ESC European Society of Cardiology

Europace (2022) **24**, 1739–1753
https://doi.org/10.1093/europace/euac096

CLINICAL RESEARCH
Atrial fibrillation

Comprehensive comparison of stroke risk score performance: a systematic review and meta-analysis among 6 267 728 patients with atrial fibrillation

Vera H.W. van der Endt [1]*, Jet Milders [1], Bas B.L. Penning de Vries [1], Serge A. Trines [2], Rolf H.H. Groenwold [1], Olaf M. Dekkers [1], Marco Trevisan [4], Juan J. Carrero [4], Merel van Diepen [1], Friedo W. Dekker [1], and Ype de Jong [1,3]

[1]Department of Clinical Epidemiology, Leiden University Medical Center, PO Box 9600, 2333 ZA Leiden, The Netherlands; [2]Department of Cardiology, Willem Einthoven Center of Arrhythmia Research and Management, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands; [3]Department of Internal Medicine, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands; and [4]Department of Medical Epidemiology and Biostatistics (MEB), Karolinska Institutet, 171 77 Stockholm, Sweden

Received 20 December 2021; accepted after revision 24 May 2022; online publish-ahead-of-print 27 July 2022

Aims Multiple risk scores to predict ischaemic stroke (IS) in patients with atrial fibrillation (AF) have been developed. This study aims to systematically review these scores, their validations and updates, assess their methodological quality, and calculate pooled estimates of the predictive performance.

Methods and results We searched PubMed and Web of Science for studies developing, validating, or updating risk scores for IS in AF patients. Methodological quality was assessed using the Prediction model Risk Of Bias ASsessment Tool (PROBAST). To assess discrimination, pooled *c*-statistics were calculated using random-effects meta-analysis. We identified 19 scores, which were validated and updated once or more in 70 and 40 studies, respectively, including 329 validations and 76 updates—nearly all on the CHA_2DS_2-VASc and $CHADS_2$. Pooled *c*-statistics were calculated among 6 267 728 patients and 359 373 events of IS. For the CHA_2DS_2-VASc and $CHADS_2$, pooled *c*-statistics were 0.644 [95% confidence interval (CI) 0.635–0.653] and 0.658 (0.644–0.672), respectively. Better discriminatory abilities were found in the newer risk scores, with the modified-$CHADS_2$ demonstrating the best discrimination [*c*-statistic 0.715 (0.674–0.754)]. Updates were found for the CHA_2DS_2-VASc and $CHADS_2$ only, showing improved discrimination. Calibration was reasonable but available for only 17 studies. The PROBAST indicated a risk of methodological bias in all studies.

Conclusion Nineteen risk scores and 76 updates are available to predict IS in patients with AF. The guideline-endorsed CHA_2DS_2-VASc shows inferior discriminative abilities compared with newer scores. Additional external validations and data on calibration are required before considering the newer scores in clinical practice.

Clinical trial registration ID CRD4202161247 (PROSPERO).

Keywords Ischaemic stroke • Risk score • Atrial fibrillation • Meta-analysis • *C*-statistic • Discrimination • Predictive performance • Calibration • Prediction model • CHA2DS2-VASc • CHADS2 • External validation • PROBAST

* Corresponding author: Tel: +31 611820984. *E-mail address*: veravanderendt@gmail.com

Introduction

Patients with atrial fibrillation (AF) are at increased risk of ischaemic stroke (IS) and face poor stroke outcomes including severe morbidity and mortality.[1] Anticoagulation reduces this risk substantially and is therefore prescribed to most patients.[2] In current guidelines, the threshold for initiating anticoagulation therapy is based on the balance between the predicted IS risk with the expected risk of bleeding and associated quality of life.[3] Therefore, accurate prediction of these outcomes is of major importance in AF management.

Since the widespread use of the first risk scores for cardiovascular disease, such as the Framingham risk score (1998), a multitude of risk scores have been developed to predict IS risk in patients with AF.[4] Examples include the $CHADS_2$ (2001) and the commonly used CHA_2DS_2-VASc (2010), with the latter being endorsed by most current clinical guidelines.[5–7] Studies on the predictive performance suggest that current risk scores have comparable but limited overall ability to predict IS in patients with AF.[8–10] Consequently, the use of these scores in clinical practice to allocate anticoagulation treatment is not without risk: overprediction of IS risk will result in overtreatment and higher bleeding rates, whereas higher IS rates will occur following underpredicted risks. To improve IS prediction, new risk scores have been developed and earlier scores updated. As much emphasis has been put on commonly available risk scores, limited research seems to be performed on newer risk scores and updates. For example, external validation, a crucial method in the assessment of predictive performances such as discrimination and calibration, is lacking the newer risk scores, despite the overall increase of external validations conducted over the past years.[8–11] As a result, there may be an undervalued but promising risk score in the literature that is currently not integrated in clinical practice but could improve decision-making.

For the clinician interested in the best risk score to inform on the patient's risk, but also for the researcher aiming to develop or validate a risk score, a comprehensive comparison of all available risk scores, their updates, and validations is essential. Therefore, the objective of this study is to(i) identify and systematically review all available risk scores predicting IS risk in patients with AF,(ii) present an overview on their external validations and updates,(iii) assess the methodological quality of these studies, and (iv) provide a pooled estimate of the predictive performance.

What's new?

- 19 risk scores, 329 external validations and 76 risk score updates have been conducted to accurately predict ischaemic stroke (IS) in patients with atrial fibrillation (AF).
- Despite the development of new risk scores, the choice on initiating antithrombotic therapy for the prevention of IS in patients with AF remains mostly based on the conventional guideline endorsed CHA_2DS_2-VASc (2010).
- Probably due to the inclusion of specific biomarkers, new risk scores on IS risk in patients with AF tend to have better predictive abilities when compared with conventional scores, yet differences are marginal and should be interpreted with caution.
- Although newer scores tend to have better predictive abilities, external validations and studies on calibration are required to confirm this.
- In future studies developing and/or validating risk scores, adhering to the PROBAST and TRIPOD guidelines is warranted in order to minimize the risk of bias.

Methods

The current study is performed and reported in line with the PRISMA,[12] TRIPOD,[13] and CHARMS[14] guidelines, which were followed where applicable. The review was registered in PROSPERO under ID CRD4202161247.

Data sources and eligibility

The present review aims to identify risk scores, and corresponding update and validation studies, on future event of IS in adults with AF. In collaboration with a medical librarian, we conducted two searches in an iterative fashion. First, on 19 May 2021, we conducted a search in PubMed for English-language studies regarding 'risk scores' and 'ischaemic stroke' using methods developed in earlier work.[15,16] Studies were screened independently by two researchers (V.H.W.v.d.E. and J.M.) and were deemed eligible if they met the following criteria:(i) development of a multivariable prognostic risk score,(ii) predicting outcomes including first event of IS from 1 month onward, assessed in a longitudinal design,(iii) in a population of adults (>18 years) with AF. Secondly, on 28 May 2021, for each of the identified risk scores, we looked for studies externally validating or updating these scores, using citation search methods in Web of Science. Studies renewing old risk scores, for example, by adding or replacing a predictor, were regarded as update studies.[17] In the end, we thus compiled three data sets: (i) development: the set of studies in which the scores were originally developed, (ii) validation: the studies in which these scores were validated, and (iii) update: the set of studies in which (at least) one of the scores was updated. Update studies that validated an original score as comparison to the updated score were included in both the validation and update data set. Finally, included articles and reviews were cross-referenced for possible relevant studies. Detailed search methods are displayed in Supplementary material online, section 'Detailed search methods'.

Data extraction and study appraisal

Titles, abstracts, and full texts were screened independently by two researchers (V.H.W.v.d.E. and J.M.). Data extraction was conducted by V.H.W.v.d.E. and J.M., with consultation of Y.d.J. when necessary. Information on the study design, population, outcome, prediction horizon (i.e. the time between prediction and the timeframe in which the outcome may occur), candidate predictors, sample size, risk score development, and risk score performance were extracted and summarized. Risk score predictive performance was evaluated using discrimination and calibration measures. For discrimination, which describes the scores' ability to discriminate between events and non-events, we extracted the *c*-statistic (area under the receiver operating characteristic curve for logistic model and Harrell's *C*-index for Cox model). In general, *c*-statistics of <0.60, 0.60–0.80, and >0.80 are interpreted as poor, reasonable, and good discrimination.[18] For calibration (i.e. the agreement between the predicted and the observed risk), we extracted the calibration-in-the-large, the calibration slope, the Hosmer–Lemeshow, or the Nam–d'Agostino statistic. Studies presenting a calibration plot (i.e. a plot with the relation between the observed and predicted risks) were qualitatively categorized as poor, medium, or good fit.[18] For the update studies, net reclassification index and the integrated discrimination improvement, measures quantifying the improvement when a score is updated, were extracted.[19] Methodological quality was assessed in all

studies using the Prediction model Risk Of Bias ASsessment Tool (PROBAST). This tool consists of 4 domains (participants, predictors, outcome, and analysis) containing 20 signalling questions to determine the risk of bias and three signalling questions to determine applicability to the review question.[20]

Statistical analysis

A random-effects meta-analysis was conducted to summarize discrimination measures of the included original risk scores. It should be noted that pooling calibration-in-the large, the expected/observed ratio and calibration slopes is possible as well.[21] However, due to the limited number of studies assessing calibration and the heterogeneity of these calibration measures, no random-effects meta-analysis was conducted on calibration. Risk scores with more than two external validations were included. *C*-statistics were logit-transformed, and confidence intervals were calculated following the Hartung–Knapp–Sidik–Jonkman approach.[22–24] A logit pooled *c*-statistic was calculated for each of the included models and then transformed back to the original scale.[21,25,26] Forest plots were drawn to visualize the estimated results of all included studies. To assess small study bias, funnel plots were made and Egger's regression tests were performed to test for funnel plot asymmetry.[27,28] All analyses were conducted using RStudio version 1.2.5033 and the metafor package.[29]

Sensitivity analyses

Regarding the validation studies included in the meta-analysis, eight sensitivity analyses were conducted. Per analysis, pooled *c*-statistics were calculated and forest plots were drawn for each individual score. Next, the yielded *c*-statistics were analysed on discrepancies. For the first sensitivity analysis, studies were stratified according to the cohorts' mean or median age: (i) <65 years, (ii) 65–75 yeras, and (iii) >75 years. In the second analysis, we evaluated the effect of outcome measure: IS, thromboembolism, or other outcomes measures. For the third analysis, we calculated *c*-statistics for three patient groups: (i) corrected for anticoagulation, (ii) not corrected for anticoagulation, and (iii) with no information on anticoagulation. In the fourth analysis, we studied the effect of ethnicity and grouped the studies conducted in predominantly Caucasian or Asian participants. For the fifth analysis, we categorized the studies in high or low risk of bias according to the PROBAST. For the sixth analysis, we compared 'true low' risk patients (CHA_2DS_2-VASc/$CHADS_2$ score ≤ 1, ATRIA score ≤ 5) with intermediate- or high-risk patients (CHA_2DS_2-VASc/$CHADS_2$ score ≥ 2, ATRIA score ≥ 6). For analysis 7, we evaluated the effect of study design: observational vs. randomized controlled trial studies. In the last analysis, studies were stratified according to the year of publication to evaluate the effect of inclusion of older study cohorts with possible different baseline risks for IS due to different treatment options at that time. Detailed methods are given in Supplementary material online, section 'Sensitivity analyses'.

