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 CHA2DS2-VASc and CHADS2. Pooled c-statistics were calculated among 6 267 728 patients and 359 373 events of IS. For the CHA2DS2-VASc and CHADS2, 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-CHADS2 demonstrating the best discrimination [c-statistic 0.715 (0.674–0.754)]. Updates were found for the CHA2DS2-VASc and CHADS2 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 CHA2DS2-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
© The Author(s) 2022. Published by Oxford University Press on behalf of the European Society of Cardiology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.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. Anticoagulation reduces this risk substantially and is therefore prescribed to most patients. 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. 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. Examples include the CHADS2 (2001) and the commonly used CHA2DS2-VASC (2010), with the latter being endorsed by most current clinical guidelines. Studies on the predictive performance suggest that current risk scores have comparable but limited overall ability to predict IS in patients with AF. 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. 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.

Methods

The current study is performed and reported in line with the PRISMA, TRIPOD, and CHARMS 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. 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. 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.), 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. For calibration (i.e. the agreement between the predicted and the observed risk), we extracted the calibration-in-the-large, 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. For the update studies, net reclassification index and the integrated discrimination improvement, measures quantifying the improvement when a score is updated, were extracted. 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.\textsuperscript{25}

**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.\textsuperscript{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.\textsuperscript{22–24} A logit pooled c-statistic was calculated for each of the included models and then transformed back to the original scale.\textsuperscript{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.\textsuperscript{22,28} All analyses were conducted using RStudio version 1.2.5033 and the metafor package.\textsuperscript{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 years, 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\textsubscript{2}-DS\textsubscript{2}-VASc/CHADS\textsubscript{2} score ≤1, ATRIA score ≤3) with intermediate- or high-risk patients (CHA\textsubscript{2}-DS\textsubscript{2}-VASc/CHADS\textsubscript{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-CHADS2 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...
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 years and 1.0–1.9 years in studies presenting a mean and median follow-up period, respectively. Except for four studies, all studies were conducted in a cohort of predominantly Caucasian patients. Six studies validated their score in a cohort of patients not on anticoagulation, whereas 12 studies 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, a decision tree, or employed artificial intelligence algorithms.46 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.46,48 The prediction horizon ranged between 31 days and 5 years and was not given for three scores.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 to 0.86.45 Calibration was presented as observed vs. expected risks,32,41,42,44,49,55,56,70–88,91 adjusted by the Hosmer–Lemeshow statistic.43,46,48,85,92 Detailed characteristics of the validation studies are presented in Supplementary material online, section ‘Characteristics of included validation 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 did not provide population size and number of events. Most of the validations were performed on the CHADS\(_2\)-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\(_2\) (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),12,44,66–74 or in cohorts with ethnicity other than predominantly Caucasian (n = 20).33,34,45,47,53,55,63,71,75–87 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 CHADS\(_2\)-VASc (2010), no discrimination lower than 0.60 was reported in the validation studies.32,40,43–45,67 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,43,46,48,85,92 calibration in the large,20 and the Nam–D’Agostino statistic.93 Except for this limited information on calibration, compared with the CHADS\(_2\)-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 CHADS\(_2\)-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 CHADS\(_2\)-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\(_3\) CHADS\(_2\)-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),79,74,82–84,86,92,97,98,100–111 echocardiographic characteristics (e.g. hypertrophic cardiomyopathy),65,67,99,112 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,107,108 or the original predictors were assessed over time (i.e. differences between the baseline and follow-up CHADS\(_2\)-VASc scores).81,94 For the CHADS\(_2\)-VASc, discrimination was available for 49 updates (89%): c-statistics improved for all but three updates when compared with the original CHADS\(_2\)-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 (CHADS\(_2\)-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 CHADS\(_2\)-VASc and CHADS\(_2\), respectively. All risk scores that were developed after the publication of the CHADS\(_2\)-VASc (2010) showed better discriminative abilities when compared with the CHADS\(_2\)-VASc (Figure 4). For the CHADS\(_2\)-VASc, CHADS\(_2\), and ATRIA, pooled c-statistics were derived from >10 studies per risk score including large sample sizes ranging
| Year of publication | Risk score name | Risk score type | Study design | Country | Outcome | Method | Follow-up | Prediction horizon | n events/n total (%) | n cand. pred (EPV) | Internal validation | Discrimination (C-statistic) | Calibration | Mean age (SD) | Male sex (%) | Treatment (%) |
|---------------------|----------------|----------------|--------------|---------|---------|--------|-----------|-------------------|---------------------|----------------|-------------------|-------------------|-------------|----------------|-------------|----------------|
| Year of publication | 2013 ATRIA | 2016 ABC stroke risk score | 2017 GARFIELD-AF | 2019 ACTS | 2019 IMRS-VASc | 2019 ABCD score |
|---------------------|-------------|-----------------------------|------------------|----------|----------------|----------------|
| Risk score name     | Risk score  | Risk score                  | Formula          | Formula  | Risk score     | Risk score     |
| Study design        | Cohort (prospective) | RCT                     | Cohort (prospective) | Cohort (prospective) | USA            | Case-control   |
| Country             | USA         | >10 countries               | USA              | 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<sup>a</sup> | 1.9 years<sup>b</sup> | —                | 1.8 years<sup>a</sup> | —             | —              |
| Prediction horizon  | 1 year      | 1 year                      | 1 year           | 2 year   | —              | —              |
| n events/n total (%)| 685/10 927 (6.3) | 391/27929<sup>c</sup> (N.A.) | 511/38 935 (1.3) | 2028/252 904<sup>c</sup> (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      | 0.73 (0.71–0.75) | 0.68 (0.65–0.71) | 0.69 (0.67–0.71)<sup>d</sup> | 0.68 (0.66–0.70) | 0.70 (0.69, 0.73) | 0.86 (0.84–0.88) |
| Calibration         | Hosmer–Lemeshow | Observed vs. expected       | Observed vs. expected | Hosmer–Lemeshow, observed/expected | —              | —              |
| Mean age (SD)       | F: 73 (SD 13) | M: 69 (SD 14)               | F: 73 (63–78)   | M: 60     | 53             | 73             |
| Male sex (%)        | 57          | 64                          | 45               | 60       | 53             | 73             |
| Treatment (%)       | Antiplatelet | DOAC                       | VKA              | 36%      | 33%            | 33%            |
|                     | —           | 50%                         | 50%              | b        | b              | b              |
|                     | Antiplatelet | DOAC                       | VKA              | 20%      | 10%            | 10%            |
|                     | —           | 50%                         | 50%              | —        | —              | —              |

