Predicting Thromboembolic and Bleeding Event Risk in Patients with Non-Valvular Atrial Fibrillation: A Systematic Review

Ethan D. Borre¹, Adam Goode¹,² Giselle Raitz¹, Bimal Shah¹,³ Angela Lowenstern⁴,⁵ Ranee Chatterjee¹, Lauren Sharan¹, Nancy M. Allen LaPointe¹,⁶ J. Kelly Davis⁸ Kathryn Lallinger¹,⁴,⁹ Robyn Schmidt¹,⁴,⁹ Andrzej Kosinski¹⁰ Sana M. Al-Khatib⁴,⁵ Gillian D. Sanders¹,⁴,⁸,⁹

¹Department of Medicine, Duke University School of Medicine, Durham, North Carolina, United States
²Department of Orthopedic Surgery, Duke University School of Medicine, Durham, North Carolina, United States
³Livongo, Mountain View, California, United States
⁴Duke Clinical Research Institute, Duke University School of Medicine, Durham, North Carolina, United States
⁵Division of Cardiology, Duke University School of Medicine, Durham, North Carolina
⁶Premier Inc., Charlotte, North Carolina, United States
⁷Department of Medicine, University of Colorado, Aurora, Colorado, United States
⁸Duke-Margolis Center for Health Policy, Duke University, Durham, North Carolina, United States
⁹Evidence-Based Practice Center, Duke Clinical Research Institute, Duke University School of Medicine, Durham, North Carolina, United States
¹⁰Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina, United States

Address for correspondence: Gillian D. Sanders, PhD, Duke Clinical Research Institute, Duke Box 3485, 7020 North Pavilion Building, Durham, NC 27710, United States (e-mail: gillian.sanders@duke.edu).

Keywords
► non-valvular atrial fibrillation
► stroke risk
► bleeding risk

Abstract

Background  Atrial fibrillation (AF) is a common cardiac arrhythmia that increases the risk of stroke. Medical therapy for decreasing stroke risk involves anticoagulation, which may increase bleeding risk for certain patients. In determining the optimal therapy for stroke prevention for patients with AF, clinicians use tools with various clinical, imaging and patient characteristics to weigh stroke risk against therapy-associated bleeding risk.

Aim  This article reviews published literature and summarizes available risk stratification tools for stroke and bleeding prediction in patients with AF.

Methods We searched for English-language studies in PubMed, Embase and the Cochrane Database of Systematic Reviews published between 1 January 2000 and 14 February 2018. Two reviewers screened citations for studies that examined tools for predicting thromboembolic and bleeding risks in patients with AF. Data regarding study design, patient characteristics, interventions, outcomes, quality, and applicability were extracted.
Results  Sixty-one studies were relevant to predicting thromboembolic risk and 38 to predicting bleeding risk. Data suggest that CHADS<sub>2</sub>, CHA<sub>2</sub>DS<sub>2</sub>-VASc and the age, biomarkers, and clinical history (ABC) risk scores have the best evidence for predicting thromboembolic risk (moderate strength of evidence for limited prediction ability of each score) and that HAS-BLED has the best evidence for predicting bleeding risk (moderate strength of evidence).

Limitations  Studies were heterogeneous in methodology and populations of interest, setting, interventions and outcomes analysed.

Conclusion  CHADS<sub>2</sub>, CHA<sub>2</sub>DS<sub>2</sub>-VASc and ABC scores have the best prediction for stroke events, and HAS-BLED provides the best prediction for bleeding risk. Future studies should define the role of imaging tools and biomarkers in enhancing the accuracy of risk prediction tools.

Primary Funding Source  Patient-Centered Outcomes Research Institute (PROSPERO #CRD42017069999)

**Introduction**

Atrial fibrillation (AF) is the most common cardiac arrhythmia seen in clinical practice, occurring in up to 6.1 million people in the United States and accounting for approximately one-third of hospitalizations for cardiac rhythm disturbances. Further, AF is associated with significant morbidity and mortality, including increased risk of embolic stroke, heart failure and cognitive impairment; reduced quality of life; and higher overall mortality.

Optimal clinical management of AF is critical to reducing this associated morbidity and mortality, and includes prevention of AF-related thromboembolic events in at-risk patients. Vitamin K antagonists or direct oral anticoagulants have been shown to reduce thromboembolic events, but long-term use of these medications puts certain patients at higher risk for serious bleeding events. As such, accurate risk stratification for both thromboembolic and bleeding risk is paramount in identifying patients for whom anti-thrombotic therapy would achieve maximum treatment benefit with the lowest risk of complications.

Unfortunately, it is challenging to estimate the trade-off between stroke risk and risk of bleeding complications from long-term anticoagulation therapy because many risk factors for stroke are also associated with increased risk of bleeding. There are several available risk stratification tools used to determine thromboembolic and bleeding risk that incorporate diagnostic imaging as well as patient factors such as age, sex and history of heart disease to aid in clinical decision-making around treatment strategies for AF. Of the many available risk stratification tools, the 2014 American Heart Association/American College of Cardiology/Heart Rhythm Society (AHA/ACC/HRS) guideline for patients with AF recommends the use of the CHA<sub>2</sub>DS<sub>2</sub>-VASc score to estimate the stroke risk and the HAS-BLED score for bleeding risk.