Results

Study selection

The study selection is described in Flowcharts 1 and 2 (*Figure 1*). The primary search for studies developing a risk score identified 1496 titles of which 213 abstracts were screened and 71 articles were screened full text for eligibility, leading to 19 studies that were included in the review. Next, we searched for update and validation studies, 2188 unique titles were identified of which 70 validation studies and 40 update studies were included in the present review. As most update studies ($n = 37$, 93%) also validated the original model as comparison for the updated model, we included a total of 107 studies in which a risk score was validated.

Methodological quality of the included studies

Methodological risk of bias was assessed for all included development ($n = 19$), validation ($n = 70$), and update studies ($n = 40$).

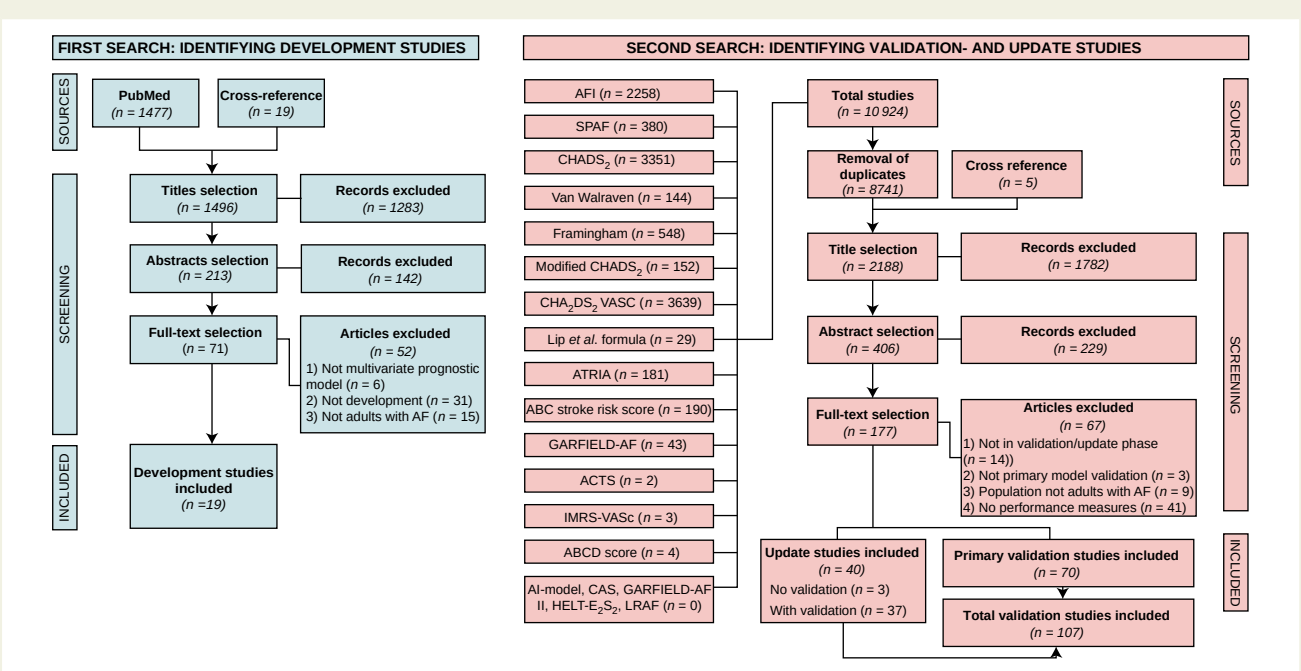

Figure 1 Flowchart study selection. Two iterative searches were performed to identify (i) development studies on risk scores for IS in patients with AF (ii) corresponding validation and update studies.

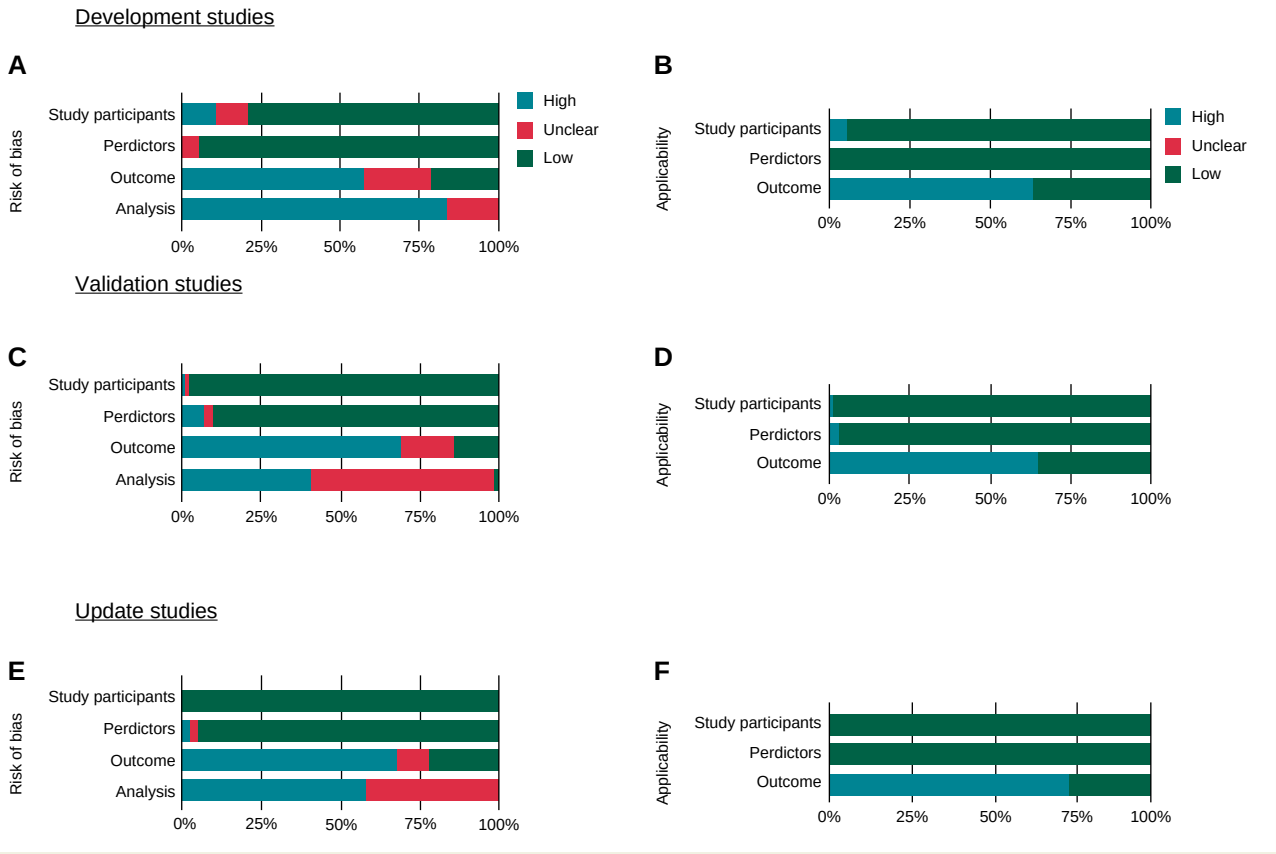

Figure 2 Risk of bias (ROB) and applicability assessment using the PROBAST. (*A* and *B*) development studies ($n = 19$), (*C* and *D*) validation studies ($n = 70$), and (*E* and *F*) update studies ($n = 40$). Blue: high risk of bias, green: low risk of bias, pink: unclear bias due to lack of information.

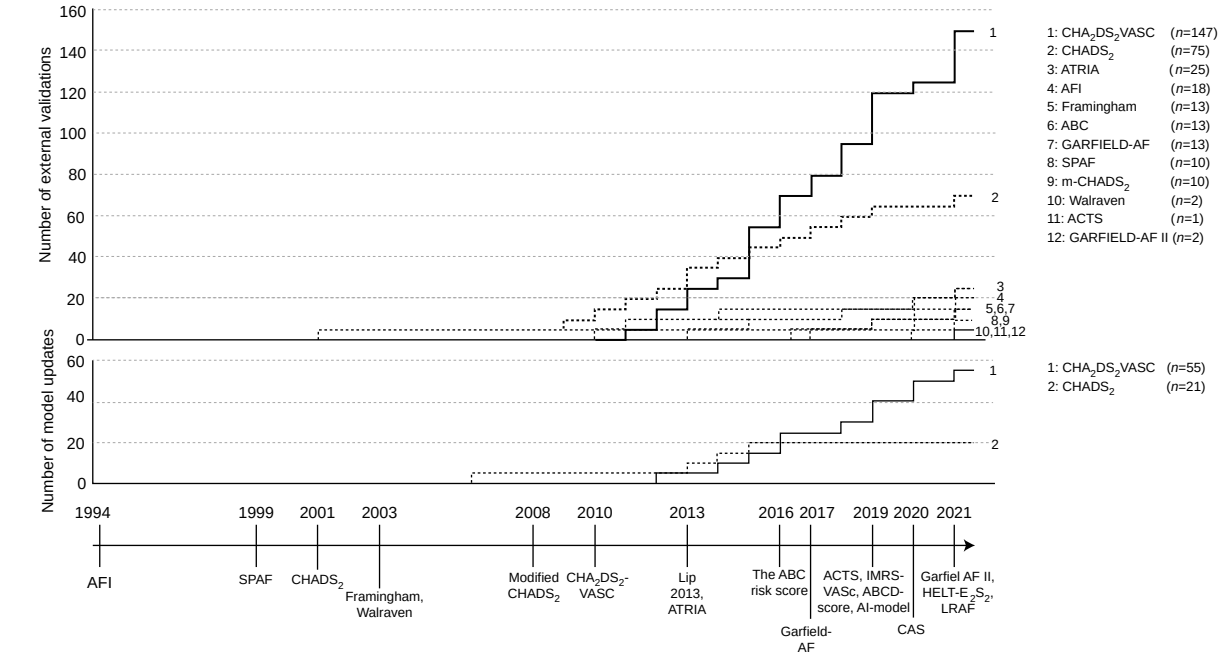

Figure 3 Timeline on the development, validation, and update of risk scores. (Upper panel) The number of external validations plotted over time; (Middle panel) The number of update studies plotted over time; (lower panel) timeline of developed risk scores. Note that the number of validations slightly differs with *Table* 3, which gives the number of validations that were included in the random effects meta-analysis.