Continued
| Year of publication | Risk score name | Formula | Study design | Country | Outcome | Method | Follow-up | Prediction horizon | Internal validation | Discrimination (C-statistic) | Calibration | Treatment (% | Male sex (%) | Mean age (SD) | Mean age (SD) | Antithrombotic | DOAC | VKA |
|---------------------|----------------|---------|--------------|---------|---------|--------|-----------|-------------------|---------------------|-----------------------|-------------|--------------|-------------|--------------|--------------|---------------|------|------|
| 2019                | Artificial intelligence model | Cox | Cohort (prospective) | China | Composite (S, TIA, systemic embolism) | Neural network | — | 30–365 days | Bootstrapping | 0.69 (0.65–0.73) | — | — | — | 72 | 72 | — | — | — |
| 2020                | CAS | Cox | Cohort (prospective) | Japan | Composite (S, TIA, undefined stroke, systemic embolism) | Cox | 1 year | 2 years | Cross-validation | 0.68 (0.66–0.70) | — | 74% | 59 | 67 | 67 | — | — | — |
| 2021                | GARFIELD-AF II | Logistic | Cohort (prospective) | Israel | Composite (S, TIA) | Cox | — | — | — | 0.69 (0.65–0.73) | — | 71 | 58 | 67 | 58 | 100% | — | — |
| 2021                | LIFA | Logistic | Cohort (prospective) | IS | Composite (S, TIA, undefined stroke, systemic embolism) | Cox | 1 year | 1 year | Bootstrapping | 0.69 (0.65–0.73) | — | 71 | 58 | 67 | 58 | 100% | — | — |

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.

- Mean.
- No information.
- Person years.
- C-statistic per subgroup: CHA2DS2-VASc: 0.61 (0.51–0.70) in AF cohort not on OAC, 0.68 (0.64–0.72) in AF cohort on OAC.
- C-statistic per subgroup: GARFIELD-AF: 0.67 (0.64–0.71) in AF cohort not on OAC, 0.69 (0.65–0.72) in AF cohort on OAC.
- C-statistic per subgroup: HELT-E: 0.65 (0.63–0.68) in AF cohort not on OAC, 0.66 (0.65–0.71) in AF cohort on OAC.
from more than a half to more than 3 million patients. Of these three, the ATRIA performed superiorly compared with the CHADS2 and CHA2DS2-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-CHADS2 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-CHADS2 [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 CHA2DS2-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 CHA2DS2-VASc or CHADS2. Of these 19 scores, 10 were included in our meta-analysis on their discriminatory abilities.
in external validations. Although all risk scores showed poor to only reasonable performance, the scores published after the CHA²DS²-VASc (2010) tended to have slightly better discriminatory abilities when compared with the scores published before the CHA²DS²-VASc, with the exception of the modified-CHADS² (published in 2008) which showed overall the best discriminative abilities. Update studies were found for the CHA²DS²-VASc and CHADS² 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 inative performances of 19 risk scores, our study may support the