However, these risk scores have been previously categorized as poor to moderate predictors of risk, and are just two of many different published and validated methods for assessing stroke and bleeding risk in patients with AF. Because patients, providers and policymakers have numerous decision tools that could inform treatment decisions and policy recommendations, there is a need for a compilation and analysis of the currently available data. This systematic review was commissioned by the Patient-Centered Outcomes Research Institute (PCORI) to update a 2013 Agency for Healthcare Research and Quality (AHRQ) review, with a focus on evaluating the comparative diagnostic accuracy and impact on clinical decision-making of available clinical and imaging tools and associated risk factors for predicting thromboembolic and bleeding risk in U.S. patients with AF. Our findings related to stroke prevention treatments are discussed in a companion paper.

**Methods**

Methods for this updated comparative effectiveness review (CER) follow the AHRQ’s Methods Guide for Effectiveness and Comparative Effectiveness Reviews (hereafter referred to as the Methods Guide) and Methods Guide for Medical Test Reviews (hereafter referred to as the Medical Test Guide). This article is part of the larger updated review; complete details of our methods, including exact search strings, and full results and conclusions can be found in the full report, available at www.effectivehealthcare.ahrq.gov.

**Defining the Key Questions**

PCORI convened two multi-stakeholder virtual workshops in December 2016 and January 2017 to (1) gather input from end users and clinical, content and methodological experts on scoping for the updated review; (2) prioritize the key questions; (3) discuss changes in the evidence base since the 2013 review; and (4) explore emerging issues in AF. The protocol for this systematic review was informed by discussion at the January 2017 workshop and builds upon the original report. The final protocol for this review is posted on the Effective Health Care (EHC) website (www.effectivehealthcare.ahrq.gov) and registered at PROSPERO (CRD42017069999).

**PCORI convened two multi-stakeholder virtual workshops in December 2016 and January 2017 to (1) gather input from end users and clinical, content and methodological experts on scoping for the updated review; (2) prioritize the key questions; (3) discuss changes in the evidence base since the 2013 review; and (4) explore emerging issues in AF. The protocol for this systematic review was informed by discussion at the January 2017 workshop and builds upon the original report. The final protocol for this review is posted on the Effective Health Care (EHC) website (www.effectivehealthcare.ahrq.gov) and registered at PROSPERO (CRD42017069999).
In this article, we summarize the evidence and findings related to two key questions (KQs): (1) In patients with non-valvular AF, what are the comparative diagnostic accuracy and impact on clinical decision-making (diagnostic thinking, therapeutic and patient outcome efficacy) of available clinical and imaging tools and associated risk factors for predicting thromboembolic risk? and (2) In patients with non-valvular AF, what are the comparative diagnostic accuracy and impact on clinical decision-making (diagnostic thinking, therapeutic and patient outcome efficacy) of clinical tools and associated risk factors for predicting bleeding events?

**Data Sources and Study Selection**

In consultation with an expert medical librarian, we searched PubMed, Embase and the Cochrane Database of Systematic Reviews for relevant literature published from 1 August 2011 to 14 February 2018 (exact search strings are given in [Supplementary Table S1](#)), available in the online version). We supplemented electronic searches with a manual search of citations from a set of systematic review articles. Our findings were combined with those from the 2013 review, and so the literature summarized here reflects evidence back through 1 January 2000. Due to updates in inclusion criteria, any studies excluded from the original review were also re-reviewed for eligibility. We used search criteria to identify relevant on-going clinical trials through ClinicalTrials.gov as well as citations to guide the conclusions ([Supplementary Table S1](#), available in the online version).

Our pre-specified inclusion and exclusion criteria are given in [Supplementary Table S2](#) (available in the online version). We included English-language studies of adults with non-valvular AF (including atrial flutter) that reported the efficacy of clinical or imaging tools, or patient risk factors, on predicting thromboembolic and/or bleeding outcomes. Clinical or imaging tools considered for predicting thromboembolic events were CHADS2 score, CHA2DS2-VASC score, Framingham risk score, age, biomarkers, and clinical history (ABC) stroke score, transarachic and transoesophageal echocardiography, computed tomography scans and cardiac magnetic resonance imagerings (MRIs). Clinical or imaging tools considered for predicting bleeding events were the HAS-BLED score, HEMORRHAGES score, Anticoagulation and Risk Factors in Atrial Fibrillation (ATRIA) score, Bleeding Risk Index (BRI) and ABC bleeding risk score. Thromboembolic outcomes included cerebrovascular infarction, transient ischaemic attack and systemic embolism (excluding pulmonary embolism and deep vein thrombosis). Bleeding outcomes included haemorrhagic stroke, intra-cranial haemorrhage (ICH) and major and minor bleeds. We excluded studies that evaluated patients exclusively from Asia, Africa or the Middle East. We also sought to identify studies which were linked to an individual study.

**Data Extraction and Quality Assessment of Individual Studies**

Pairs of investigators screened all citations and abstracts for eligibility, and those considered relevant by either investigator advanced to full-text review. Paired investigators then reviewed all full-text articles and resolved disagreements through discussion or adjudication by a third investigator. Paired investigators independently abstracted data and assessed study quality. Disagreements were resolved by consensus or arbitration by a third investigator. Articles that represented evidence from the same overall study were linked to avoid duplication of patient cohorts.

We assessed methodological quality, or risk of bias, for each individual study using tools specific to the study’s characteristics. For studies assessing diagnostic accuracy, we used the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. Our outcome-specific quality assessment classified study outcomes as containing low, medium or high risk of bias as defined by QUADAS-2.

**Data Synthesis and Analysis**

We summarized key features of the included studies for each KQ, including information on study design; patient characteristics; clinical settings; diagnostic tools; and intermediate, final and adverse event outcomes. We ordered our findings by diagnostic comparison, and then within these comparisons by outcome, with long-term final outcomes emphasized.