For all these studies, risk of bias was low in the *participants* and *predictors* domains. In the *outcome* domain, however, the risk of bias was high or unclear as study outcomes were often ill-defined or, in the case of validation and update studies, determined differently from the original risk scores. In the *analysis* domain, all studies were of high risk of bias, mostly due to statistical considerations that were not correctly addressed, such as omission of calibration, failure to take competing risk into account, or inappropriate methods to handle incomplete data. The applicability of the *participant* domain was of low concern for all studies, and for most studies in the *predictor* domain. For the *outcome* domain, applicability was a concern for most studies due to the use of composite outcomes, whereas the present study focused on the predictive performance of the non-composite outcome of IS. More detailed information on the risk of bias per signalling question is provided in *Figure 2* and in Supplementary material online, section 'Risk of bias of included validation and update studies' for the validation and update studies. The funnel plots showed no clear asymmetry. This was confirmed by the Egger's regression test which showed a *P*-value of >0.05 in all risk scores but two: the modified-$CHADS_2$ showed a *P*-value of <0.05 and for the GARFIELD-AF, no Egger's test could be performed due to the inclusion of only two studies (see Supplementary material online, section 'Analysis on publication bias').

Development studies

Study characteristics

The first of the 19 studies developing a prediction model for IS in AF patients was published in 1994[30] and the most recent included studies were published in 2021[31–34] (*Figure 3*). Risk scores were developed in

Table 1 Predictor use in development studies (*N* =19)

	Patient characteristics							Comorbidities								Treatment				Laboratory findings						Clinical findings	
	Age	Sex	World religion	Race	Smoking	BMI	Type of AF	Previous stroke/TIA^{a}	Previous bleeding	Vascular disease	Heart failure	Hypertension	Diabetes mellitus	Chronic kidney disease	Dementia	Antiplatelet	Oral anticoagulant	Other treatment^{b}	Time in therapeutic range	IMRS^{a}[a]	cTnI/T-hs^{a}[a]	NT-proBNP^{a}[a]	Proteinuria	Creatinine clearance	PT-INR^{a}[a]	Blood pressure	Left atrial dimension
AFI investigators, 1994	x							x				x	x														
Hart, 1999	x	x										x														x	
Gage, 2001	x							x			x	x	x														
Van Walraven, 2003								x		x			x													x	
Wang, 2003	x	x						x					x													x	
Rietbrock, 2008	x	x						x					x														
Lip, 2010	x2	x						x		x	x	x	x														
Lip, 2013	x							x								x			x								
Singer, 2013	x	x									x	x	x										x	x			
Hijazi, 2016	x							x												x	x						
Fox, 2017			x	x				x	x		x			x			x										
Claxton, 2019	x	x						x	x							x	x3	x5									
Horne, 2019	x	x						x				x	x				x			x7								
Shin, 2019	x																					x		x			x
Goto, 2019																									x		
Jiang 2020	x							x			x																
Fox, 2021	x				x			x	x	x	x	x	x	x	x		x										
Okumara, 2021	x2					x	x	x				x															
Arnson, 2021	x									x		x	x	x2													

The inclusion of a predictor is shown as 'x'. The subscript under x indicates the number of predictors included from that category (e.g. 'x2' implies that 2 predictors were used from the same category).

[a]cTnI/T-hs, high-sensitive cardiac troponin I/T; IMRS, intermountain risk score (consists of complete blood count parameters and basic metabolic factors); NT-proBNP, N-terminal-pro hormone Brain natriuretic peptide; PT-INR, prothrombin time international normalized ratio; TIA, transient ischemic attack.

[b]Antiarrhythmic, calcium channer blocker, beta Blocker, lipid lowering medication, or anti-diabetic medication.

cohorts with a sample sizes ranging between 705 and 52 032 patients; event rates ranged between 1.3 and 11.8%. In 11 studies, the follow-up period was described, ranging from 0.9 to 5.5[33–42] and 1.0–1.9[43,44] years in studies presenting a mean and median follow-up period, respectively. Except for four studies,[31,33,34,45], all studies were conducted in a cohort of predominantly Caucasian patients. Six studies[31,36,37,40,43,46] developed their score in a cohort of patients not on anticoagulation, whereas 12 studies[30,32–35,38,39,41,42,44,47,48] included patients on anticoagulation. In one study, no information on anticoagulation was given.[45]

Risk score characteristics

Of the 19 risk scores, most ($n = 13$) studies developed a point-based risk score. In total, 27 different predictors were used of which age ($n = 15$), history of stroke/transient ischaemic attack (TIA) ($n = 12$), diabetes mellitus ($n = 10$), and sex ($n = 7$) were most common (*Table 1*). Other studies developed a mathematical formula,[32,39,41,42] a decision tree,[36] or employed artificial intelligence algorithms.[48] Regarding the statistical analysis method used to develop a model, 13 studies used Cox proportional hazard regression,[30–33,35,37,40–44,46,47] 4 logistic regression,[34,38,39,45] and 2 studies used another regression method.[36,48] The prediction horizon ranged between 31 days and 5 years and was not given for three scores.[33,42,45]

The outcome measure was IS ($n = 5$)[30,33–35,42], undefined stroke ($n = 2$),[45,47] thromboembolic events ($n = 5$, including IS, TIA, venous thrombosis, and/or pulmonary embolism)[31,38,40,41,48] or the composite of thromboembolic events with bleeding events ($n = 7$, including haemorrhagic stroke and/or major bleeding).[32,36,37,39,43,44,46] Discrimination was reported in 16 studies with *c*-statistics ranging from 0.61[38] to 0.86.[45] Calibration was presented as observed vs. expected risks for the ABC-score, GARFIELD-AF, ACTS, and GARFIELD-AF II, demonstrating adequate calibration.[32,41,42,44] The Hosmer–Lemeshow statistic was given in five studies, showing no evidence of poor calibration for ACTS, ATRIA, HELT-E_2S_2, and the ABCD score (χ^2-statistics with a *P*-value of >0.05).[33,40,42,45] Information on characteristics per individual score is presented in *Table 2*; detailed information in Supplementary material online, section 'Overview of included studies'.

Validation and update studies

Validation studies

A total of 107 validation studies were included, of which 60 validated multiple scores, resulting in a total of 327 validations. In total, 359 373 events occurred in 6 267 728 patients; 12 studies[45,49–59] did not provide population size and number of events. Most of the validations were performed on the CHA_2DS_2-VASc ($n = 147$) and $CHADS_2$ ($n = 75$) (*Figure 3*). For the newer scores, the number of external validations was limited: the IMRS-VASc (2019), ABCD (2019), AI-model (2019), CAS (2020), HELT-E_2S_2 (2021), and LRAF (2021) were not validated, and the ACTS (2019) was validated only once. Most outcome measures ($n = 66$) were either defined as IS or thromboembolic events (including IS, TIA, systemic embolism, and pulmonary embolism), whereas other study outcomes ($n = 41$) were undefined stroke or the composite of stroke and bleeding events. The majority of the studies validated a risk score in a general predominantly Caucasian AF population, with the exception of studies validating prior or post-surgery ($n = 6$),[60–65] in cohorts with a specific secondary disease ($n = 11$),[42,44,66–74] or in cohorts with ethnicity other than predominantly Caucasian ($n = 20$).[33,34,45,47,53,55,65,71,75–87] Discrimination was presented in all studies ($n = 107$) and indicated poor (<0.60) to reasonable (0.60–0.80) and in exceptional cases good (>0.80) discrimination; these values are visualized in the forest plots given in Supplementary material online, section 'Random-effects meta-analysis'. For the risk scores developed after the publication of the CHA_2DS_2-VASc (2010), no discrimination lower than 0.60 was reported in the validation studies.[32,40,43–45,47] For calibration, only 14 studies (13%) presented one or more calibration measures: 12 studies presented observed vs. expected risks,[32,41,42,44,49,55,56,70,88–91] 4 the Hosmer–Lemeshow statistic,[40,42,56,92] calibration in the large,[70] and the Nam–D'Agostino statistic.[49] With this limited information on calibration, compared with the CHA_2DS_2-VASc, the modified-$CHADS_2$, ABC-score, and GARFIELD-AF showed improved calibration measures.[70,88,90,91] Detailed characteristics of the validation studies are presented in Supplementary material online, section 'Characteristics of included validation studies'.

Update studies

All of the 40 update studies updated the CHA_2DS_2-VASc ($n = 25$), the $CHADS_2$ ($n = 7$), or both scores ($n = 8$)—no update studies were found on other risk scores. Ten studies updated these scores multiple times, resulting in a total of 55 updated scores for the CHA_2DS_2-VASc and 21 for $CHADS_2$.[59,65,83,85,93–99] Except for seven scores, all scores were updated by adding one or more predictors to the original score (e.g. R_2 CHA_2DS_2-VASc: the original score with the additive predictor 'renal failure'). New predictors included blood or urine biomarkers (e.g. D-dimer, IL-6, soluble fibrin monomer complex),[59,74,82–84,86,92,97,98,100–111] echocardiographic characteristics (e.g. hypertrophic cardiomyopathy),[65,87,99,112] electrocardiographic markers (e.g. P-wave indices),[92] genetics (e.g. microRNAs),[113] or socio-economic status.[114] In the remaining scores, either predictors were left out,[85,95] replaced,[80,85] or the original predictors were assessed over time (i.e. differences between the baseline and follow-up CHA_2DS_2-VASc scores).[81,94] For the CHA_2DS_2-VASc, discrimination was available for 49 updates (89%): *c*-statistics improved for all but three updates when compared with the original CHA_2DS_2-VASc.[57,115,116] For the $CHADS_2$, discrimination was presented in 17 updates (80%): *c*-statistics improved for all but four updates when compared with the original $CHADS_2$.[85,115,116] More detailed information on the new scores and their predictive performances are shown in Supplementary material online, section 'Characteristics of included update studies'.