### Comparison to literature

Our findings confirm the modest overall discriminatory abilities of commonly used risk scores regarding IS in AF patients. This finding may be attributed to multiple factors, such as differences in study designs, measurement methods, and study case mix. For example, the CHADS² is developed in a heterogeneous population of patients with AF and shows good discriminative abilities in the development cohort, in contrast to moderate discrimination in the validation cohorts. When validating the CHADS² in less heterogeneous but clinically relevant populations, for example, in patients with AF and impaired kidney function, risk score performance drops. 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. 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 and TRIPOD12 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 contradictory to each other and need different treatment strategies, for example, IS and haemorrhagic stroke. In our study, however, outcomes predominately regarded non-contradictory 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 as well. 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 CHADS2 and CHA2DS2-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 CHADS2 and CHA2DS2-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 1996;97:1837–47.
5. Hindricks G, Potpara T, Dagres N, Arbilo E, Bax JJ, Blomstrom-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, Benerjee A, Boriani G, Chang 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, Calhoun H, Chen LY, Giarrova 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:e123–51.
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, Raatz G, Shah B, Lowenstein A, Chatterjee R, et al. Predicting thromboembolic and bleeding event risk in patients with non-valvular atrial fibrillation: a systematic review. Thromb Haemost 2010;103:2711–87.
10. Fang MC, Go AS, Chang Y, Borowwky L, Pomerczaki 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.
Meta-analysis of the performance of stroke risk scores

13. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multi-variable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMJ Med 2015;13:1.

14. Moons KG, de Groot JA, Bouwmeester W, Vergouw E, 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 Enck VH, Rookmaaker M, Blankenstein PJ, Vermeij RJVM, 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 risk stratification system for ischemic stroke in Japanese patients with non-valvular atrial fibrillation. Chin Med J (Engl) 2021;144:580–71.

17. Moons KG, Keneppe AP, Grobbbee DE, Royston P, Vergouw E, Altman DG, et al. Risk prediction models: I. 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 Jr, D’Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: data from the Framingham Heart Study. Stat Med 2008;27:157–72.

20. Moons KGM, Wolf RF, Riley RD, Whiting PF, Woodward 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;21:353–9.

24. Hothouft J, Iosipidis JP, Born 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, Hoof L, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res 2019;28:2786–88.

26. Debray TP, Damen JA, Snell KI, Ensor J, Hoof 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 the fail-safe N test. BMJ 1997;315:629–34.

29. Vehtoaibauer 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 antiplatelet 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, Lu Y, 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 J Med (Engl) 2021;134:2293–8.

32. Fox KAA, Virding S, 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.

33. Claxton J, Maclehose RF, Lussey 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.

34. 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 J Heart 2008;156:57–64.

35. Hijazi Z, Lindblad A, 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;37:11500–11.

36. Shin SH, Han SJ, Kim SJ, Im SL, Shin JH, et al. Identification of markers associated with development of stroke in “Clinically Low-Risk” atrial fibrillation patients. J Am Heart Assoc 2019;8:e012697.

37. 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;285:2644–70.

38. Hooves BD, Jacobs V, Colin P, Bunch KJ. Augmented intelligence decision tool for stroke prediction combines factors from CHA2DS2-VASc and the intermountain risk score for patients with atrial fibrillation. J Cardiovasc Electrophysiol 2019;10:1452–61.

39. 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 2019;5:1–12.

40. Berg DD, Ruff CT, Jarcimovl Galligano RP, Nordfio 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;139:70–71.

41. Kim YG, Choi JI, Boo KY, Kim DY, Hong Y, Kim MS, et al. Impact of age on thromboembolic events in patients with non-valvar atrial fibrillation. Circ Cardiovasc Qual Outcomes 2020;13:78–85.

42. Larsen TB, Lip GY, Skjeth F, Due KM, Overkaer V, Hilstved Rasmussen L. Added predictive ability of the CHA2DS2-VASc 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.

43. Li YG, Miyazawa K, Wolf A, Zibaud M, Alsheik-Ali AA, Suliman 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;274:217–25.

44. Lin LY, Lee CH, Yu CC, Tsai CT, Lai LP, Hwang JJ, et al. Risk factors and incidence of ischemic stroke in Taiwanese with nonvalvar atrial fibrillation—a nation wide database analysis. Atherosclerosis 2011;217:292–5.

45. 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.

46. 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.

47. Wang SY, Franklin JH, 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:i607.

48. Frierig LLH. Heart failure: a weak link in CHA2DS2-VASc ESC Heart Fail 2018;5:231–9.

49. Graves KG, May HT, Knowlton KU, Mulestien JB, Jacobs V, Lappe DL, et al. Improving CHA2DS2-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.