Grouping interventions by prediction tool, we determined the feasibility of completing a quantitative synthesis (i.e. meta-analysis) based on the volume of relevant literature (at least three appropriate studies), conceptual homogeneity of the studies in terms of study population and outcomes and completeness of the reporting of results. When at least three comparable studies reported the same outcome, we used the R statistical package (version 3.1.2) (The R Foundation) with the ‘metafor’ meta-analysis library (version 1.9–7) to synthesize available c-statistics, which quantify the discrimination ability of the studied tools, for each appropriate thromboembolic or bleeding risk prediction tool. We used the random-effects DerSimonian and Laird estimator to generate summary values. In addition, we used the Knapp–Hartung approach to adjust the standard errors of the estimated coefficients. Since the diagnostic tools considered are not binary, it was not possible to consider summary receiver operating characteristic curves. When possible, the c-statistics were pooled by considering their estimated values (point estimates) and confidence intervals (CIs), and the ‘generic point estimates’ effect specification option in the Comprehensive Meta-Analysis software. For a clinical prediction rule, we assumed that a c-statistic of < 0.6 had no clinical value, 0.6 to 0.7 had limited value, 0.7 to 0.8 had modest value and > 0.8 had discrimination adequate for genuine clinical utility.

**Strength of Evidence**

We assigned strength of evidence scores for each diagnostic tool using the approach described in the AHROG’s Methods Guide. We assessed five domains: study limitations; consistency; directness; precision; and reporting bias, which includes publication bias, outcome reporting and analysis reporting bias. These domains were considered qualitatively,
and a summary rating of high, moderate or low strength of evidence was assigned for each outcome after independent assessment and discussion by two reviewers. In cases where ratings were impossible or imprudent to make, a grade of ‘insufficient’ was assigned.

Role of the Funding Source
This topic was nominated and funded by PCORI for systematic review by an Evidence-based Practice Center in partnership with AHRQ. A representative from AHRQ served as a Contracting Officer’s Representative (COR) and provided technical assistance during the conduct of the full evidence report. The AHRQ COR and PCORI program officers provided comments on draft versions of the protocol and full evidence report. PCORI and AHRQ did not directly participate in the literature search; determination of study eligibility criteria; data analysis or interpretation; or preparation, review or approval of the manuscript for publication.

Results
We screened 11,274 publications and found 45 articles (25 studies) for KQ1 and 34 articles (18 studies) for KQ2 that investigated our included tools for determining stroke or bleeding risk in patients with non-valvular AF and that met the other inclusion criteria. We combined these newly identified studies with those included in the 2013 review, yielding a total of 83 articles (61 studies) for KQ1 and 57 articles (38 studies) for KQ2 included in this updated review (Fig. 1). Complete results of the review, including long-term stroke and bleeding risk summaries, are in the full report.

Predicting Thromboembolic Risk in Patients with AF
We considered findings from the 61 studies reporting the predictive value of the CHADS2, CHA2DS2-VASc, Framingham and ABC stroke clinical tools for thromboembolic risk (Table 1). Twenty-nine studies directly compared the predictive ability for thromboembolic events of the CHADS2 risk score with other risk scores,17–24,35,36,54,55 24 compared CHA2DS2-VASc,18–21,23,24,26,37,39–54 6 compared Framingham18,24,33,34,37,55 and 4 compared ABC stroke.54,56–58 e-Statistics for predicting thromboembolic risk, when available, are reported in Supplementary Table S3 (available in the online version). Sufficient data existed to permit meta-analysis of studies evaluating e-statistics for the CHADS2 score using a continuous score (Fig. 2A) and categorical score (Fig. 2B), the CHA2DS2-VASc continuous score (Fig. 2C) and categorical score (Fig. 2D), the Framingham categorical score (Fig. 2E) and the ABC stroke categorical score (Fig. 2F). For both the continuous and categorical CHADS2 scores (continuous: 14 studies with 489,335 patients; categorical: 16 studies, 548,464 patients; Table 2), there was moderate strength of evidence that the scores provide limited prediction of stroke events (continuous: e-statistic of 0.69; 95% CI, 0.66–0.73; categorical: e-statistic of 0.66; 95% CI, 0.63–0.69). There was also moderate strength of evidence (16 studies, 511,481 patients) that the continuous CHA2DS2-VASc score provides limited prediction of stroke events (e-statistic of 0.66; 95% CI, 0.63–0.69). For the categorical CHA2DS2-VASc score (13 studies, 496,683 patients), there was low strength evidence of its ability to predict stroke risk (e-statistic of 0.64; 95% CI, 0.58–0.70). Based on a meta-analysis of 6 studies (282,572 patients), we found moderate strength of evidence that the categorical Framingham score provides limited prediction of stroke events (e-statistic of 0.63; 95% CI, 0.62–0.65). For the categorical ABC score (4 studies, 25,614 patients), we found a moderate strength of evidence of limited prediction of stroke events (e-statistic of 0.67; 95% CI, 0.63–0.71 (Table 2).

Seven imaging studies examined specific anatomical findings and their association with stroke risk in patients with AF.59–65 Imaging studies included MRI, magnetic resonance angiography quantification of left atrial appendage dimensions, transoesophageal echocardiography and transthoracic echocardiography. There was insufficient evidence for the relationship between findings on echocardiography (trans-thoracic) and subsequent stroke based on 7 studies (4 low risk of bias, 3 medium risk of bias; 4,962 patients) that reported discrepant results.

We found 20 studies that evaluated either the predictive role of international normalized ratio (INR), pattern of AF, renal impairment or other risk factors.31,43,48,57,66–75 There was insufficient evidence, however, for further meta-analysis of the results. These abstracted data are in the full report.