Pooled *c*-statistic

For 10 scores (CHA_2DS_2-VASc, $CHADS_2$, AFI, Framingham, ATRIA, SPAF, ABC, modified-$CHADS_2$, GARFIELD-AF, and GARFIELD-AF II) pooled *c*-statistics were calculated in a random-effects meta-analysis. Pooled *c*-statistics ranged from 0.598 [95% confidence interval (CI) 0.558–0.636, 12 validations] for the AFI to 0.715 (0.674–0.754, 10 validations) for the modified-$CHADS_2$ (*Table 3* and *Figure 4*). Pooled *c*-statistics were 0.644 (0.635–0.653) and 0.658 (0.644–0.672) for the CHA_2DS_2-VASc and $CHADS_2$, respectively. All risk scores that were developed after the publication of the CHA_2DS_2-VASc (2010) showed better discriminative abilities when compared with the CHA_2DS_2-VASc (*Figure 4*). For the CHA_2DS_2-VASc, $CHADS_2$, and ATRIA, pooled *c*-statistics were derived from >10 studies per risk score including large sample sizes ranging

Table 2 Baseline characteristics and predictive performances of the risk scores

Year of publication	1994	1999	2001	2003	2003	2008	2010	2013
Risk score name	**AFI**	**SPAF**	**$CHADS_2$**	**—**	**Framingham**	**Modified-$CHADS_2$**	**CHA_2DS_2-VASc**	**—**
Risk score type	Risk score	Risk score	Risk score	Prediction rule	Risk score	Risk score	Risk score	Formula
Study design	RCT	RCT	Cohort (retrospective)	RCT	Cohort (prospective)	Case–control	Cohort (prospective)	RCT
Country	USA	USA	USA	USA, Canada, Denmark, and The Netherlands	USA	UK	>10 countries	>10 countries
Outcome	IS	IS	Composite (IS, TIA, undefined cerebral event)	Composite (IS, haemorrhagic stroke, TIA)	Composite (IS, haemorrhagic stroke)	Composite (IS, haemorrhagic stroke)	Composite (IS, systemic embolim, pulmonary embolism)	Composite (IS, systemic embolus, major bleeding)
Method	Cox	Cox	Cox	Recursive partitioning	Cox	Cox	Logistic	Logistic
Follow-up	—	2 years[a]	1 years[b]	1.9 years[a]	4.3 years[a]	—	1 year[a]	0.9 years[a]
Prediction horizon	1 year	1 year	1 year	1 year	5 years	5 years	1 year	1 year
n events/*n* total (%)	208/5955[c] (3.5)	130/2012 (6.5)	94/1733 (5.4)	103/1661 (6.2)	83/705 (11.8)	5526/51807 (10.7)	25/1084 (2.3)	50/2292 (2.2)
n cand. pred (EPV)	15 (14)	30 (4)	5 (19)	18 (6)	9 (9)	9 (614)	9 (3)	14 (4)
Internal validation	—	—	Bootstrapping	Split sample	Bootstrap	—	—	—
Discrimination (*C*-statistic)	—	—	0.82 (0.80–0.84)	—	0.66 (SD 0.03)	0.72 (0.72–0.73)	0.61 (0.51–0.70)[d]	0.73 (0.57–0.73)
Calibration	—	—	—	—	Hosmer–Lemeshow	—	—	—
Mean age (SD)	69[a]	69[a] (SD 10)	81[a]	70[a] (SD 10)	75	[e]	66[a] (SD 14)	70 (SD 9)
Male sex (%)	—	72	42	67	52	51	59	65
Treatment (%)								
Antiplatelet	50%	100%	31%	100%	[e]	[e]	[e]	24%
DOAC	—	—	—	—	—	[e]	—	—
VKA	50%	14%	—	—	—	[e]	18%	100%

Continued

Table 2 *Continued*

Year of publication **Risk score name**	**2013** **ATRIA**	**2016** **ABC stroke risk score**	**2017** **GARFIELD-AF**	**2019** **ACTS**	**2019** **IMRS-VASc**	**2019** **ABCD score**
Risk score type	Risk score	Risk score	Formula	Formula	Risk score	Risk score
Study design	Cohort (prospective)	RCT	Cohort (prospective)	Cohort (prospective)	Cohort (prospective)	Case–control
Country	USA	>10 countries	>10 countries	USA	USA	South Korea
Outcome	Composite (IS, systemic embolus)	Composite (IS, haemorrhagic stroke, systemic embolus)	Composite (IS, systemic embolus)	IS	Stroke (undefined)	Stroke (undefined)
Method	Cox	Cox	Cox	Cox	Cox	Logistic
Follow-up	3 years[a]	1.9 years[b]	—	1.8 years[a]	—	—
Prediction horizon	1 year	1 year	1 year	—	2 year	—
n events/*n* total (%)	685/10 927 (6.3)	391/27 929[c] (N.A.)	511/38 935 (1.3)	2028/252 904[c] (N.A.)	1506/55 970 (2.7)	583/– (–)
n cand. pred (EPV)	13 (53)	15 (26)	30 (17)	44 (46)	32 (47)	15 (39)
Internal validation	Bootstrapping	Bootstrapping	Cross-validation	Bootstrapping	—	Bootstrapping
Discrimination (*C*-statistic)	0.73 (0.71–0.75)	0.68 (0.65–0.71)	0.69 (0.67–0.71)[g]	0.68 (0.66–0.70)	0.70 (0.69, 0.73)	0.86 (0.84–0.88)
Calibration	Hosmer–Lemewhow	Observed vs. expected	Observed vs. expected	Hosmer–Lemeshow, observed/expected	—	Hosmer–Lemewhow
Mean age (SD)	[e]	70[b]	71[f] (63–78)	[e]	F: 73 (SD 13) M: 69 (SD 14)	61 (SD8)
Male sex (%)	57	64	45	60	53	73
Treatment (%)						
Antiplatelet	[e]	[e]	36%	[b]	20%	[e]
DOAC	—	50%	23%	33%	[b]	[e]
VKA	—	50%	42%	67%	10%	[e]

Continued

Table 2 *Continued*

Year of publication **Risk score name**	**2019** **Artificial intelligence (AI) model**	**2020** **CAS**	**2021** **GARFIELD-AF II**	**2021** **HELT-E_2S_2**	**2021** **LRAF**
Risk score type	Artificial intelligence	Risk score	Formula	Risk score	Risk score
Study design	Cohort (prospective)	Cohort (prospective)	Cohort (prospective)	Cohort (prospective)	Cohort (retrospective)
Country	>10 countries	China	>10 countries	Japan	Israel
Outcome	Composite (IS, TIA, systemic embolism)	Composite (IS, systemic embolism)	Composite (IS, TIA, undefined stroke, systemic embolism)	IS	IS
Method	Neural network	Cox	Cox	Cox	Logistic
Follow-up	—	—	—	1.8 years[a]	5.5 years[a]
Prediction horizon	30–365 days	1 year	2 years	[b]	1 year
n events/*n* total (%)	[b]	163/6601 (2.5%)	957/52 032 (1.8%)	241/12 289 (2.0%)	304/15 621 (2.0%)
n cand. pred (EPV)	[b]	3 (54)	11 (87)	6 (40)	6 (51)
Internal validation	Bootstrapping	Bootstrapping	Cross-validation	Bootstrapping, cross-validation	—
Discrimination (*C*-statistic)	0.70 (0.56–0.83)	0.69 (0.65–0.73)	0.68 (0.66–0.70)	0.68 (0.65–0.71)[h]	0.65 (0.63–0.68)
Calibration	—	—	Observed vs. expected	Hosmer Lemeshow	—
Mean age (SD)	72[a]	67[a]	71[a]	70[a]	54
Male sex (%)	55	58	56	69	57
Treatment (%)					
Antiplatelet	[b]	78%	21%	[b]	37%
DOAC	[b]	—	28%	10%	2%
VKA	100%	—	39%	64%	23%

DOAC, direct oral anticoagulant; EPV, events per variable; NA, not applicable; RCT, randomized controlled trial; SD, standard deviation; USA, United States of America; UK, United Kingdom; VKA, vitamin K antagonists.
[a]Mean.
[b]No information.
[c]Person years.
[d]*C*-statistic per subgroup: CHA_2DS_2-VASc: 0.61 (0.51–0.70) in AF cohort on OAC, 0.58 (0.44–0.73) in AF cohort not on OAC.
[e]No information.
[f]Not stated.
[g]*C*-statistic per subgroup: GARFIELD-AF: 0.67 (0.64–0.71) in AF cohort on OAC, 0.69 (0.65–0.72) in AF cohort not on OAC.
[h]*C*-statistic per subgroup: HELT-E_2S_2: 0.70 (0.65–0.76) in AF cohort on OAC, 0.69 (0.64–0.73) in AF cohort not on OAC.

