the ABH score (MCC = 0.031; Table 2).

3.3.3 | Calibration plots

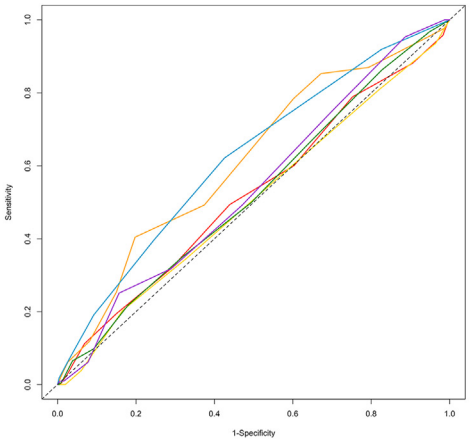
Calibration plots were employed to assess the alignment between the incidence rates of major bleeding events per 100 person-years

A



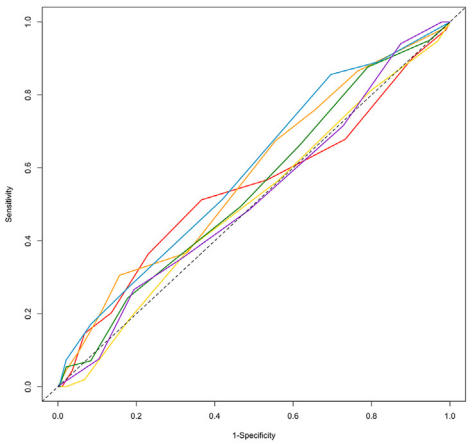
AUROC (95%CI)	ABH 0.60 (0.48-0.72)	ATRIA 0.64 (0.53-0.76)	HAS-BLED 0.61 (0.49-0.73)	HEMORR$_2$HAGES 0.60 (0.48-0.73)	ORBIT-AF 0.69 (0.59-0.79)	CHA$_2$DS$_2$-VASc 0.59 (0.47-0.71)
ABH 0.60 (0.48-0.72)	—	p = 0.243	p = 0.794	p = 0.894	**p = 0.013**	p = 0.823
ATRIA 0.64 (0.53-0.76)	p = 0.243	—	p = 0.603	p = 0.456	p = 0.241	p = 0.292
HAS-BLED 0.61 (0.49-0.73)	p = 0.794	p = 0.603	—	p = 0.845	p = 0.066	p = 0.714
HEMORR$_2$HAGES 0.60 (0.48-0.73)	p = 0.894	p = 0.456	p = 0.845	—	**p = 0.018**	p = 0.777
ORBIT-AF 0.69 (0.59-0.79)	**p = 0.013**	p = 0.241	p = 0.066	**p = 0.018**	—	p = 0.067
CHA$_2$DS$_2$-VASc 0.59 (0.47-0.71)	p = 0.823	p = 0.292	p = 0.714	p = 0.777	p = 0.067	—

B



AUROC (95%CI)	ABH 0.52 (0.44-0.60)	ATRIA 0.61 (0.54-0.69)	HAS-BLED 0.50 (0.43-0.58)	HEMORR$_2$HAGES 0.52 (0.45-0.60)	ORBIT-AF 0.62 (0.55-0.69)	CHA$_2$DS$_2$-VASc 0.54 (0.47-0.61)
ABH 0.52 (0.44-0.60)	—	**p = 0.004**	p = 0.428	p = 0.982	**p <0.001**	p = 0.588
ATRIA 0.61 (0.54-0.69)	**p = 0.004**	—	**p = 0.007**	**p = 0.003**	p = 0.734	p = 0.056
HAS-BLED 0.50 (0.43-0.58)	p = 0.428	**p = 0.007**	—	p = 0.386	**p <0.001**	p = 0.277
HEMORR$_2$HAGES 0.52 (0.45-0.60)	p = 0.982	**p = 0.003**	p = 0.386	—	**p <0.001**	p = 0.605
ORBIT-AF 0.62 (0.55-0.69)	**p <0.001**	p = 0.734	**p <0.001**	**p <0.001**	—	**p = 0.028**
CHA$_2$DS$_2$-VASc 0.54 (0.47-0.61)	p = 0.588	p = 0.056	p = 0.277	p = 0.605	**p = 0.028**	—