Predicting Bleeding Risk in Patients with AF
Of the 38 studies which explored bleeding risk in patients with AF, 26 studies evaluated various risk scores (BRI, HEMORR2HAGES, HAS-BLED, ATRIA, ABC) for estimating the outcome of major bleeding risk in patients with AF, including patients on warfarin, aspirin and no anti-thrombotic therapy.9,18,21,22,46,54,76–97 Thirteen studies (10 low risk of bias, 2 medium risk of bias, 1 high risk of bias; 351,985 patients) compared different risk scores (BRI, HEMORR2HAGES, HAS-BLED, ATRIA, ABC) in predicting major bleeding events in AF patients on warfarin. These studies differed markedly in population, major bleeding rates and statistics reported for evaluating risk prediction scores for major bleeding events.

Assessment of major bleeding events based on individual risk factors was reported by 17 studies (Supplementary Table S4, available in the online version). Eight of these (7 low risk of bias, 1 medium risk of bias; 322,010 patients) evaluated the risk of major bleeding in patients with chronic kidney disease (CKD). All studies demonstrated increased risk of bleeding in patients with CKD (moderate strength of evidence). Other risk factors abstracted included the impact of INR, age, prior stroke, presence of heart disease, diabetes mellitus, sex, cancer, race/ethnicity and cognitive impairment; however, the evidence was insufficient to support findings (results in full report).

Most available studies for KQ2 included ICH within the outcome ‘major bleeding’, but three studies presented this outcome separately. One of these studies evaluated both HAS-BLED and HEMORR2HAGES,18 another study evaluated...
both HAS-BLED and ATRIA\textsuperscript{57} and a third study evaluated the INR.\textsuperscript{66} The single included study comparing HAS-BLED and HEMORR\textsubscript{2}HAGES did not show a statistically significant difference between the risk scores in prediction abilities for ICH in any patient population. Better understanding of ICH risk prediction will be particularly important, because this represents the most devastating variety of major bleeding event that patients on anticoagulation suffer.

The comparative risk discrimination abilities of each clinical tool was evaluated, when data were available, for (1) major bleeding risk in AF patients on warfarin, (2) AF patients on aspirin alone, (3) AF patients not on therapy and (4) ICH risk in AF patients on warfarin (see \textsuperscript{\textit{Supplementary Table S5}} for c-statistics, available in the online version). For AF patients on warfarin, evidence favoured HAS-BLED based on two studies demonstrating that it has significantly higher prediction (by c-statistic) for major bleeding events than other scores among patients on warfarin, but the majority of studies showed no statistically significant differences in prediction abilities, reducing the strength of evidence (moderate; \textit{\textsuperscript{-Table 3}}). For AF patients on aspirin alone, three studies (2 low risk of bias, 1 medium risk of bias; 177,538

---

\textbf{Fig. 1} Literature flow diagram. KQ, key question.
Table 1  Description and interpretation of included risk scores

<table>
<thead>
<tr>
<th>Thromboembolic risk score</th>
<th>Reference</th>
<th>Risk factors included</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHADS2</td>
<td>Gage et al, 2001(^35)</td>
<td>Congestive heart failure, hypertension, age (\geq 75), diabetes mellitus, prior stroke/transient ischaemic attack [2 points]</td>
<td>Low (0), moderate (1–2), high (3–6)</td>
</tr>
<tr>
<td>CHA(_2)DS(_2)-VASc</td>
<td>Lip et al, 2010(^37)</td>
<td>Congestive heart failure/left ventricular ejection fraction (&lt; 40%), hypertension, age (\geq 75) [2 points], diabetes mellitus, prior stroke/transient ischaemic attack/thromboembolism [2 points], vascular disease, age 65–74, sex category female</td>
<td>Low (0), moderate (1), high (2–9)</td>
</tr>
<tr>
<td>Framingham</td>
<td></td>
<td>Advancing age, female sex, increasing systolic blood pressure, prior stroke or transient ischaemic attack and diabetes</td>
<td></td>
</tr>
<tr>
<td>ABC</td>
<td>Hijazi et al, 2016(^93)</td>
<td>Age, biomarkers (cTnI-hs and NT-proBNP), and clinical history (prior stroke/TIA)</td>
<td>Low &lt; 1%, moderate 1–2%, high &gt; 2%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bleeding risk score</th>
<th>Reference</th>
<th>Risk factors included</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>Hijazi et al, 2016(^93)</td>
<td>Age, biomarkers [GDF-15, cTnT-hs, and haemoglobin], and clinical history (previous bleeding)</td>
<td>Low &lt; 1%, medium 1–2%, high &gt; 2%</td>
</tr>
<tr>
<td>ATRIA</td>
<td>Fang et al, 2011(^76)</td>
<td>Anaemia, renal disease (CrCl &lt; 30) (3 points each); age (\geq 75) (2 points); any prior bleeding, hypertension (1 point each)</td>
<td>Low (0–3), moderate (4), high (5–10)</td>
</tr>
<tr>
<td>BRI</td>
<td>Beyth et al, 1998(^109)</td>
<td>Age (\geq 65), GI bleed in past 2 wk, previous stroke, co-morbidities (recent MI, haematocrit &lt; 30%, diabetes, creatinine &gt; 1.5), with 1 point for presence of each condition and 0 if absent</td>
<td>Low (0), moderate (1–2), high (3–4)</td>
</tr>
<tr>
<td>HAS-BLED</td>
<td>Pisters et al, 2010(^9)</td>
<td>Hypertension, abnormal renal (CrCl &lt; 50) or liver function (1 point each); stroke, bleeding history or predisposition, labile INR (TTR &lt; 60%), age &gt; 65, drugs of interest/alcohol (1 point each)</td>
<td>Low (0), moderate (1–2), high ((\geq 3))</td>
</tr>
<tr>
<td>HEMORR(_2)HAGES</td>
<td>Gage et al, 2006(^29)</td>
<td>Liver/renal disease, ethanol abuse, malignancy, age (\geq 75), low platelet count or function, re-bleeding risk, uncontrolled hypertension, anaemia, genetic factors (CYP2C9), risk of fall or stroke (1 point for each risk factor present with 2 points for previous bleed)</td>
<td>Low (0–1), moderate (2–3), high ((\geq 4))</td>
</tr>
</tbody>
</table>