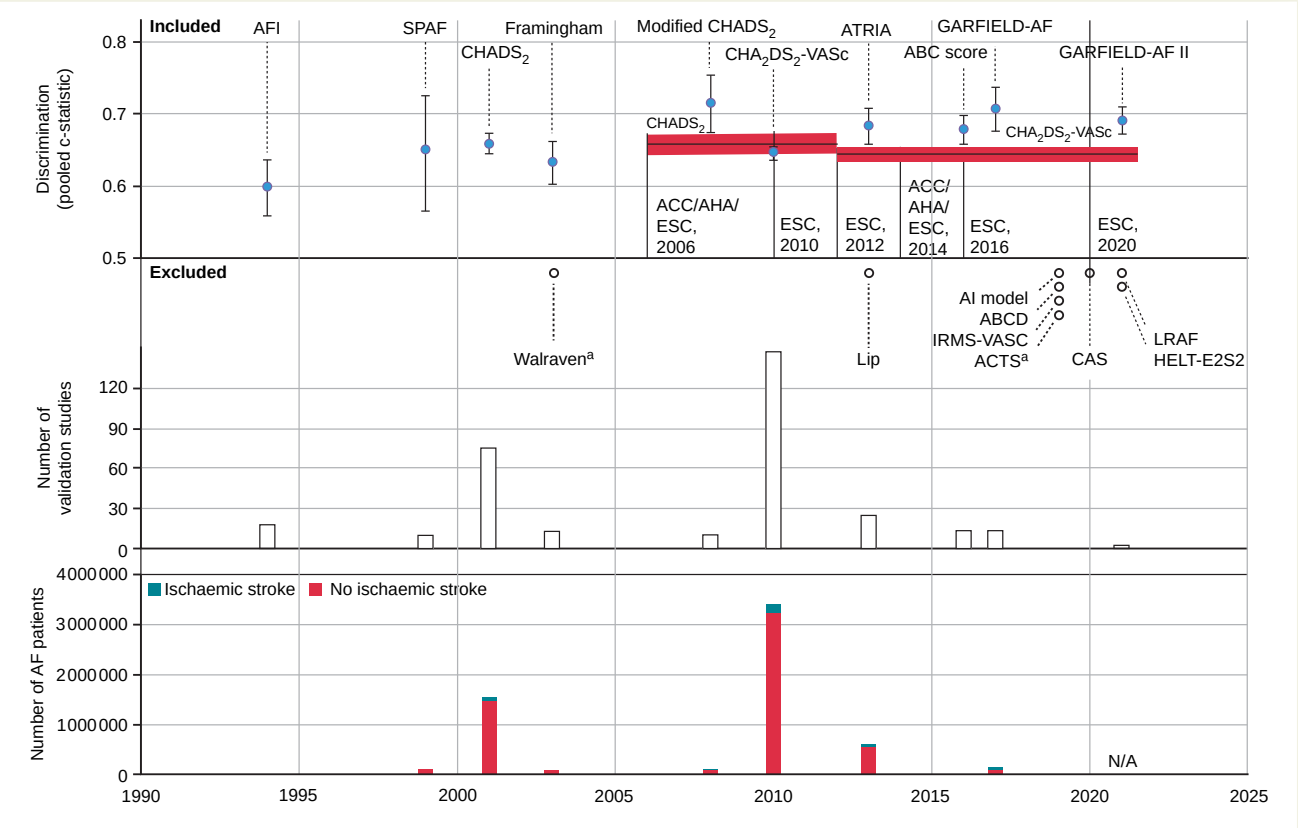

Figure 4 Discriminative performances over time. In this figure, the risk scores' years of publication (*x*-axis) are plotted against the number of AF patients (*y*-axis 1, lower panel), the number of validation studies (*y*-axis 2, middle panel), and results of the pooled *c*-statistic (*y*-axis 3, upper panel). The horizontal lines and the marked red area in the upper panel indicate the c-statistic and 95% confidence interval of the $CHADS_2$ (first horizontal line) and CHA_2DS_2-VASc (second line) during the perioded these scores were endorsed by the relevant guidelines (resp 2006 to 2012 and 2012 to present). The pooled *c*-statistics were generated from the random-effect meta-analysis. Ten risk scores were included in the random-effects analysis and are therefore displayed in the figure. [a]Walraven and ACTS were externally validated, yet not included in the analysis as the yielded discrimination measures (*c*-statistic) did not include confidence intervals, which were required for the random-effects meta-analysis (more details in the Statistical analysis section). AF, atrial fibrillation.

from more than a half to more than 3 million patients. Of these three, the ATRIA performed superiorly compared with the $CHADS_2$, and CHA_2DS_2-VASc. Nevertheless, the differences between the pooled *c*-statistics were marginal and all corresponded with poor to reasonable model performance.

Sensitivity analyses

In the 10 risk scores that were used for the random-effects meta-analysis, multiple sensitivity analyses were conducted. The outcomes of most sensitivity analyses showed no, or only marginal differences, indicating that age, anticoagulation, ethnicity, PROBAST score, prespecified IS risk, and year of publication (respectively, sensitivity analyses 1, 3–6, and 8) have no or only limited effect on most risk scores' discriminative abilities. For sensitivity analysis 2 on the effect of different outcome definitions, the modified-$CHADS_2$ showed improved discriminative abilities for the outcome IS [0.749 (0.710–0.785)], when compared with the outcome of thromboembolism [0.626 (0.561–0.686)], indicating that the original pooled *c*-statistic of the modified-$CHADS_2$ [0.715 (0.674–0.754)] might be an underestimation of the scores true performance on predicting IS only. In analysis 7 on study design, overall observational studies performed marginally better [e.g. for the CHA_2DS_2-VASc, the *c*-statistic was 0.649 (95% CI 0.639–0.658) for observational studies and 0.619 (0.603–0.634) for RCTs]. The number of validation studies in RCTs, however, was substantially lower than for observational studies. Consequently, we expect the effect of the inclusion of RCTs in our pooled calculations to be minimal. More detailed information on the sensitivity analyses can be found in the Supplementary material online, section 'Sensitivity analyses'.

Discussion

Summary

In this systematic review and meta-analysis, we reviewed risk scores predicting IS in AF patients using data of over 6 million individuals. We identified 19 original scores predicting IS in AF patients, which were validated a total of 327 times, and updated 76 times—nearly all on either the CHA_2DS_2-VASc or $CHADS_2$. Of these 19 scores, 10 were included in our meta-analysis on their discriminatory abilities

Table 3 Results of the random-effects meta-analysis: pooled *c*-statistic ranging from largest to smallest validation cohort

Risk score characteristics	Development					Validation			
	Original sample size	*N* of events	Original *c*-statistic	Model type	Time frame	*N* of studies[b] (*N* of validations)	Sample size	*N* of events	Pooled *c*-statistic
CHA_2DS_2-VASc	1084	25	0.61	Risk score	1 year	82 ($n=135$)	3 229 267	169 199	0.644 (0.635–0.653)
$CHADS_2$	1733	94	0.82	Risk score	1 year	46 ($n=68$)	1 479 228	71 644	0.658 (0.644–0.672)
ATRIA	10 927	685	0.73	Risk score	1 year	11 ($n=24$)	562 443	45 444	0.683 (0.658–0.708)
AFI	5955	208	—	Risk score	1 year	7 ($n=12$)	153 530	6879	0.598 (0.558–0.636)
GARFIELD-AF	38 935	511	0.69	Formula	1 year	4 ($n=13$)	149 848	5427	0.707 (0.676–0.737)
SPAF	2012	130	—	Risk score	1 year	5 ($n=5$)	116 864	3778	0.650 (0.564–0.726)
Modified-$CHADS_2$	51 807	5526	0.72	Risk score	5 years	5 ($n=10$)	109 313	10 083	0.715 (0.674–0.754)
Framingham	705	83	0.66	Risk score	5 years	6 ($n=8$)	95 145	2716	0.633 (0.602–0.662)
The ABC stroke risk score	27 929[a]	391	0.68	Risk score	1 year	5 ($n=11$)	40 340	1441	0.678 (0.658–0.697)
GARFIELD-AF II	52 032	957	0.68	Formula	NA	1 ($n=2$)	NA	NA	0.690 (0.672–0.707)

NA, data not available. Note that the number of validations slightly differs with *Figure* 3 since some validations needed to be excluded from the random effect meta-analysis due to studies omitting confidence intervals.
[a]Person years.
[b]Validation studies were included if discrimination measures (*c*-statistic) together with confidence intervals were available.

in external validations. Although all risk scores showed poor to only reasonable performance, the scores published after the CHA_2DS_2-VASc (2010) tended to have slightly better discriminatory abilities when compared with the scores published before the CHA_2DS_2-VASc, with the exception of the modified-$CHADS_2$ (published in 2008) which showed overall the best discriminative abilities. Update studies were found for the CHA_2DS_2-VASc and $CHADS_2$ only, showing improved discrimination. Information on calibration was omitted in nearly all studies developing, validating or updating a risk score. We observed methodological biases in all studies.

Clinical implication

By presenting a comprehensive comparison of the pooled discriminative performances of 19 risk scores, our study may support the choice of the best discriminative risk score in clinical or research settings. Faced with a patient diagnosed with new-onset AF, the clinician using such a risk score will more correctly classify the patient as low, intermediate, or high risk for IS.[5–7] With this knowledge, one can argue that the clinician might consider the use of other risk scores than the CHA_2DS_2-VASc, for example, the modified-$CHADS_2$ or one of the newer risk scores instead to classify his or her patients. However, though differences were found on discriminatory abilities, conclusions on the risk score's overall predictive performances should be taken cautiously. First, differences in discriminative abilities were only marginal and all corresponded with poor to only reasonable score performance. Secondly, not only discrimination but also calibration (i.e. the agreement between the predicted and the observed risk) is a key element in risk score assessment.[15,70] Yet, nearly all studies developing, validating or updating a risk score omitted information on this essential aspect of predictive modelling. Without knowing whether a risk score over- or underpredicts the observed risk of IS, misclassification and thus over- or undertreatment may be a serious issue but unbeknownst to the clinician in the absence of reliable data on calibration.[15,70] Finally, it should be noted that these promising newer risk scores updates have not, or only limited, been externally validated.

Comparison to literature

Our findings confirm the modest overall discriminatory abilities of commonly used risk scores regarding IS in AF patients.[8–10] This finding may be attributed to multiple factors, such as differences in study designs, measurement methods, and study case mix.[117,118] For example, the $CHADS_2$ is developed in a heterogeneous population of patients with AF and shows good discriminative abilities in the development cohort,[46] in contrast to moderate discrimination in the validation cohorts. When validating the $CHADS_2$ in less heterogeneous but clinically relevant populations, for example, in patients with AF and impaired kidney function, risk score performance drops.[70, 95,115,119] While the components of the risk score may predict well in heterogeneous general AF cohorts, other risk factors more specific for this homogeneous high-risk population may improve the discriminative abilities.[11,15,20,70] Indeed, supporting this mechanism, probably due to the inclusion of more specific predictors, a slight improvement in discriminative ability was found in the newer risk scores and updates

Strengths and limitations

Our study comes with strengths and limitations. The main strength regards the assessment of the predictive abilities of these risk scores by performing a random-effect meta-analysis including a well-

powered cohort of more than 6 million patients, yielding robust pooled *c*-statistics for the 10 most commonly validated risk scores. The study has, however, several limitations. First, no random-effects meta-analysis was conducted on calibration, due to the limited number of studies assessing calibration and the heterogeneity of these calibration measures. As a consequence, conclusions on overall predictive performances should be taken with caution as not only discrimination but also calibration is essential in risk score assessment. Future development and validation studies should include assessments on calibration to enable pooled calibration measures. Secondly, all studies were indicative for methodological bias based on the PROBAST, questioning the reliability of the presented discriminative abilities.[118] High risk of bias and concerns regarding applicability were mostly observed in the outcome and analysis domain and is common in prediction research, regardless of publication year.[120] Due to the similarity in the PROBAST scores, we performed a subgroup analysis based on the median PROBAST score to gain more insight in the effect of high vs. low risk of bias. Though no substantial effects were found, our risk of bias assessment underlines the work that needs to be done in the outcome and analysis domains. In future studies, adhering to the PROBAST[20] and TRIPOD[13] guidelines is warranted. Thirdly, a substantial part of the studies based the risk scores' discriminative abilities on the prediction of composite outcomes, instead of IS prediction only. The use of composite outcomes is debated, especially if the outcomes are contradictive to each other and need different treatment strategies, for example, IS and haemorrhagic stroke. In our study, however, outcomes predominantly regarded non-contradictive composite outcomes (e.g. IS together with TIA or systemic embolism) which may be defendable from the perspective of a clinician. Moreover, we studied the effect of composite outcome usage by means of a sensitivity analysis: no substantial effects were found, indicating that the inclusion of composite outcomes did not, or only marginally, influence our study results. Next, most studies validated the risk scores in patients already on anticoagulation treatment, whereas in clinical practice the risk scores are used as a tool for therapy decision. Validating in patients already using anticoagulation might have led to bias: high risk patients are more likely to receive anticoagulation, and thus paradoxically, a reduced risk of developing IS may be observed in them, leading to suboptimal performance of stroke risk scores.[121,122] To assess whether this treatment paradox was present, we performed a sensitivity analysis on anticoagulation use, which showed only marginal differences. Another limitation regards the exclusion of studies in the random-effect meta-analysis when no confidence interval was presented next to the *c*-statistic, possibly leading to selection bias. Yet, due to the large number of studies that included confidence intervals ($n = 95$, 89%), we do not expect the exclusion of the small subset ($n = 12$, 11%) to have an influential impact on the average performances. Also, we excluded non-English studies, which might have had influence on the limited number of studies that were included regarding ethnicities other than Caucasian. Finally, we limited our study to IS risk prediction and did not focus on bleeding risk prediction. It should be noted that for the decision-making of anticoagulation therapy in AF patients, bleeding risk scores and their predictive performances are of importance as well.