C



AUROC (95%CI)	ABH 0.54 (0.44-0.64)	ATRIA 0.58 (0.50-0.66)	HAS-BLED 0.51 (0.42-0.59)	HEMORR$_2$HAGES 0.54 (0.46-0.62)	ORBIT-AF 0.59 (0.51-0.67)	CHA$_2$DS$_2$-VASc 0.52 (0.43-0.60)
ABH 0.54 (0.44-0.64)	—	p = 0.481	p = 0.165	p = 0.969	p = 0.291	p = 0.438
ATRIA 0.58 (0.50-0.66)	p = 0.481	—	p = 0.159	p = 0.346	p = 0.570	p = 0.260
HAS-BLED 0.51 (0.42-0.59)	p = 0.165	p = 0.159	—	p = 0.170	p = 0.062	p = 0.730
HEMORR$_2$HAGES 0.54 (0.46-0.62)	p = 0.969	p = 0.346	p = 0.170	—	p = 0.157	p = 0.488
ORBIT-AF 0.59 (0.51-0.67)	p = 0.291	p = 0.570	p = 0.062	p = 0.157	—	p = 0.145
CHA$_2$DS$_2$-VASc 0.52 (0.43-0.60)	p = 0.438	p = 0.260	p = 0.730	p = 0.488	p = 0.145	—

FIGURE 2 Comparative time-dependent areas under the receiver operating characteristic curves (AUROCs) of the 6 scores in discriminating the risk of direct oral anticoagulant–related major bleeding at (A) 100, (B) 1100, and (C) 2100 days of follow-up. ABH, Age, history of Bleeding, and non-bleeding related Hospitalisation; ATRIA, Anticoagulation and Risk Factors in Atrial Fibrillation; CHA$_2$DS$_2$-VASc, Congestive heart failure, Hypertension, Age ≥75 (doubled), Diabetes, Stroke (doubled)-Vascular disease, Age 65–74, Sex category; HAS-BLED, Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly (>65 years), Drugs/alcohol concomitantly; HEMORR$_2$HAGES, Hepatic or Renal Disease, Ethanol Abuse, Malignancy, Older Age, Reduced Platelet Count or Function, Re-Bleeding, Hypertension, Anemia, Genetic Factors, Excessive Fall Risk and Stroke; ORBIT-AF, Outcomes Registry for Better Informed Treatment of Atrial Fibrillation.

reported in the derivation cohorts and the actual incidence rate observed within our cohort. The results are shown in Supplementary Figure S4.

Based on our findings, it was observed that the incidence rates of major bleeding per 100 person-years in the low-risk category within our cohort were slightly underestimated when compared to the derivation cohort. This pattern is consistent across all risk scores except for ATRIA and HAS-BLED, where there is a slight over-estimation of the incidence rates. Shifting to the intermediate risk category, our cohort displayed a significant underestimation of the incidence rates of major bleeding per 100 person-years for ABH, HEMORR$_2$HAGES, and ORBIT-AF scores. However, ATRIA,

TABLE 2 Distribution of cumulative incidence across risk categories and diagnostic efficiency measures of risk scores for major bleeding from direct oral anticoagulants.

	Cumulative incidence of major bleeds (number of events/total of individuals)			Diagnostic efficiency measures (low + intermediate vs high risk)					
Prediction score	Low	Intermediate	High	Sensitivity	Specificity	NPV	PPV	Youden's index	Matthew's correlation coefficient
ABH	8/224 (3.6%)	54/1449 (3.7%)	35/691 (5.1%)	36.1%	71.1%	96.3%	5.1%	0.07	0.031
ATRIA	23/872 (2.6%)	30/528 (5.7%)	44/964 (4.6%)	45.4%	59.4%	96.2%	4.6%	0.05	0.019
HAS-BLED	5/90 (5.6%)	33/1017 (3.2%)	59/1257 (4.7%)	60.8%	47.2%	96.6%	4.7%	0.08	0.032
$HEMORR_2HAGES$	13/408 (3.2%)	34/807 (4.2%)	50/1149 (4.4%)	51.6%	51.5%	96.1%	4.4%	0.03	0.012
ORBIT-AF	38/1272 (3.0%)	24/405 (5.9%)	35/687 (5.1%)	36.1%	71.2%	96.3%	5.1%	0.07	0.032
CHA_2DS_2-VASc	1/39 (2.6%)	9/283 (3.2%)	87/2042 (4.3%)	89.7%	13.8%	96.9%	4.3%	0.04	0.020

Classification of risk score categories: ABH: low risk, 0 to 2; intermediate risk, 3 to 6; high risk, ≥7; ATRIA: low risk, 0 to 3; intermediate risk, 4; high risk, ≥5; HAS-BLED: low risk, 0; intermediate risk, 1 to 2; high risk, ≥3; $HEMORR_2HAGES$: low risk, 0 to 1; intermediate risk, 2 to 3; high risk, ≥4; ORBIT-AF: low risk, 0 to 2; intermediate risk, 3; high risk, ≥4; and CHA_2DS_2-VASc: low risk, 0 for men and 0 to 1 for women; intermediate risk, 1 for men and 2 for women; high risk, ≥2 for men and ≥3 for women.