Abbreviations: ABC, age, biomarkers, clinical history; ATRIA, Anticoagulation and Risk Factors in Atrial Fibrillation; BRI, Bleeding Risk Index; CrCl, creatinine clearance; cTnT-hs, high-sensitivity cardiac troponin T; GDF, growth differentiation factor-15; GI, gastrointestinal; HAS-BLED, Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile INR (TTR < 60%), age > 65, drugs of interest/alcohol; HEMORR\(_2\)HAGES, Hepatic or renal disease, Ethanol abuse, Malignancy, Older (age \(\geq 75\) years), Reduced platelet count or function, Re-bleeding risk (2 points), Hypertension (uncontrolled), Anaemia, Genetic factors, Excessive fall risk; Stroke; INR, international normalized ratio; MI, myocardial infarction; TTR, time in therapeutic range.
patients) comparing different combinations of bleeding risk scores (BRI, HEMORR2HAGES and HAS-BLED) in predicting major bleeding events showed no statistically significant differences (low strength of evidence). Among AF patients not on therapy, six studies (4 low risk of bias, 2 medium risk of bias; 310,607 patients) comparing different combinations of bleeding risk scores (BRI, HEMORR2HAGES, HAS-BLED and ATRIA) in predicting major bleeding events showed no statistically significant differences (low strength of evidence). Evaluating ICH in AF patients on warfarin, one study (low risk of bias; 48,599 patients) compared HEMORR2HAGES and HAS-BLED in predicting ICH. This study showed no statistically significant difference in prediction abilities between the two scores (low strength of evidence).

**Discussion**

Our review included studies comparing the diagnostic accuracy and impact on clinical decision-making of available clinical tools, imaging tools and associated risk factors for
predicting thromboembolic and bleeding risk in patients with AF. For predicting thromboembolic risk, the CHADS2, CHA2DS2-VASc and ABC scores appeared similar and had the best predictive abilities given the available evidence, but this advantage was not substantial on an absolute basis. Imaging risk tools, however, found conflicting results when the presence of a left atrial thrombus was assessed, and there was insufficient evidence to support conclusions regarding the predictive ability of the presence of a left atrial thrombus. Among the tools for predicting risk of major bleeding and ICH, there was a suggestion that HAS-BLED is the best score for predicting major bleeds in patients on warfarin, although it only has modest prediction abilities. However, the majority of studies for other patient scenarios showed no statistically significant differences in predictive accuracy among tools.

**Findings in Relation to What is Already Known**

ESC guidelines recommend using the CHA2DS2-VASc score, and AHA guidelines recommend using the CHADS2 or CHA2DS2-VASc to categorize thromboembolic risk when making treatment decisions in patients with AF. Additionally, recent American College of Clinical Pharmacy (ACCP), Australian and New Zealand (ANZ), and Asia Pacific Heart Rhythm Society (APHRS) guidelines endorse using the CHA2DS2-VASc score (excluding sex in the calculation under ACCP and ANZ guidelines) to identify low-risk patients that can be excluded from anticoagulation. In the current CER, we found that of the available risk scores, the CHADS2 and CHA2DS2-VASc scores are the most commonly studied and that the CHADS2, CHA2DS2-VASc and ABC risk scores appeared to be similar and to have the highest predictive ability for stroke events. While some studies have explored the inclusion of biomarkers in stroke risk scores such as the ABC stroke risk score, and preliminary evidence supports the ABC score being comparable to CHADS2 and CHA2DS2-VASc, the experience with ABC is limited and more data are needed on the contribution of these and other biomarkers to the overall risk assessment. Further, few comparisons of the ABC score in predicting thromboembolic risk have been completed in ‘real-world’ populations, which may better clarify its predictive ability.

In predicting bleeding risk, our review found limited evidence favouring the HAS-BLED risk score based on two studies demonstrating that it has a significantly higher predictive ability for major bleeding events than other scores among patients on warfarin. The majority of studies, however, showed no statistically significant differences in prediction, which reduced the strength of evidence. Recent evidence suggests that inclusion of time to therapeutic range (TTR), included in the HAS-BLED score, might enhance the predictive ability of other bleeding scores. Bleeding risk scores are not included in the most recent AHA/ACC guideline recommendations on AF, and they are generally not used to decide whether to prescribe an oral anticoagulant to individual patients. However, bleeding risk scores may inform shared decision-making discussions of the risks of stroke and bleeding incorporating patients’ values and preferences.
preferences. As more data on stroke and bleeding risk scores emerge, it is possible that improvement in the tools and methods for risk stratification of both stroke and bleeding will be important to better individualize treatment using different oral anticoagulants in patients with AF.