Conclusion

We identified 19 primary risk scores regarding IS risk in patients with AF along with 327 validations and 76 updates, mostly on the $CHADS_2$ and CHA_2DS_2-VASc. All risk scores showed largely similar, poor to reasonable, and discriminative performance; information on calibration was not reported for most studies. Compared with the $CHADS_2$ and CHA_2DS_2-VASc, newer risk scores and updates showed improved discrimination and might therefore be considered for use in clinical practice. To confirm this positive trend, external validations to assess discrimination but especially calibration of these newer risk scores are needed.

Supplementary material

Supplementary material is available at *Europace* online.

Funding

The work on this study by J.M. and M.v.D. was supported by a grant from the Dutch Kidney Foundation (20Ok016).

Conflicts of interest: None declared.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

1. Kannel WB, Wolf PA, Benjamin EJ, Levy D. Prevalence, incidence, prognosis, and predisposing conditions for atrial fibrillation: population-based estimates. *Am J Cardiol* 1998;**82**:2N–9N.
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. Lip G, Freedman B, De Caterina R, Potpara TS. Stroke prevention in atrial fibrillation: past, present and future. Comparing the guidelines and practical decision-making. *Thromb Haemost* 2017;**117**:1230–9.
4. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;**97**:1837–47.
5. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, *et al.* 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association of Cardio-Thoracic Surgery (EACTS). *Eur Heart J* 2021;**42**:373–498.
6. Lip GYH, Banerjee A, Boriani G, Chiang CE, Fargo R, Freedman B, *et al.* Antithrombotic therapy for atrial fibrillation: CHEST Guideline and Expert Panel Report. *Chest* 2018;**154**:1121–201.
7. January CT, Wann LS, Calkins H, Chen LY, Cigarroa JE, Cleveland J, *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.
8. Van Staa TP, Setakis E, Di Tanna GL, Lane DA, Lip GY. A comparison of risk stratification schemes for stroke in 79,884 atrial fibrillation patients in general practice. *J Thromb Haemost* 2011;**9**:39–48.
9. Borre ED, Goode A, Raitz G, Shah B, Lowenstern A, Chatterjee R, *et al.* Predicting thromboembolic and bleeding event risk in patients with non-valvular atrial fibrillation: a systematic review. *Thromb Haemost* 2018;**118**:2171–87.
10. Fang MC, Go AS, Chang Y, Borowsky L, Pomernacki NK, Singer DE. Comparison of risk stratification schemes to predict thromboembolism in people with nonvalvular atrial fibrillation. *J Am Coll Cardiol* 2008;**51**:810–5.
11. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* 2021;**14**:49–58.
12. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;**339**:b2700.

13. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 2015;**13**:1.
14. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, *et al.* Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014;**11**:e1001744.
15. de Jong Y, Ramspek CL, van der Endt VHW, Rookmaaker MB, Blankestijn PJ, Vernooij RWM, *et al.* A systematic review and external validation of stroke prediction models demonstrates poor performance in dialysis patients. *J Clin Epidemiol* 2020;**123**:69–79.
16. Ramspek CL, de Jong Y, Dekker FW, van Diepen M. Towards the best kidney failure prediction tool: a systematic review and selection aid. *Nephrol Dial Transplant* 2020;**35**:1527–38.
17. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, *et al.* Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;**98**:691–8.
18. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, *et al.* Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA* 2017;**318**:1377–84.
19. Pencina MJ, D'Agostino R, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;**27**:157–72.
20. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, *et al.* PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med* 2019;**170**:W1–33.
21. Snell KI, Ensor J, Debray TP, Moons KG, Riley RD. Meta-analysis of prediction model performance across multiple studies: which scale helps ensure between-study normality for the C-statistic and calibration measures? *Stat Methods Med Res* 2018;**27**:3505–22.
22. Hartung J, Knapp G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Stat Med* 2001;**20**:1771–82.
23. Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. *Stat Med* 2002;**21**:3153–9.
24. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol* 2014;**14**:25.
25. Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, Hooft L, *et al.* A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Stat Methods Med Res* 2019;**28**:2768–86.
26. Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, *et al.* A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;**356**:i6460.
27. Mavridis D, Salanti G. How to assess publication bias: funnel plot, trim-and-fill method and selection models. *Evid Based Ment Health* 2014;**17**:30.
28. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**:629–34.
29. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Soft* 2010;**36**:1–48.
30. Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Analysis of pooled data from five randomized controlled trials. *Arch Intern Med* 1994;**154**:1449–57.
31. Jiang C, Chen TG, Du X, Li X, He L, Lai YW, *et al.* A simple and easily implemented risk model to predict 1-year ischemic stroke and systemic embolism in Chinese patients with atrial fibrillation. *Chin Med J (Engl)* 2021;**134**:2293–8.
32. Fox KAA, Virdone S, Pieper KS, Bassand JP, Camm AJ, Fitzmaurice DA, *et al.* GARFIELD-AF risk score for mortality, stroke and bleeding within 2 years in patients with atrial fibrillation. *Eur Heart J Qual Care Clin Outcomes* 2022;**8**:214–27.
33. Okumura K, Tomita H, Nakai M, Kodani E, Akao M, Suzuki S, *et al.* A novel risk stratification system for ischemic stroke in Japanese patients with non-valvular atrial fibrillation. *Circ J* 2021;**85**:1254–62.
34. Arnson Y, Senderey AB, Hoshen M, Reges O, Balicer R, Alnsasra H, *et al.* Identifying patients with atrial fibrillation with a single CHA(2)DS(2)-VASC risk factor who are at higher risk of stroke. *Ir J Med Sci* 2022;**191**:705–11.
35. Hart RG, Pearce LA, McBride R, Rothbart RM, Asinger RW. Factors associated with ischemic stroke during aspirin therapy in atrial fibrillation: analysis of 2012 participants in the SPAF I-III clinical trials. The Stroke Prevention in Atrial Fibrillation (SPAF) Investigators. *Stroke* 1999;**30**:1223–9.
36. van Walraven C, Hart RG, Wells GA, Petersen P, Koudstaal PJ, Gullov AL, *et al.* A clinical prediction rule to identify patients with atrial fibrillation and a low risk for stroke while taking aspirin. *Arch Intern Med* 2003;**163**:936–43.
37. Wang TJ, Massaro JM, Levy D, Vasan RS, Wolf PA, D'Agostino RB, *et al.* A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: the Framingham Heart Study. *JAMA* 2003;**290**:1049–56.
38. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest* 2010;**137**:263–72.
39. Lip GYH, Lane DA, Buller H, Apostolakis S. Development of a novel composite stroke and bleeding risk score in patients with atrial fibrillation: the AMADEUS Study. *Chest* 2013;**144**:1839–47.
40. Singer DE, Chang Y, Borowsky LH, Fang MC, Pomernacki NK, Udaltsova N, *et al.* A new risk scheme to predict ischemic stroke and other thromboembolism in atrial fibrillation: the ATRIA study stroke risk score. *J Am Heart Assoc* 2013;**2**:e000250.
41. Fox KAA, Lucas JE, Pieper KS, Bassand JP, Camm AJ, Fitzmaurice DA, *et al.* Improved risk stratification of patients with atrial fibrillation: an integrated GARFIELD-AF tool for the prediction of mortality, stroke and bleed in patients with and without anticoagulation. *BMJ Open* 2017;**7**:e017157.
42. Claxton JS, MacLehose RF, Lutsey PL, Norby FL, Chen LY, O'Neal WT, *et al.* A new model to predict ischemic stroke in patients with atrial fibrillation using warfarin or direct oral anticoagulants. *Heart Rhythm* 2019;**16**:820–6.
43. Rietbrock S, Heeley E, Plumb J, van Staa T. Chronic atrial fibrillation: incidence, prevalence, and prediction of stroke using the congestive heart failure, hypertension, age >75, diabetes mellitus, and prior stroke or transient ischemic attack (CHADS2) risk stratification scheme. *Am Heart J* 2008;**156**:57–64.
44. Hijazi Z, Lindback J, Alexander JH, Hanna M, Held C, Hylek EM, *et al.* The ABC (age, biomarkers, clinical history) stroke risk score: a biomarker-based risk score for predicting stroke in atrial fibrillation. *Eur Heart J* 2016;**37**:1582–90.
45. Shin SY, Han SJ, Kim JS, Im SI, Shim J, Ahn J, *et al.* Identification of markers associated with development of stroke in "Clinically Low-Risk" atrial fibrillation patients. *J Am Heart Assoc* 2019;**8**:e012697.
46. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *JAMA* 2001;**285**:2864–70.
47. Horne BD, Jacobs V, May HT, Graves KG, Bunch TJ. Augmented intelligence decision tool for stroke prediction combines factors from CHA2 DS2-VASc and the intermountain risk score for patients with atrial fibrillation. *J Cardiovasc Electrophysiol* 2019;**30**:1452–61.
48. Goto S, Goto S, Pieper KS, Bassand JP, Camm AJ, Fitzmaurice DA, *et al.* New artificial intelligence prediction model using serial prothrombin time international normalized ratio measurements in atrial fibrillation patients on vitamin K antagonists: GARFIELD-AF. *Eur Heart J Cardiovasc Pharmacother* 2020;**6**:301–9.
49. Berg DD, Ruff CT, Jarolim P, Giugliano RP, Nordio F, Lanz HJ, *et al.* Performance of the ABC scores for assessing the risk of stroke or systemic embolism and bleeding in patients with atrial fibrillation in ENGAGE AF-TIMI 48. *Circulation* 2019;**139**:760–71.
50. Kim YG, Choi JI, Boo KY, Kim DY, Hong Y, Kim MS, *et al.* Impact of age on thromboembolic events in patients with non-valvular atrial fibrillation. *Clin Cardiol* 2020;**43**:78–85.
51. Larsen TB, Lip GY, Skjøth F, Due KM, Overvad K, Hvilsted Rasmussen L. Added predictive ability of the CHA_2DS_2VASc risk score for stroke and death in patients with atrial fibrillation: the prospective Danish Diet, Cancer, and Health cohort study. *Circ Cardiovasc Qual Outcomes* 2012;**5**:335–42.
52. Li YG, Miyazawa K, Wolff A, Zubaid M, Alsheikh-Ali AA, Sulaiman K, *et al.* One-year risks of stroke and mortality in patients with atrial fibrillation from different clinical settings: the Gulf SAFE registry and Darlington AF registry. *Int J Cardiol* 2019;**274**:158–62.
53. Lin LY, Lee CH, Yu CC, Tsai CT, Lai LP, Hwang JJ, *et al.* Risk factors and incidence of ischemic stroke in Taiwanese with nonvalvular atrial fibrillation– a nation wide database analysis. *Atherosclerosis* 2011;**217**:292–5.
54. Olesen JB, Lip GY, Hansen ML, Hansen PR, Tolstrup JS, Lindhardsen J, *et al.* Validation of risk stratification schemes for predicting stroke and thromboembolism in patients with atrial fibrillation: nationwide cohort study. *BMJ* 2011;**342**:d124.
55. Reps JM, Williams RD, You SC, Falconer T, Minty E, Callahan A, *et al.* Feasibility and evaluation of a large-scale external validation approach for patient-level prediction in an international data network: validation of models predicting stroke in female patients newly diagnosed with atrial fibrillation. *BMC Med Res Methodol* 2020;**20**:102.
56. Wang SV, Franklin JM, Glynn RJ, Schneeweiss S, Eddings W, Gagne JJ. Prediction of rates of thromboembolic and major bleeding outcomes with dabigatran or warfarin among patients with atrial fibrillation: new initiator cohort study. *BMJ* 2016;**353**:i2607.
57. Friberg LL LH. Heart failure: a weak link in CHA(2) DS(2)-VASc. *ESC Heart Fail* 2018;**5**:231–9.
58. Graves KG, May HT, Knowlton KU, Muhlestein JB, Jacobs V, Lappé DL, *et al.* Improving CHA(2)DS(2)-VASc stratification of non-fatal stroke and mortality risk using the Intermountain Mortality Risk Score among patients with atrial fibrillation. *Open Heart* 2018;**5**:e000907.
59. Ruff CT, Giugliano RP, Braunwald E, Murphy SA, Brown K, Jarolim P, *et al.* Cardiovascular biomarker score and clinical outcomes in patients with atrial