ABH, Age, history of Bleeding, and non-bleeding related Hospitalisation; ATRIA, Anticoagulation and Risk Factors in Atrial Fibrillation; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age ≥75 (doubled), Diabetes, Stroke (doubled)-Vascular disease, Age 65–74, Sex category; HAS-BLED, Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly (>65 years), Drugs/alcohol concomitantly; $HEMORR_2HAGES$, Hepatic or Renal Disease, Ethanol Abuse, Malignancy, Older Age, Reduced Platelet Count or Function, Re-Bleeding, Hypertension, Anemia, Genetic Factors, Excessive Fall Risk and Stroke; NPV, negative predictive value; ORBIT-AF, Outcomes Registry for Better Informed Treatment of Atrial Fibrillation; PPV, positive predictive value.

HAS-BLED, and CHA_2DS_2-VASc scores demonstrated well-calibrated incidence rates in this category. Within the high-risk category, our cohort consistently displayed a significant underestimation of the incidence rates. This contrasted sharply with the derivation cohort, as evidenced by the pronounced slope observed in all 6 risk scores (Supplementary Figure S4).

3.3.4 | Decision curves

The clinical utility of the 6 risk assessment tools in predicting the probability of experiencing DOAC-related major bleeding over different time points (100, 1100, and 2100 days) was evaluated through decision curve analysis, depicted in Figure 3. Within our patient cohort, the ATRIA and ORBIT-AF scores exhibited the most pronounced net benefit in predicting the occurrence of major bleeding risk from DOAC at day 100 of observation, as illustrated by the orange and blue trends in Figure 3A. At the 1100^{th} day mark, the ATRIA score demonstrated superior net benefit alone in predicting the probability of experiencing DOAC-related major bleeding events (Figure 3B). However, when considering the 2100-day horizon, none of the risk scores exhibited discernible superiority in net benefit in predicting the probability of major bleeding events from DOACs (Figure 3C).

4 | DISCUSSION

To the best of our knowledge, this study is the first to evaluate the performance of the ABH score in predicting major bleeding risk within a real-world cohort of patients with AF undergoing DOAC treatment. The purpose of this research was to assess the predictive performance of the DOAC-specific ABH score compared with 5 other DOAC–unspecific risk scores in predicting major bleeding risk among patients with AF undergoing treatment with rivaroxaban or apixaban within real-world settings. We employed the time-dependent ROC analysis as the primary metric to compare the performance of the 6 continuous risk scores in predicting DOAC-related major bleeding risk. While some scores consistently fell below the acceptable threshold (AUROC <0.7) over the entire observation period, our findings underscored the sustained predictive performance of both ATRIA and ORBIT-AF scores. In contrast to the remaining metrics, these 2 risk scores exhibited a persistent AUROC surpassing 0.6 over 3000 days of monitoring duration, occasionally even exceeding 0.7 at specific intervals. The 95% CI findings highlight that both the ATRIA and ORBIT-AF scores exhibited the narrowest intervals, consistently mirroring the AUROC trend over time with minimal exaggerated fluctuations.

During the initial 100-day follow-up period, ORBIT-AF exhibited significantly superior performance in predicting major bleeding risk from DOACs compared to ABH and $HEMORR_2HAGES$ scores. Extending the evaluation to 1100 days, ORBIT-AF maintained its statistical significance over all other scores, except for ATRIA. Notably, ATRIA also demonstrated significantly higher predictive performance than all scores except for ORBIT-AF and CHA_2DS_2-VASc scores. However, as the observation extended further into the long term, the differences in performance between ATRIA and ORBIT-AF ceased to hold statistical significance compared to all other scores. Hence, our findings suggest that both ORBIT-AF and ATRIA scores