Limitations of the Evidence Base and the Comparative Effectiveness Review Process
Comparisons across studies were difficult due to varying categorical arrangements of stroke risk scores, inter-study differences in approach to calculating some of the bleeding risk scores, limited comparison of bleeding risk scores across populations, heterogeneous patient populations and the variability in treating patients with anti-platelets and oral anticoagulants. It is known that risk scores correlate to differing event rates based on patient setting and treatment, such as whether they are in a clinical trial or in the outpatient setting, which further added to between study event rate discrepancies. Additionally, there was inconsistency among individual studies in reporting measures of calibration, strength of association and diagnostic accuracy. While the nature of a meta-analysis precludes the ability to directly account for individual study-level bias, we were able to
Table 2  Strength of evidence domains for prediction of thromboembolic risk

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Number of studies (subjects)</th>
<th>Risk of bias</th>
<th>Consistency</th>
<th>Directness</th>
<th>Precision</th>
<th>SOE and effect (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHADS₂ (Categorical)</td>
<td>16 (548,464)</td>
<td>Observational/ Moderate</td>
<td>Inconsistent</td>
<td>Direct</td>
<td>Precise</td>
<td>SOE = Moderate Limted risk prediction ability (c-statistic 0.66, 95% CI, 0.63–0.69)</td>
</tr>
<tr>
<td>CHADS₂ (Continuous)</td>
<td>14 (489,335)</td>
<td>Observational/ Moderate</td>
<td>Inconsistent</td>
<td>Direct</td>
<td>Precise</td>
<td>SOE = Moderate Limted risk prediction ability (c-statistic = 0.69; 95% CI, 0.66–0.73)</td>
</tr>
<tr>
<td>CHA2DS2-VASc (Categorical)</td>
<td>13 (496,683)</td>
<td>Observational/ Moderate</td>
<td>Inconsistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>SOE = Low Limit risk prediction ability (c-statistic = 0.64; 95% CI, 0.58–0.70)</td>
</tr>
<tr>
<td>CHA2DS2-VASc (Continuous)</td>
<td>16 (511,481)</td>
<td>Observational/ Moderate</td>
<td>Inconsistent</td>
<td>Direct</td>
<td>Precise</td>
<td>SOE = Moderate Limit risk prediction ability (c-statistic = 0.66; 95% CI, 0.63–0.69)</td>
</tr>
<tr>
<td>Framingham (Categorical)</td>
<td>6 (282,572)</td>
<td>Observational/ Moderate</td>
<td>Consistent</td>
<td>Direct</td>
<td>Precise</td>
<td>SOE = Moderate Limit risk prediction ability (c-statistic = 0.63; 95% CI, 0.62–0.65)</td>
</tr>
<tr>
<td>Framingham (Continuous)</td>
<td>4 (274,538)</td>
<td>Observational/ Moderate</td>
<td>Consistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>SOE = Low Limit risk prediction ability (c-statistic ranges between 0.64 and 0.69 across studies)</td>
</tr>
<tr>
<td>ABC (Categorical)</td>
<td>4 (25,614)</td>
<td>Observational/ Moderate</td>
<td>Consistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>SOE = Moderate Limit risk prediction ability (c-statistic = 0.67; 95% CI, 0.63–0.71)</td>
</tr>
<tr>
<td>Imaging risk tools</td>
<td>7 (4,962)</td>
<td>Observational/ Moderate</td>
<td>Inconsistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>SOE = Insufficient</td>
</tr>
</tbody>
</table>

Abbreviations: CHA2DS2-VASc, Congestive heart failure/left ventricular ejection fraction ≤ 40%, Hypertension, Age ≥ 75 (2 points), Diabetes mellitus, prior Stroke/transient ischaemic attack/thromboembolism (2 points), Vascular disease, Age 65–74, Sex category female; CHADS₂, Congestive heart failure, Hypertension, Age ≥ 75, Diabetes mellitus, prior Stroke/transient ischaemic attack (2 points); CI, confidence interval; SOE, strength of evidence.
Table 3 Strength of evidence domains for prediction of bleeding risk

<table>
<thead>
<tr>
<th>Outcome Description</th>
<th>Number of studies (subjects)</th>
<th>Risk of bias</th>
<th>Consistency</th>
<th>Directness</th>
<th>Precision</th>
<th>SOE and effect (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary statistic (patients on warfarin)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRI</td>
<td>4 (11,939)</td>
<td>Observational/Moderate</td>
<td>Consistent</td>
<td>Direct</td>
<td>Precise</td>
<td>SOE = Moderate Limited risk discrimination ability (c-statistic ranging from 0.56 to 0.65)</td>
</tr>
<tr>
<td>HEMORR2HAGES</td>
<td>10 (115,348)</td>
<td>Observational/Moderate</td>
<td>Consistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>SOE = Moderate Limited risk discrimination ability (c-statistic ranging from 0.53 to 0.78)</td>
</tr>
<tr>
<td>HAS-BLED</td>
<td>11 (194,839)</td>
<td>Observational/Moderate</td>
<td>Consistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>SOE = Moderate Modest risk discrimination ability (c-statistic ranging from 0.50 to 0.80)</td>
</tr>
<tr>
<td>ATRIA</td>
<td>7 (76,163)</td>
<td>Observational/Moderate</td>
<td>Inconsistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>SOE = Insufficient</td>
</tr>
<tr>
<td>ABC</td>
<td>1 (22,998)</td>
<td>Observational/Moderate</td>
<td>NA</td>
<td>Direct</td>
<td>Precise</td>
<td>SOE = Low Limited risk discrimination (c-statistic of 0.65 in validation study)</td>
</tr>
<tr>
<td><strong>Comparative risk discrimination abilities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major bleeding events among patients with AF on warfarin</td>
<td>13 (351,985)</td>
<td>Observational/Moderate</td>
<td>Consistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>SOE = Moderate Favours HAS-BLED</td>
</tr>
<tr>
<td>Intra-cranial haemorrhage among patients with AF on warfarin</td>
<td>2 (71,597)</td>
<td>Observational/Moderate</td>
<td>NA</td>
<td>Direct</td>
<td>Precise</td>
<td>SOE = Low No evidence of a difference</td>
</tr>
<tr>
<td>Major bleeding events among patients with AF on aspirin alone</td>
<td>3 (177,538)</td>
<td>Observational/Moderate</td>
<td>Inconsistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>SOE = Low No evidence of a difference</td>
</tr>
<tr>
<td>Major bleeding events among patients with AF not on anti-thrombotic therapy</td>
<td>6 (310,607)</td>
<td>Observational/Moderate</td>
<td>Consistent</td>
<td>Direct</td>
<td>Imprecise</td>
<td>SOE = Low No evidence of a difference</td>
</tr>
</tbody>
</table>