fibrillation: a subanalysis of the ENGAGE AF-TIMI 48 randomized clinical trial. *JAMA Cardiol* 2016;**1**:999–1006.
60. Chao TF, Lin YJ, Chang SL, Lo LW, Hu YF, Tuan TC, *et al.* R2CHADS2 score and thromboembolic events after catheter ablation of atrial fibrillation in comparison with the CHA2DS2-VASc score. *Can J Cardiol* 2014;**30**:405–12.
61. Fauchier L, Lecoq C, Ancedy Y, Stamboul K, Saint Etienne C, Ivanes F, *et al.* Evaluation of 5 prognostic scores for prediction of stroke, thromboembolic and coronary events, all-cause mortality, and major adverse cardiac events in patients with atrial fibrillation and coronary stenting. *Am J Cardiol* 2016;**118**:700–7.
62. García-Fernández A, Marín F, Roldán V, Gómez-Sansano JM, Hernández-Romero D, Valdés M, *et al.* Long-term predictors of thromboembolic events in nonvalvular atrial fibrillation patients undergoing electrical cardioversion. *Circ J* 2016;**80**: 605–12.
63. Kornej J, Hindricks G, Kosiuk J, Arya A, Sommer P, Husser D, *et al.* Renal dysfunction, stroke risk scores (CHADS(2), CHA(2)DS(2)-VASc, and R(2)CHADS(2)), and the risk of thromboembolic events after catheter ablation of atrial fibrillation: the Leipzig Heart Center AF Ablation Registry. *Circ Arrhythm Electrophysiol* 2013;**6**: 868–74.
64. Puurunen MK, Kiviniemi T, Schlitt A, Rubboli A, Dietrich B, Karjalainen P, *et al.* $CHADS_2$, CHA_2DS_2-VASc and HAS-BLED as predictors of outcome in patients with atrial fibrillation undergoing percutaneous coronary intervention. *Thromb Res* 2014;**133**:560–6.
65. Cho DH, Choi JI, Choi J, Kim YG, Oh SK, Kook H, *et al.* Impact of carotid atherosclerosis in CHA2DS2-VASc-based risk score on predicting ischemic stroke in patients with atrial fibrillation. *Korean J Intern Med* 2021;**36**:342–51.
66. Chao TF, Liu CJ, Wang KL, Lin YJ, Chang SL, Lo LW, *et al.* Incidence and prediction of ischemic stroke among atrial fibrillation patients with end-stage renal disease requiring dialysis. *Heart Rhythm* 2014;**11**:1752–9.
67. Nielsen PB, Overvad TF, Andersen SD, Larsen TB, Skjøth F, Søgaard M, *et al.* Risk stratification for ischemic cerebrovascular events and mortality among intracerebral hemorrhage patients with and without atrial fibrillation: a nationwide cohort study. *Cerebrovasc Dis* 2019;**48**:236–43.
68. Shih CJ, Ou SM, Chao PW, Kuo SC, Lee YJ, Yang CY, *et al.* Risks of death and stroke in patients undergoing hemodialysis with new-onset atrial fibrillation: a competing-risk analysis of a nationwide cohort. *Circulation* 2016;**133**:265–72.
69. Bel-Ange A, Itskovich SZ, Avivi L, Stav K, Efrati S, Beberashvili I. Prior ischemic strokes are non-inferior for predicting future ischemic strokes than CHA(2) DS(2)-VASc score in hemodialysis patients with non-valvular atrial fibrillation. *BMC Nephrol* 2021;**22**:179.
70. de Jong Y, Fu EL, van Diepen M, Trevisan M, Szummer K, Dekker FW, *et al.* Validation of risk scores for ischaemic stroke in atrial fibrillation across the spectrum of kidney function. *Eur Heart J* 2021;**42**:1476–85.
71. Benz AP, Healey JS, Chin A, Commerford P, Marsden T, Karthikeyan G, *et al.* Stroke risk prediction in patients with atrial fibrillation with and without rheumatic heart disease. *Cardiovasc Res* 2022;**118**:295–304.
72. Pastori D, Marang A, Bisson A, Menichelli D, Herbert J, Lip GYH, *et al.* Thromboembolism, mortality, and bleeding in 2,435,541 atrial fibrillation patients with and without cancer: a nationwide cohort study. *Cancer* 2021;**127**:2122–9.
73. Lip GYH, Jensen M, Melgaard L, Skjøth F, Nielsen PB, Larsen TB. Stroke and bleeding risk scores in patients with atrial fibrillation and valvular heart disease: evaluating 'valvular heart disease' in a nationwide cohort study. *Europace* 2019;**21**:33–40.
74. Saliba WB-G O, Elias M, Rennert G. The association between red cell distribution width and stroke in patients with atrial fibrillation. *Am J Med* 2015;**128**:192.e11–8.
75. Chao TF, Liu CJ, Tuan TC, Chen SJ, Wang KL, Lin YJ, *et al.* Comparisons of $CHADS_2$ and CHA_2DS_2-VASc scores for stroke risk stratification in atrial fibrillation: which scoring system should be used for Asians? *Heart Rhythm* 2016;**13**: 46–53.
76. Chao TF, Liu CJ, Wang KL, Lin YJ, Chang SL, Lo LW, *et al.* Using the CHA2DS2-VASc score for refining stroke risk stratification in 'low-risk' Asian patients with atrial fibrillation. *J Am Coll Cardiol* 2014;**64**:1658–65.
77. Guo Y, Apostolakis S, Blann AD, Wang H, Zhao X, Zhang Y, *et al.* Validation of contemporary stroke and bleeding risk stratification scores in non-anticoagulated Chinese patients with atrial fibrillation. *Int J Cardiol* 2013;**168**:904–9.
78. Kang SH, Choi EK, Han KD, Lee SR, Lim WH, Cha MJ, *et al.* Risk of ischemic stroke in patients with non-valvular atrial fibrillation not receiving oral anticoagulants-Korean nationwide population-based study. *Circ J* 2017;**81**:1158–64.
79. Cha MJC, Cho Y, Oh IY, Choi EK, Oh S. Validation of conventional thromboembolic risk factors in a Korean atrial fibrillation population - suggestion for a novel scoring system, CHA_2DS_2-VAK. *Circ J* 2018;**82**:2970–5.
80. Chao TF, Lip GYH, Liu CJ, Tuan TC, Chen SJ, Wang KL, *et al.* Validation of a modified CHA2DS2-VASc score for stroke risk stratification in Asian patients with atrial fibrillation: a nationwide cohort study. *Stroke* 2016;**47**:2462–9.
81. Chao TF, Lip GYH, Liu CJ, Lin YJ, Chang SL, Lo LW, *et al.* Relationship of aging and incident comorbidities to stroke risk in patients with atrial fibrillation. *J Am Coll Cardiol* 2018;**71**:122–32.
82. Nakamura M, Koeda Y, Tanaka F, Onoda T, Itai K, Ohsawa M, *et al.* Plasma B-type natriuretic peptide as a predictor of cardiovascular events in subjects with atrial fibrillation: a community-based study. *PLoS ONE* 2013;**8**:e81243.
83. Sadanaga T, Kohsaka S, Ogawa S. D-dimer levels in combination with clinical risk factors can effectively predict subsequent thromboembolic events in patients with atrial fibrillation during oral anticoagulant therapy. *Cardiology* 2010;**117**:31–6.
84. Saito Y, Okumura Y, Nagashima K, Fukamachi D, Yokoyama K, Matsumoto N, *et al.* Impact of the fibrosis-4 index on risk stratification of cardiovascular events and mortality in patients with atrial fibrillation: findings from a Japanese multicenter registry. *J Clin Med* 2020;**9**:584.
85. Tomita H, Okumura K, Inoue H, Atarashi H, Yamashita T, Origasa H, *et al.* Validation of risk scoring system excluding female sex from CHA2DS2-VASc in Japanese patients with nonvalvular atrial fibrillation—subanalysis of the J-RHYTHM registry. *Circ J* 2015;**79**:1719–26.
86. Tsai CT, Chang SH, Chang SN, Hwang JJ, Wu CK, Wang YC, *et al.* Additive effect of the metabolic syndrome score to the conventional $CHADS_2$ score for the thromboembolic risk stratification of patients with atrial fibrillation. *Heart Rhythm* 2014;**11**:352–7.
87. Tsuda T, Hayashi K, Fujino N, Konno T, Tada H, Nomura A, *et al.* Effect of hypertrophic cardiomyopathy on the prediction of thromboembolism in patients with nonvalvular atrial fibrillation. *Heart Rhythm* 2019;**16**:829–37.
88. Dalgaard F, Pieper K, Verheugt F, Camm AJ, Fox KA, Kakkar AK, *et al.* GARFIELD-AF model for prediction of stroke and major bleeding in atrial fibrillation: a Danish nationwide validation study. *BMJ Open* 2019;**9**:e033283.
89. Hijazi Z, Lindahl B, Oldgren J, Andersson U, Lindbäck J, Granger CB, *et al.* Repeated measurements of cardiac biomarkers in atrial fibrillation and validation of the ABC stroke score over time. *J Am Heart Assoc* 2017;**6**:e004851.
90. Oldgren J, Hijazi Z, Lindbäck J, Alexander JH, Connolly SJ, Eikelboom JW, *et al.* Performance and validation of a novel biomarker-based stroke risk score for atrial fibrillation. *Circulation* 2016;**134**:1697–707.
91. Benz AP, Hijazi Z, Lindbäck J, Connolly SJ, Eikelboom JW, Oldgren J, *et al.* Biomarker-based risk prediction with the ABC-AF scores in patients with atrial fibrillation not receiving oral anticoagulation. *Circulation* 2021;**143**:1863–73.
92. Maheshwari A, Norby FL, Roetker NS, Soliman EZ, Koene RJ, Rooney MR, *et al.* Refining prediction of atrial fibrillation-related stroke using the P(2)-CHA(2) DS(2)-VASc score. *Circulation* 2019;**139**:180–91.
93. Banerjee AF L, Bernard-Brunet A, Clementy N, Lip GY. Composite risk scores and composite endpoints in the risk prediction of outcomes in anticoagulated patients with atrial fibrillation. The Loire Valley Atrial Fibrillation Project. *Thromb Haemost* 2014;**111**:549–56.
94. Fauchier LB A, Bisson A, Herbert J, Spiesser P, Clementy N, Babuty D, *et al.* Incident comorbidities, aging and the risk of stroke in 608,108 patients with atrial fibrillation: a nationwide analysis. *J Clin Med* 2020;**9**:1234.
95. Friberg L, Benson L, Lip GYH. Balancing stroke and bleeding risks in patients with atrial fibrillation and renal failure: the Swedish Atrial Fibrillation Cohort study. *Eur Heart J* 2015;**36**:297–306.
96. Kim MN, Kim SA, Choi JI, Park SM, Park SW, Kim YH, Shim WJ. Improvement of predictive value for thromboembolic risk by incorporating left atrial functional parameters in the CHADS2 and CHA2DS2-VASc scores. *Int Heart J* 2015;**56**: 286–92.
97. Rivera-Caravaca JM, Marín F, Vilchez JA, Gálvez J, Esteve-Pastor MA, *et al.* Refining stroke and bleeding prediction in atrial fibrillation by adding consecutive biomarkers to clinical risk scores. *Stroke* 2019;**50**:1372–9.
98. Roldán V, Arroyo AB, Salloum-Asfar S, Manzano-Fernández S, García-Barberá N, Marín F, *et al.* Prognostic role of MIR146A polymorphisms for cardiovascular events in atrial fibrillation. *Thromb Haemost* 2014;**112**:781–8.
99. Liao JN, Chao TF, Kuo JY, Sung KT, Tsai JP, Lo CI, *et al.* Global left atrial longitudinal strain using 3-beat method improves risk prediction of stroke over conventional echocardiography in atrial fibrillation. *Circ Cardiovasc Imaging* 2020;**13**:e010287.
100. Banerjee AF L, Vourc'h P, Andres CR, Taillandier S, Halimi JM, Lip GY. Renal impairment and ischemic stroke risk assessment in patients with atrial fibrillation: the Loire Valley Atrial Fibrillation Project. *J Am Coll Cardiol* 2013;**61**:2079–87.
101. Chao TF, Tsao HM, Ambrose K, Lin YJ, Lin WS, Chang SL, *et al.* Renal dysfunction and the risk of thromboembolic events in patients with atrial fibrillation after catheter ablation – the potential role beyond the CHA_2DS_2-VASc score. *Heart Rhythm* 2012;**9**:1755–60.
102. García-Fernández A, Roldán V, Rivera-Caravaca JM, Hernández-Romero D, Valdés M, Vicente V, *et al.* Does von Willebrand factor improve the predictive ability of current risk stratification scores in patients with atrial fibrillation? *Sci Rep* 2017;**7**: 41565.
103. He H, Guo J, Zhang A. The value of urine albumin in predicting thromboembolic events for patients with non-valvular atrial fibrillation. *Int J Cardiol* 2016;**221**: 827–30.
104. Hijazi Z, Wallentin L, Siegbahn A, Andersson U, Christersson C, Ezekowitz J, *et al.* N-terminal pro-B-type natriuretic peptide for risk assessment in patients with atrial