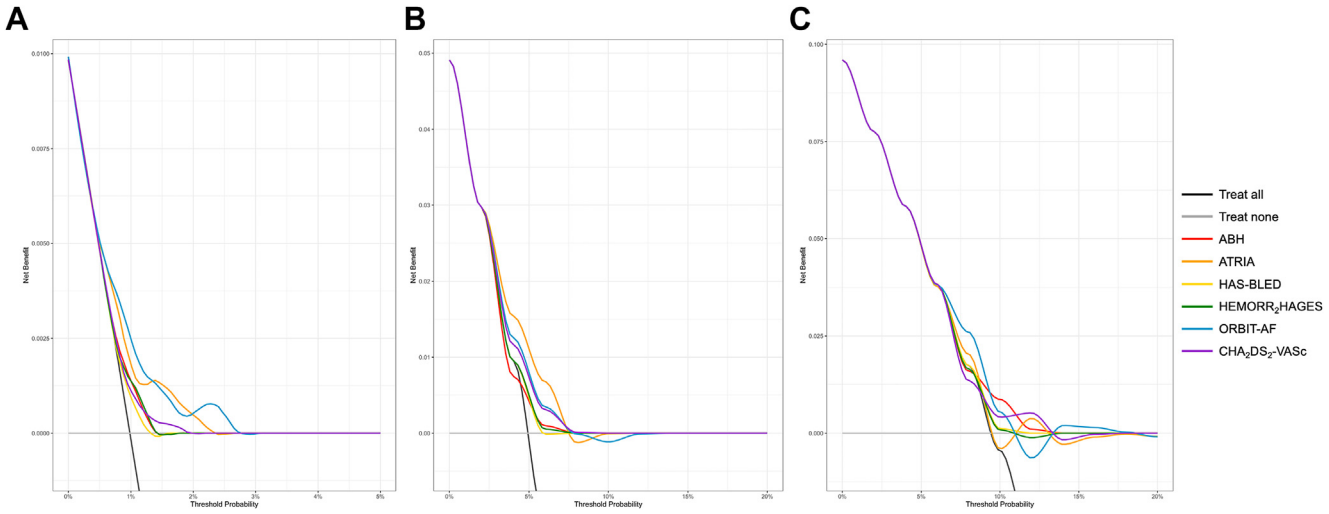


FIGURE 3 Decision curve analysis plots for the probability of experiencing direct oral anticoagulant–related major bleeding events at (A, B) 100, (C) 1100, and (D) 2100 days of follow-up according to each of the 6 risk predicting scores. The axes scale in Figure 3A is condensed compared to those in Figure 3B–D to zoom in on score performance at 100 days. ABH, Age, history of Bleeding, and non-bleeding related Hospitalisation; ATRIA, Anticoagulation and Risk Factors in Atrial Fibrillation; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age ≥75 (doubled), Diabetes, Stroke (doubled)-Vascular disease, Age 65–74, Sex category; HAS-BLED, Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly (>65 years), Drugs/alcohol concomitantly; $HEMORR_2HAGES$, Hepatic or Renal Disease, Ethanol Abuse, Malignancy, Older Age, Reduced Platelet Count or Function, Re-Bleeding, Hypertension, Anemia, Genetic Factors, Excessive Fall Risk and Stroke; ORBIT-AF, Outcomes Registry for Better Informed Treatment of Atrial Fibrillation.

exhibit the best performance in predicting the risk of major bleeding throughout the first 1100 days of DOAC treatment.

Using the Cox proportional hazards regression model, we estimated the individual probabilities of experiencing major bleeding from DOACs. These estimates made it possible to assess the net benefits of each score through decision curve analysis. In the initial 100-day follow-up period, the analysis confirmed the AUROC findings, demonstrating that ATRIA had a net benefit in predicting DOAC-related major bleeding, paralleled the ORBIT-AF score. When extending the follow-up to 1100 days, the ATRIA score consistently exhibited superior net benefit in predicting the probability of major bleeds, followed by the ORBIT-AF and CHA_2DS_2-VASc scores. If analyzed together, the outcomes from time-dependent AUROC and decision curves consistently indicate superior performance for predicting major bleeding using the ORBIT-AF and ATRIA scores during the initial 100 days of exposure to DOACs and ATRIA scores for DOAC exposures ranging between 100 and 1100 days.

Although the risk scores are a combination of clinical variables, their assigned weights vary greatly. This array of clinical variables and weights has the potential to elucidate the reasons behind the enhanced predictive performance of ATRIA and ORBIT-AF scores in assessing bleeding risk within our study cohort. Both ATRIA and ORBIT-AF consist of 5 clinical variables each, of which age ≥75 years, kidney disease, previous bleeding, and anemia are shared by them. However, they diverge in the assessment of arterial hypertension, use of antiplatelet drugs, and assigned weights of variables [16,19]. Out of the 6 tested scores, 3 ascribe 2 points to individuals aged ≥75 years. Notably, these specific scores—CHA_2DS_2-VASc, ORBIT-AF, and ATRIA—have showcased superior performance in predicting DOAC-related major bleeding within our patient cohort. Since most of our patient cohort consists of the older adults, with a median age of 68.3 years and a substantial subset (25.1%) aged ≥75 years, it is conceivable that attributing a higher weight of 2 points to this age group might reasonably account for the distinct improvement in the predictive performance of these scores observed in our study.

Additionally, we posit that the significant improvement in predictive performance observed both in ATRIA and ORBIT-AF could be attributed to their distinct approaches to evaluating anemia. Although ABH, ATRIA, $HEMORR_2HAGES$, and ORBIT-AF all consider anemia as a clinical predictor of bleeding events [16,18,19,21], it is noteworthy that ATRIA and ORBIT-AF uniquely allocate more substantial weights of 3 and 2 points, respectively, to this parameter [16,19]. Notably, 68.2% of our patient cohort exhibited evidence of some degree of anemia at baseline, substantiated by at least 1 laboratory or International Classification of Diseases code record, as defined by the scoring criteria. Hence, we postulate that this heightened emphasis placed on this clinical variable might have played a pivotal role in bolstering the predictive performance of these 2 risk scores, particularly when they are applied to a patient population where anemia holds a high prevalence, as in our study.