50. Ruff CT, Galligano 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, Yu YF, Tuan TC, et al. R2CHADS2 score and thromboembolic events after catheter ablation of atrial fibrillation in comparison with the CHADS2-VASc score. Can J Cardiol 2014;30:405–12.

61. Faucher L, Lecoq C, Ancedy Y, Stamboul K, Saint Etienne C, Iannes 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, Marin 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. Konecny JJ, Hindricks G, Koskij, Arya A, Sommer P, Husser D, et al. Renal dysfunction, stroke risk scores (CHADS2, CHA2DS2-VASc, and R2CHADS2), 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. Puurnen MK, Kivimäki T, Schlitt A, Rubboli A, Dietrich B, Karjalainen P, et al. CHADS2, CHA2DS2-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 JJ, Choi J, Kim YG, Oh SK, Kook H, et al. Impact of carotid atherosclerosis in CHADS2-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, Skjøtt S, 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. Shin CJ, Ou SM, Chao PW, Kuo SC, Lee YJ, Yang CY, et al. Risks of death and stroke in patients undergoing hemiablation with new-onset atrial fibrillation: a competing-risk analysis of a nationwide cohort. Circulation 2016;133:365–72.

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. PLoS One 2020;15:e0230287.

70. Banerjee AF L, Bernard-Brunet A, Clementy N, Lip GY. Composite risk scores and HAS-BLED as predictors of outcome in patients with atrial fibrillation undergoing percutaneous coronary intervention. Thromb Res 2021;199:137–9.

71. Tsuda T, Hayashi K, Fujino N, Konno T, Tada H, Nomura A, et al. Effect of hyper tension on the prediction of thromboembolism in patients with nonvalvular atrial fibrillation. Heart Rhythm 2019;16:829–37.

72. Dalgaard F, Pfeifer 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 2019:e033283.

73. Hijazi Z, Lindahl B, Oldgren J, Andersons U, Lindbäck J, Granger CB, et al. Repeated measurement of cardiac biomarkers in atrial fibrillation and validation of the ABC and P(2)-CHA(2)-DS(2)-VASc scores for predicting recurrent ischemic stroke over time. J Am Heart Assoc 2017;6:e004851.

74. Oldgren J, Hijazi Z, Lindbäck J, Alexander JT, 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.

75. 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 anticoagulants. Circulation 2021;143:1863–73.

76. Maheshwari A, Norby FL, Roekters NS, Soliman EZ, Koene RJ, Rooney MB, et al. Refining prediction of atrial fibrillation-related stroke using the P(2)-CHA(2)-DS(2)-VASc score. Circulation 2019;139:80–91.

77. 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.

78. Faucher LB, Bisson A, Herbert J, Spieesser 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.

79. Friberg L, Benson L, Lip GY. 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.

80. Kim MM, Kim SA, Choi JI, Park SM, Park SW, Kim YH, Shin WM. Improvement of predictive value for thromboembolic risk by incorporating left atrial functional parameters in the CHADS2 and CHA2DS2-VASC scores. Int Heart J 2015;56:826–92.

81. Rivera-Caravaca JM, Marín F, Víchlez 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;50:1373–9.

82. Roldán V, Arroyo AB, Salisou-Atfar S, Manzano-Fernández S, García-Barberà N, Marín F, et al. Prognostic role of MIR146A polymorphisms for cardiovascular events in atrial fibrillation. Thromb Haemost 2014;112:781–8.

83. Liu JB, Zhao TT, Cui YQ, Song K, Zhou JL, Li CJ, 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.

84. Banerjee AF L, Vourch N, Andreas D, Tammelinder 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.

85. Chao TF, Tsoo 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 CHA2DS2-VASc score. Heart Rhythm 2019;16:1753–60.

86. 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.

87. He H, Guo J, Zhang A. The value of urine albumin in predicting thromboembolic events for patients with nonvalvar atrial fibrillation. Int J Cardiol 2016;221:395–400.

88. 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.
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 CHA2DS2-VASc and HAS-BLED scores using multiple biomarkers. Ann Med 2018;50:26–34.

109. Roldán V, Vilchez JA, Manzano-Fernández S, Jover E, Galvez 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-Ginnes 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 as a risk of 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, Etierney A, Ederhy S, Ancey 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;113:525–37.

113. Roldán V, Marin 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. Ravaz K, Weissert JA, Jahangir A, Huff 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, Retisma 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, Lokhmetyga 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)CHADS2 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 PW, 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. Groenewold RH, Moons KG, Pajouheshnia R, Altmann 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, Pleeten LM, Moons KGM, Retisma JB, Groenewold RH. Accounting for treatment use when validating a prognostic model: a simulation study. BMC Med Res Methodol 2017;17:103.