Abbreviations: ABC, age, biomarkers, clinical history; AF, atrial fibrillation; ATRIA, Anticoagulation and Risk Factors in Atrial Fibrillation; BRI, Bleeding Risk Index; CHA2DS2-VASc, Congestive heart failure/left ventricular ejection fraction ≤ 40%, Hypertension, Age ≥ 75 (2 points), Diabetes mellitus, prior Stroke/transient ischaemic attack/thromboembolism (2 points), Vascular disease, Age 65–74, Sex category female; CHADS2, Congestive heart failure, Hypertension, Age ≥ 75, Diabetes mellitus, prior Stroke/transient ischaemic attack (2 points); CI, confidence interval; HAS-BLED, Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly (> 65 years), Drugs/alcohol concomitantly; HEMORR2HAGES, Hepatic or renal disease, Ethanol abuse, Malignancy, Older (age > 75 years), Reduced platelet count or function, Re-bleeding risk (2 points), Hypertension (uncontrolled), Anaemia, Genetic factors, Excessive fall risk, Stroke; INR, international normalized ratio; KQ, key question; NA, not applicable; SOE, strength of evidence.

c-Statistics given are for categorical risk scores unless otherwise noted.
carefully assess for risk of bias, consistency, directness, precision and strength of evidence as outlined by best practice guidelines in systematic review methodology.

Further, our conclusions may be limited by the limitations in the development and validation of risk scores. Specifically, although many of the studies use clinical data sources to derive or validate these risk scores, some studies relied on billing data and institutional electronic medical records to identify patients with AF and co-morbidity information, which could under-estimate stroke risk due to lack of clinical adjudication of events. Likewise, lack of validated results or common event definitions for the endpoints of thromboembolism and bleeding could have under-estimated the performance of these risk scores. Additionally, lack of standard definitions for co-morbidities such as heart failure, diabetes mellitus and hypertension could also lead to discrepancies across studies validating the various risk scores. Moreover, our review included both ambulatory and hospitalized patients, which inherently introduces bias in comparing studies and results in heterogeneity with regards to stability of covariates, concomitant medications, stroke inducing procedures, etc.

Our review methods also had limitations. Our study was limited to English-language publications and excluded studies conducted exclusively in Asia, Africa or the Middle East. We also limited our analysis to studies published since 2000 as the recent literature was considered the most relevant to today's clinical and policy uncertainties. Lastly, we were unable to include systematic review of all available clinical risk score tools for stroke and bleeding risk. We are aware of other tools, such as QStroke and ORBIT scores, but our scope was focused on the scores used most frequently in clinical settings and prioritized through the stakeholder panel and topic refinement process with PCORI.

**Research Recommendations**

In our analyses, we have identified several areas for recommended future research. Given the aforementioned limitations of the currently available studies, further studies are needed that: (1) utilize complete data; (2) use validated clinical outcomes; and (3) compare all available risk scores using consistent and appropriate statistical evaluations.

Despite the availability and validation of numerous tools for both stroke and bleeding risk assessment in patients with non-valvular AF, meaningful comparisons of the tools could not be performed in this CER. Although the 2014 AHA/ACC guideline recommends using the CHA2DS2-VASc score for stroke risk stratification and that all patients with a CHA2DS2-VASc score of ≥ 2 be considered for oral anticoagulant therapy, the guideline acknowledged the limitation of current risk tools, including the CHA2DS2-VASc score, to identify patients at high risk for thromboembolic risk. As a response to this poor predictive ability in high-risk patients, recently published ACCP, ANZ and APHRS guidelines suggest using the CHA2DS2-VASc score to identify low-risk patients in the initial step of determining whether anti-thrombotic therapy should be offered.99–101 Whether biomarkers such as brain natriuretic peptide, C-reactive protein or troponin can enhance the CHA2DS2-VASc score and as a result be incorporated in guideline recommendations remains to be seen.

Also, the current ACC/AHA guidelines7 do not recommend use of bleeding risk scores, but rather focusing on modifiable bleeding risks. Our results found moderate strength of evidence for modest risk discrimination of the HAS-BLED score; how this modestly predictive score could potentially be utilized in clinical treatment decisions has yet to be investigated. Preliminary data in non-clinical trial populations show that biomarkers may not enhance risk scores' predictive ability of bleeding risk and further research is needed to conclusively determine whether biomarkers (e.g. brain natriuretic peptide, C-reactive protein or troponin) can enhance these scores.102,104

With the growing prevalence of digitized medical records, there is an opportunity to continue to evaluate and modify risk prediction tools to improve their accuracy in predicting stroke and bleeding risk, particularly with newer anticoagulants diffusing into clinical practice. These records might also facilitate research investigating risk as a non-static variable, observing changes in risk factors as predictive for stroke or bleeding events.105,106 Also, newer clinical markers (e.g. MRI to assess scar), co-morbidities (i.e. renal failure, etc.) and biomarkers should be tested and validated with or alongside current risk tools to improve their prediction of both stroke and bleeding risks. Additionally, more specific guidelines on how to use risk scores and apply necessary therapies, possibly in the form of physician decision-support tools, will be important for clinical decision-making. Efforts to create computer-based clinical decision-making supporting tools are on-going and may represent a way to better integrate clinical risk tools into practice.107,108 Preliminary evidence of such decision support systems is discussed within the full AHRQ report.