Weighting previous bleeding events as 2 points may be an additional factor that likely enhanced the predictive performance of ORBIT-AF in our cohort [19], whose prevalence was 20.8% at baseline. Alternatively, in the context of ATRIA, we conjecture that the 3-point weight attributed to renal dysfunction, specifically when the glomerular filtration rate is <30 mL/min or in cases requiring

dialysis [16], would also have a substantial impact on the predictive performance of this score. However, due to our study's exclusion criteria encompassing either a creatinine clearance of <15 mL/min or the need for renal replacement therapy, only 1% of our patient cohort exhibited glomerular filtration rates ranging from 15 to 30 mL/min. Consequently, it is reasonable to postulate that, although this variable may contribute to a better predictive performance of the ATRIA score, its influence supposedly remained marginal in our investigation.

Several primary studies have attempted to assess the performance of various scoring systems in predicting the risk of major bleeding in patients with AF undergoing anticoagulation treatment. However, a consensus on this matter has not yet been reached. To tackle this issue, we conducted a comparative analysis of our findings with more robust evidence derived from meta-analytical approaches despite acknowledging the limitations of such comparisons. Among the 6 existing meta-analyses, all have taken HAS-BLED as the standard score and assessed its predictive performance in comparison to the others. In 4 of these meta-analyses, the ATRIA and ORBIT-AF scores were tested against the HAS-BLED score [37–42]. The findings of these meta-analytic studies diverge from our results as they concluded that there is no significant difference in the predictive performance of the ATRIA and ORBIT-AF scores when compared to the HAS-BLED score [38–40,42]. It is imperative to note, however, that the patient cohorts in these studies differ from ours. For instance, Zhu et al. [42] compiled primary studies involving patients solely on VKAs. Meanwhile, Gao et al. [39], Wang et al. [38], and Liu et al. [40] encompassed patients on both DOACs and VKAs without conducting a subanalysis to discern the individual predictive performance of scores for each drug. Although preliminary evidence suggests a similar performance of predictive risk scores in patients experiencing bleeding events, whether from VKAs or DOACs [39,40], further comprehensive investigations are necessary to appropriately address any potential limitations that may arise when comparing these findings against ours.

Regarding the calibration of risk category scores, our findings exhibited partial discrepancies in comparison to the outcomes presented by Zeng et al. [37]. Although our approach computed annualized incidence rates, the outcomes of this comprehensive meta-analysis study were originally presented as odds ratios. To make cross-study comparisons possible, we calculated the odds ratios of risk categories for the ATRIA, HEMORR$_2$HAGES, and ORBIT-AF scores against the HAS-BLED score (data not shown). Our discoveries align with those by Zeng et al. [37] within the high-risk categories of the 3 scores, where the *P* values for odds ratios were not statistically significant. Furthermore, our findings were also aligned with theirs in the intermediate-risk category of ORBIT-AF, revealing odds ratio in favor of a significantly higher risk of major bleeding in the ORBIT-AF category than in the HAS-BLED category (our result: odds ratio, 0.53; 95% CI, 0.31-0.91; $P = .022$). There was no consistency of odds ratios across our findings and those of Zeng et al. [37] when comparing the low-risk categories of the 3 scores and the intermediate-risk categories for HEMORR$_2$HAGES and ATRIA scores.

To enable cross-comparison of our results with those of the sixth existing meta-analysis, we performed additional calculations of predictive sensitivity and specificity following the methodology described by Chang et al. [41] (data not shown). Although our numerical values differed from theirs, the pattern observed in our clustered ranking plots almost entirely reflected the results reported by Chang et al. [41]. When assessing predictive sensitivity, our results were perfectly consistent with theirs, revealing ORBIT-AF, ATRIA, HEMORR$_2$HAGES, and HAS-BLED scores in ascending order. However, when assessing predictive specificity, our results ranked ORBIT-AF, ATRIA, and HAS-BLED equally, followed by HEMORR$_2$HAGES scores in decreasing order, which contradicts the assertion by Chang et al. [41] that HAS-BLED should have a lower predictive specificity than ATRIA.