fibrillation: insights from the ARISTOTLE Trial (Apixaban for the Prevention of Stroke in Subjects With Atrial Fibrillation). *J Am Coll Cardiol* 2013;**61**:2274–84.
105. Hijazi Z, Wallentin L, Siegbahn A, Andersson U, Alexander JH, Atar D, *et al.* High-sensitivity troponin T and risk stratification in patients with atrial fibrillation during treatment with apixaban or warfarin. *J Am Coll Cardiol* 2014;**63**:52–61.
106. Lip GYH, Lane D, Van Walraven C, Hart RG. Additive role of plasma von Willebrand factor levels to clinical factors for risk stratification of patients with atrial fibrillation. *Stroke* 2006;**37**:2294–300.
107. Rivera-Caravaca JM, Roldán V, Romera M, Esteve-Pastor MA, Valdés M, Lip GYH, *et al.* Soluble fibrin monomer complex and prediction of cardiovascular events in atrial fibrillation: the observational Murcia Atrial Fibrillation project. *J Gen Intern Med* 2018;**33**:847–54.
108. Roldán V, Rivera-Caravaca JM, Shantsila A, García-Fernández A, Esteve-Pastor MA, Vilchez JA, *et al.* Enhancing the 'real world' prediction of cardiovascular events and major bleeding with the CHA(2)DS(2)-VASc and HAS-BLED scores using multiple biomarkers. *Ann Med* 2018;**50**:26–34.
109. Roldán V, Vílchez JA, Manzano-Fernández S, Jover E, Gálvez J, Puche CM, *et al.* Usefulness of N-terminal pro-B-type natriuretic Peptide levels for stroke risk prediction in anticoagulated patients with atrial fibrillation. *Stroke* 2014;**45**:696–701.
110. Saliba W, Barnett-Griness O, Elias M, Rennert G. Glycated hemoglobin and risk of first episode stroke in diabetic patients with atrial fibrillation: a cohort study. *Heart Rhythm* 2015;**12**:886–92.
111. Saliba WB-G O, Elias M, Rennert G. Neutrophil to lymphocyte ratio and risk of a first episode of stroke in patients with atrial fibrillation: a cohort study. *J Thromb Haemost* 2015;**13**:1971–9.
112. Soulat-Dufour L, Lang S, Etienney A, Ederhy S, Ancedy Y, Adavane S, *et al.* Correlation between left atrial spontaneous echocardiographic contrast and 5-year stroke/death in patients with non-valvular atrial fibrillation. *Arch Cardiovasc Dis* 2020;**113**:525–33.
113. Roldán V, Marín F, Díaz J, Gallego P, Jover E, Romera M, Manzano-Fernández S, *et al.* High sensitivity cardiac troponin T and interleukin-6 predict adverse cardiovascular events and mortality in anticoagulated patients with atrial fibrillation. *J Thromb Haemost* 2012;**10**:1500–7.
114. Ravvaz K, Weissert JA, Jahangir A, Ruff CT. Evaluating the effects of socioeconomic status on stroke and bleeding risk scores and clinical events in patients on oral anticoagulant for new onset atrial fibrillation. *PLoS ONE* 2021;**16**:e0248134.
115. O'Brien EC, Holmes DN, Thomas L, Singer DE, Fonarow GC, Mahaffey KW, *et al.* Incremental prognostic value of renal function for stroke prediction in atrial fibrillation. *Int J Cardiol* 2019;**274**:152–7.
116. Roldán V, Marín F, Manzano-Fernández S, Fernández H, Gallego P, Valdés M, *et al.* Does chronic kidney disease improve the predictive value of the CHADS2 and CHA2DS2-VASc stroke stratification risk scores for atrial fibrillation? *Thromb Haemost* 2013;**109**:956–60.
117. Dekkers OM, Mulder JM. When will individuals meet their personalized probabilities? A philosophical note on risk prediction. *Eur J Epidemiol* 2020;**35**:1115–21.
118. Damen J, Debray TPA, Pajouheshnia R, Reitsma JB, Scholten R, Moons KGM, *et al.* Empirical evidence of the impact of study characteristics on the performance of prediction models: a meta-epidemiological study. *BMJ Open* 2019;**9**:e026160.
119. Piccini JP, Stevens SR, Chang Y, Singer DE, Lokhnygina Y, Go AS, *et al.* Renal dysfunction as a predictor of stroke and systemic embolism in patients with nonvalvular atrial fibrillation: validation of the R(2)CHADS(2) index in the ROCKET AF (Rivaroxaban Once-daily, oral, direct factor Xa inhibition Compared with vitamin K antagonism for prevention of stroke and Embolism Trial in Atrial Fibrillation) and ATRIA (AnTicoagulation and Risk factors In Atrial fibrillation) study cohorts. *Circulation* 2013;**127**:224–32.
120. de Jong Y, Ramspek CL, Zoccali C, Jager KJ, Dekker FW, van Diepen M. Appraising prediction research: a guide and meta-review on bias and applicability assessment using the Prediction model Risk Of Bias ASsessment Tool (PROBAST). *Nephrology (Carlton)* 2021;**26**:939–47.
121. Groenwold RH, Moons KG, Pajouheshnia R, Altman DG, Collins GS, Debray TP, *et al.* Explicit inclusion of treatment in prognostic modeling was recommended in observational and randomized settings. *J Clin Epidemiol* 2016;**78**:90–100.
122. Pajouheshnia R, Peelen LM, Moons KGM, Reitsma JB, Groenwold RHH. Accounting for treatment use when validating a prognostic model: a simulation study. *BMC Med Res Methodol* 2017;**17**:103.