In addition, although we are able to identify patients at risk for stroke, many of these patients are also at a high risk for bleeding. Thus, there is a need for a score that could be used for decision-making about anti-thrombotic therapy in AF patients taking into account both thromboembolic and bleeding risks. Scores that identify only patients at risk for stroke or only those at risk for bleeding are not so helpful since the clinical factors in these scores are usually similar and treatments which reduce one or the other risk may increase the other for the same patient. Another challenge is that both stroke events and bleeding events are on a spectrum of severity and therefore predicting overall stroke might not align with outcomes that matter most to patients. For example, some strokes may have symptoms lasting < 24 hours with complete resolution, whereas others can cause death. It may be good for future risk tools to account for differences in severity of outcomes. Another research need specific to bleeding risk is a prospective comparison of the standard deviation of transformed INR (SDT(INR)) and TTR to establish which variable has better predictive accuracy for major bleeding including ICH.

Additionally, even assuming an optimal risk prediction score can be identified, further work is needed to clarify how scores should be used prospectively in clinical practice.
Clinical risk scores must take into account the balance between simplicity and practicality versus accurate prediction, especially in a high-capacity clinical environment. While clinical risk scores are necessarily reductionist and cannot feasibly consider all patient parameters, our results here show moderate predictive ability of risk scores that can be calculated relatively easily from patient history and demographics. Future research might explore this trade-off between ease of implementation and increasing the predictive value of clinical risk scores with more difficult-to-obtain parameters such as biomarkers.

Conclusion

Overall, we found that CHADS$_2$, CHA$_2$DS$_2$-VASc and ABC stroke scores have the best prediction for stroke events in patients with AF among the risk scores we reviewed, whereas HAS-BLED provides the best prediction for bleeding risk. Imaging tools require further evidence in regard to their appropriate use in clinical decision-making. Additionally, simple clinical decision tools are needed that incorporate both stroke risk and bleeding risk to assist providers treating patients with AF. Additional work will be required to develop risk tools for patients to discriminate those individuals with AF where the bleeding risk may be high enough to warrant more intensive follow-up and monitoring. These tools could be embedded into electronic medical record systems for point-of-care decision-making, developed into applications for smartphones and tablets or be delivered via web-based interfaces. Additional evidence of the use of these stroke and bleeding risk scores (and clinical decision tools which balance these risks) among patients on therapy is also required.

What is known about this topic?

- The comparative diagnostic accuracy and impact on clinical decision-making of available clinical and imaging tools for predicting thromboembolic and bleeding risk in patients with atrial fibrillation (AF) are uncertain.

What does this paper adds?

- CHADS$_2$, CHA$_2$DS$_2$-VASc, and ABC risk scores have the best evidence to support prediction of stroke events.
- HAS-BLED has the best evidence to support prediction of bleeding risk.
- Imaging tools for stroke prediction require further evidence.

Funding

This project was funded under Contract No. 290–2015–00004-I from the Agency for Healthcare Research and Quality (AHRQ), U.S. Department of Health and Human Services. The authors of this manuscript are responsible for its content. Statements in the manuscript should not be construed as endorsement by the Patient-Centered Outcomes Research Institute (PCORI), AHRQ and the U.S. Department of Health and Human Services. AHRQ retains a license to display, reproduce and distribute the data and the report from which this manuscript was derived under the terms of the agency’s contract with the author.

Conflict of Interest

None.

Acknowledgements

The authors thank Jamie Conklin, MSLIS, for help with the literature search and retrieval; Samantha E. Bowen, PhD, and Amanda J. McBroom, PhD, for assistance with project leadership; and Liz Wing, MA, for editorial assistance.

References

observing outcomes associated with varying levels of INR control. Thromb Res 2009;124(1):37–41
34 Olesen JB, Lip GY, Hansen ML, et al. CHADS2-VASc score in atrial fibrillation: a comparison with the CHA2DS2-VASc score. Thorac Cardiovasc Surg 2011;59(03):139–142
risk before cardioversion of non-valvular atrial fibrillation. Am J Cardiol 2013;112(05):678–683
78 Shireman TI, Mahnkopf JD, Howard PA, Kresowik TF, Hou Q, Ellerbeck EF. Development of a contemporary bleeding risk model for elderly warfarin recipients. Chest 2006;130(05):1390–1396
Predicting Stroke and Bleeding in AF Patients  Borre et al.

[References cited in the text are not included in this representation.]


Phelps E, Delate T, Witt DM, Shaw PB, McCool KH, Clark NP. Effect of increased time in the therapeutic range on atrial fibrillation outcomes within a centralized anticoagulation service. Thromb Res 2018;163:54–59


Hankey GJ, Stevens SR, Piccini JP, et al; ROCKET AF Steering Committee and Investigators. Intracranial hemorrhage among patients with atrial fibrillation anticoagulated with warfarin or rivaroxaban: the rivaroxaban once daily, oral direct factor Xa inhibition compared with vitamin K antagonism for prevention of stroke and embolism trial in atrial fibrillation. Stroke 2014;45 (05):1304–1312