Our study yields noteworthy implications for clinical practice. During the initial 100 days of DOAC treatment, ATRIA, HAS-BLED, ORBIT-AF, and CHA$_2$DS$_2$-VASc scores demonstrated comparable performance in predicting major bleeding events in patients with AF. However, we strongly advocate that clinicians prioritize the use of either ORBIT-AF or ATRIA scores during this period since these 2 scoring systems stood out with the highest net benefit on the decision curves. In the medium term, spanning from 100 to 1100 days of DOAC treatment, while both ORBIT-AF and ATRIA scores exhibited similar predictive performance for major bleeding risk in patients with AF, our recommendation leans toward the preferential use of the ATRIA score due to its superior net benefit demonstrated on the decision curve within this timeframe. The 6 scores exhibited poor predictive performance over a span of 1100 to 2100 days of follow-up, suggesting a likely inadequacy in capturing the risk of bleeding events associated with long-term anticoagulation therapy.

While our study has several strengths, such as a cohort of patients solely on DOACs, confirmation of major bleeding cases through chart reviews by a trained investigator rather than International Classification of Diseases codes, and a time-dependent approach to testing the predictive performance of these scores, certain limitations should be considered. First, this was an observational study with fewer than 100 total bleeding events, and the findings are limited to a single healthcare system that primarily serves southeastern Michigan. The study's retrospective nature may have introduced a temporality bias, which was mitigated by ascertaining risk scores using clinical data up to 1 year before the first prescription of DOACs and confirming that bleeding cases occurred within the start and stop dates of DOACs through medical chart reviews. Furthermore, since all data were obtained from the electronic health record, data on certain variables outside the scope of DOACs, such as labile international normalized ratio and genetic variants related to VKAs, were not collected, which may have affected the accuracy of risk scores. Although we followed standard protocols for data collection, the clinical care routine lacked standardization, and, as with any observational study, unmeasured residual confusion may have been included. Also, we did not have access to pharmacy records to ascertain adherence to DOAC therapy. Finally, our study focused solely on evaluating the performance of 6 bleeding risk prediction risk scores, acknowledging the existence of additional scores that were not incorporated into our analysis.

Although our study made valuable contributions, we acknowledge the need for further research to substantiate our findings within population-based cohorts. Additionally, it is imperative to assess the clinical implications of the ATRIA and ORBIT-AF scores in routine practice, aiming to improve the prediction of major bleeding risk from DOACs. Looking ahead, research endeavors should explore the performance of alternative bleeding risk prediction scores or the development of innovative ones, ultimately enhancing the precision of major bleeding risk assessment in patients with AF undergoing treatment with DOACs.

In conclusion, our study reveals the consistent superiority of ORBIT-AF and ATRIA over other scoring systems in predicting DOAC-related major bleeding in patients with AF within the first 100 days of treatment. Notably, ATRIA maintained this high predictive performance within 100 and 1100 days of follow-up, surpassing the ABH risk score. These results emphasize the practical value of ATRIA and ORBIT-AF scores as essential tools for clinicians in real-world settings, aiding in the estimation of major bleeding risk in patients with AF receiving DOAC therapy.

FUNDING

A.M.C.-S. was funded by the Michigan Institute for Clinical & Health Research (U074551) and the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (1F32HL162231-01A1). G.D.B. is funded by Boston Scientific, NHLBI, Agency for Healthcare Research and Quality, and Blue Cross Blue Shield of Michigan. J.A.L. is funded by NHLBI of the National Institutes of Health (K08 HL146990).

AUTHOR CONTRIBUTIONS

Conceptualization: A.M.C.-S., J.A.L., G.D.B., and N.A.L.. Design: A.M.C.-S., J.A.L., G.D.B., and N.A.L.. Data analysis: A.M.C.-S.. Writing: A.M.C.-S.. Critical review: M.P.D., G.D.B., N.A.L., and J.A.L.

RELATIONSHIP DISCLOSURE

M.P.D. has received research funding from the Agency for Healthcare Research and Quality; National Heart, Lung, and Blood Institute (NHLBI); National Institute of Aging (NIA); and the American Heart Association in the past 2 years. G.D.B. reports a grant from or a contract with Boston Scientific as the site principal investigator for the Higher-Risk Pulmonary Embolism Thrombolysis (HI-PEITHO) Trial; is a consultant for Pfizer, Bristol-Myers Squib, Janssen, Boston Scientific, Bayer, AstraZeneca, Sanofi, Anthos, and Abbott Vascular; and is on the Event Adjudication Committee for Translational Sciences and on the Board of Directors for Anticoagulation Forum. J.A.L. is funded by the National Heart, Lung, and Blood Institute of the National Institutes of Health (K08 HL146990). All other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY

The data will not be made publicly available to ensure patient privacy and confidentiality of protected health information.

TWITTER

Alessandra M. Campos-Staffico @LekaStaffico
Geoffrey D. Barnes @GBarnesMD
Jasmine A. Luzum @JasmineLuzum

REFERENCES

[1] Alkhouli M, Alqahtani F, Aljohani S, Alvi M, Holmes DR. Burden of atrial fibrillation–associated ischemic stroke in the United States. *JACC Clin Electrophysiol*. 2018;4:618–25.
[2] Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. *Ann Intern Med*. 2007;146:857–67.
[3] January CT, Wann LS, Calkins H, Chen LY, Cigarroa JE, Cleveland JC, et al. 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines and the Heart Rhythm Society in collaboration with the Society of Thoracic Surgeons. *Circulation*. 2019;140:e125–51.
[4] Van Ganse E, Danchin N, Mahé I, Hanon O, Jacoud F, Nolin M, et al. Comparative safety and effectiveness of oral anticoagulants in nonvalvular atrial fibrillation: the NAXOS study. *Stroke*. 2020;51:2066–75.
[5] Munir MB, Hlavacek P, Keshishian A, Guo JD, Mallampati R, Ferri M, et al. Oral anticoagulant underutilization among elderly patients with atrial fibrillation: insights from the United States Medicare database. *J Interv Card Electrophysiol*. 2023;66:771–82.
[6] Shoeb M, Fang MC. Assessing bleeding risk in patients taking anticoagulants. *J Thromb Thrombolysis*. 2013;35:312–9.
[7] Shehab N, Lovegrove MC, Geller AI, Rose KO, Weidle NJ, Budnitz DS. US emergency department visits for outpatient adverse drug events, 2013-2014. *JAMA*. 2016;316:2115–25.
[8] Budnitz DS, Shehab N, Lovegrove MC, Geller AI, Lind JN, Pollock DA. US emergency department visits attributed to medication harms, 2017-2019. *JAMA*. 2021;326:1299–309.
[9] Troy A, Anderson TS. National trends in use of and spending on oral anticoagulants among US medicare beneficiaries from 2011 to 2019. *JAMA Health Forum*. 2021;2:e211693. https://doi.org/10.1001/jamahealthforum.2021.1693
[10] Granger CB, Alexander JH, McMurray JJV, Lopes RD, Hylek EM, Hanna M, et al. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2011;365:981–92.
[11] Giugliano RP, Ruff CT, Braunwald E, Murphy SA, Wiviott SD, Halperin JL, et al. Edoxaban versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2013;369:2093–104.
[12] Halperin JL, Hankey GJ, Wojdyla DM, Piccini JP, Lokhnygina Y, Patel MR, et al. Efficacy and safety of rivaroxaban compared with warfarin among elderly patients with nonvalvular atrial fibrillation in the rivaroxaban once daily, oral, direct factor Xa inhibition compared with vitamin K antagonism for prevention of stroke and embolism trial in atrial fibrillation (ROCKET AF). *Circulation*. 2014;130:138–46.
[13] Geller AI, Shehab N, Lovegrove MC, Rose KO, Weidle NJ, Goring SK, et al. Emergency visits for oral anticoagulant bleeding. *J Gen Intern Med*. 2020;35:371–3.

[14] Kachroo S, Hamilton M, Liu X, Pan X, Brixner D, Marrouche N, et al. Oral anticoagulant discontinuation in patients with nonvalvular atrial fibrillation. *Am J Manag Care*. 2016;22:e1–8.

[15] Königsbrügge O, Simon A, Domanovits H, Pabinger I, Ay C. Thromboembolic events, bleeding, and drug discontinuation in patients with atrial fibrillation on anticoagulation: a prospective hospital-based registry. *BMC Cardiovasc Disord*. 2016;16:254. https://doi.org/10.1186/s12872-016-0438-5

[16] Fang MC, Go AS, Chang Y, Borowsky LH, Pomernacki NK, Udaltsova N, et al. A new risk scheme to predict warfarin-associated hemorrhage: the ATRIA (Anticoagulation and Risk Factors in Atrial Fibrillation) study. *J Am Coll Cardiol*. 2011;58:395–401.

[17] 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;138:1093–100.

[18] Gage BF, Yan Y, Milligan PE, Waterman AD, Culverhouse R, Rich MW, et al. Clinical classification schemes for predicting hemorrhage: results from the National Registry of Atrial Fibrillation (NRAF). *Am Heart J*. 2006;151:713–9.

[19] O'Brien EC, Simon DN, Thomas LE, Hylek EM, Gersh BJ, Ansell JE, et al. The ORBIT bleeding score: a simple bedside score to assess bleeding risk in atrial fibrillation. *Eur Heart J*. 2015;36:3258–64.

[20] Heestermans M, Poenou G, Hamzeh-Cognasse H, Cognasse F, Bertoletti L. Anticoagulants: a short history, their mechanism of action, pharmacology, and indications. *Cells*. 2022;11:3214. https://doi.org/10.3390/cells11203214

[21] Rutherford O-CW, Jonasson C, Ghanima W, Holst R, Halvorsen S. New score for assessing bleeding risk in patients with atrial fibrillation treated with NOACs. *Open Heart*. 2018;5:e000931. https://doi.org/10.1136/openhrt-2018-000931

[22] Campos-Staffico AM, Dorsch MP, Barnes GD, Zhu H-J, Limdi NA, Luzum JA. Eight pharmacokinetic genetic variants are not associated with the risk of bleeding from direct oral anticoagulants in nonvalvular atrial fibrillation patients. *Front Pharmacol*. 2022;13:1007113. https://doi.org/10.3389/fphar.2022.1007113

[23] Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron*. 1976;16:31–41.

[24] University of Michigan. Michigan Genomics Initiative; 2022. https://precisionhealth.umich.edu/our-research/michigangenomics/. [accessed May 18, 2022].

[25] Hanigan S, Das J, Pogue K, Barnes GD, Dorsch MP. The real world use of combined P-glycoprotein and moderate CYP3A4 inhibitors with rivaroxaban or apixaban increases bleeding. *J Thromb Thrombolysis*. 2020;49:636–43.

[26] Hanauer DA, Mei Q, Law J, Khanna R, Zheng K. Supporting information retrieval from electronic health records: a report of University of Michigan's nine-year experience in developing and using the Electronic Medical Record Search Engine (EMERSE). *J Biomed Inform*. 2015;55:290–300.

[27] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42:377–81.

[28] Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208. https://doi.org/10.1016/j.jbi.2019.103208

[29] IBM Micromedex Solutions; 2023. https://www.micromedexsolutions.com/home/dispatch/. [accessed March 13, 2023].

[30] Schulman S, Angerås U, Bergqvist D, Eriksson B, Lassen MR, Fisher W. Subcommittee on Control of Anticoagulation of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in surgical patients. *J Thromb Haemost*. 2010;8:202–4.

[31] Yao X, Gersh BJ, Sangaralingham LR, Kent DM, Shah ND, Abraham NS, et al. Comparison of the CHA_2DS_2-VASc, $CHADS_2$, HAS-BLED, ORBIT, and ATRIA risk scores in predicting non–vitamin K antagonist oral anticoagulants-associated bleeding in patients with atrial fibrillation. *Am J Cardiol*. 2017;120:1549–56.

[32] Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983;148:839–43.

[33] Hosmer DW Jr, Lemeshow S, Sturdivant RX. Assessing the fit of the model. Applied logistic regression. John Wiley & Sons; 2013:177.

[34] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26:565–74.

[35] Fluss R, Faraggi D, Reiser B. Estimation of the Youden index and its associated cutoff point. *Biom J*. 2005;47:458–72.

[36] Chicco D, Jurman G. An invitation to greater use of Matthews correlation coefficient in robotics and artificial intelligence. *Front Robot AI*. 2022;9:876814. https://doi.org/10.3389/frobt.2022.876814

[37] Zeng J, Yu P, Cui W, Wang X, Ma J, Zeng C. Comparison of HAS-BLED with other risk models for predicting the bleeding risk in anticoagulated patients with atrial fibrillation: a PRISMA-compliant article. *Medicine*. 2020;99:e20782. https://doi.org/10.1097/MD.0000000000020782

[38] Wang C, Yu Y, Zhu W, Yu J, Lip GYH, Hong K. Comparing the ORBIT and HAS-BLED bleeding risk scores in anticoagulated atrial fibrillation patients: a systematic review and meta-analysis. *Oncotarget*. 2017;8:109703–11.

[39] Gao X, Cai X, Yang Y, Zhou Y, Zhu W. Diagnostic accuracy of the HAS-BLED bleeding score in VKA- or DOAC-treated patients with atrial fibrillation: a systematic review and meta-analysis. *Front Cardiovasc Med*. 2021;8:757087. https://doi.org/10.3389/fcvm.2021.757087

[40] Liu X, Wang S, He W, Guo L. HAS-BLED vs. ORBIT scores in anticoagulated patients with atrial fibrillation: a systematic review and meta-analysis. *Front Cardiovasc Med*. 2022;9:1042763. https://doi.org/10.3389/fcvm.2022.1042763

[41] Chang G, Xie Q, Ma L, Hu K, Zhang Z, Mu G, et al. Accuracy of HAS-BLED and other bleeding risk assessment tools in predicting major bleeding events in atrial fibrillation: a network meta-analysis. *J Thromb Haemost*. 2020;18:791–801.

[42] Zhu W, He W, Guo L, Wang X, Hong K. The HAS-BLED score for predicting major bleeding risk in anticoagulated patients with atrial fibrillation: a systematic review and meta-analysis. *Clin Cardiol*. 2015;38:555–61.

SUPPLEMENTARY MATERIAL

The online version contains supplementary material available at https://doi.org/10.1016/j.rpth.2023.102285.