

University of Groningen

Prediction models for atrial fibrillation applicable in the community

Himmelreich, Jelle C L; Veelers, Lieke; Lucassen, Wim A M; Schnabel, Renate B; Rienstra, Michiel; van Weert, Henk C P M; Harskamp, Ralf E

Published in:
Europace

DOI:
[10.1093/europace/euaa005](https://doi.org/10.1093/europace/euaa005)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Himmelreich, J. C. L., Veelers, L., Lucassen, W. A. M., Schnabel, R. B., Rienstra, M., van Weert, H. C. P. M., & Harskamp, R. E. (2020). Prediction models for atrial fibrillation applicable in the community: a systematic review and meta-analysis. *Europace*, 22(5), 684-694. <https://doi.org/10.1093/europace/euaa005>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Prediction models for atrial fibrillation applicable in the community: a systematic review and meta-analysis

Jelle C.L. Himmelreich ^{1*}, Lieke Veelers¹, Wim A.M. Lucassen¹,
Renate B. Schnabel², Michiel Rienstra ³, Henk C.P.M. van Weert¹, and
Ralf E. Harskamp¹

¹Department of General Practice, Amsterdam UMC, University of Amsterdam, Amsterdam Public Health, Meibergdreef 9, 1105 AZ Amsterdam, Netherlands; ²Department of General and Interventional Cardiology, University Heart Center Hamburg/German Center for Cardiovascular Research, Partner Site Hamburg/Kiel/Luebeck, Martinistrasse 52, 20246 Hamburg, Germany; and ³Department of Cardiology, University of Groningen, University Medical Center Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands

Received 25 September 2019; editorial decision 18 December 2019; accepted 6 January 2020; online publish-ahead-of-print 3 February 2020

Aims

Atrial fibrillation (AF) is a common arrhythmia associated with an increased stroke risk. The use of multivariable prediction models could result in more efficient primary AF screening by selecting at-risk individuals. We aimed to determine which model may be best suitable for increasing efficiency of future primary AF screening efforts.

Methods and results

We performed a systematic review on multivariable models derived, validated, and/or augmented for AF prediction in community cohorts using Pubmed, Embase, and CINAHL (Cumulative Index to Nursing and Allied Health Literature) through 1 August 2019. We performed meta-analysis of model discrimination with the summary C-statistic as the primary expression of associations using a random effects model. In case of high heterogeneity, we calculated a 95% prediction interval. We used the CHARMS (Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies) checklist for risk of bias assessment. We included 27 studies with a total of 2 978 659 unique participants among 20 cohorts with mean age ranging from 42 to 76 years. We identified 21 risk models used for incident AF risk in community cohorts. Three models showed significant summary discrimination despite high heterogeneity: CHARGE-AF (Cohorts for Heart and Aging Research in Genomic Epidemiology) [summary C-statistic 0.71; 95% confidence interval (95% CI) 0.66–0.76], FHS-AF (Framingham Heart Study risk score for AF) (summary C-statistic 0.70; 95% CI 0.64–0.76), and CHA₂DS₂-VASc (summary C-statistic 0.69; 95% CI 0.64–0.74). Of these, CHARGE-AF and FHS-AF had originally been derived for AF incidence prediction. Only CHARGE-AF, which comprises easily obtainable measurements and medical history elements, showed significant summary discrimination among cohorts that had applied a uniform (5-year) risk prediction window.

Conclusion

CHARGE-AF appeared most suitable for primary screening purposes in terms of performance and applicability in older community cohorts of predominantly European descent.

Keywords

Atrial fibrillation • Risk model • Community • Screening • Systematic review • Meta-analysis

Introduction

Atrial fibrillation (AF) is a common cardiac arrhythmia affecting over 33 million people worldwide.¹ Its incidence increases with age, with a lifetime risk of over 30%.² Due to ageing populations, the number of AF cases in Europe is expected to double to >17 million by 2060.³

Atrial fibrillation is associated with a five-fold increased risk of ischaemic stroke, which can be largely prevented by antithrombotic prophylaxis in at-risk patients.^{4,5}

Screening for AF in the community has been proposed as an approach to optimize early AF detection and to prevent AF-associated sequelae.⁶ Prior research has shown that AF screening is cost-

* Corresponding author. Tel: +31 205667683; fax: +31 205669186. E-mail address: j.c.himmelreich@amsterdamumc.nl

© The Author(s) 2020. 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 Non-Commercial License (<http://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

What's new?

- This is the first systematic review and meta-analysis designed to capture and evaluate a broad range of prognostic models used for incident atrial fibrillation (AF) risk prediction, and the first to focus specifically on models that are applicable in and have been derived, validated and/or augmented in community cohorts.
- This work was open to any model used for incident AF prediction in the community, which also enabled inclusion of models that had not been developed for incident AF but that may have merits for this aim. We hereby identified 21 models used for incident AF prediction in community cohorts.
- This work suggests that the CHARGE-AF model is likely most robust for incident AF prediction in terms of performance as well as applicability in the community.

effective when selecting patients at older age, with thresholds for screening eligibility varying from 65 to 75 years.⁷⁻⁹ The screening regimes in these primary care studies often involved single-point screening, while multiple-point screening could result in a higher yield of new AF cases.^{9,10} Multiple-point or prolonged rhythm monitoring schemes are, however, likely to be more costly for society and more burdensome to patients.⁷⁻⁹ Multivariable prediction models for incident AF could contribute to AF screening by determining a risk category for each patient.¹¹ The more intensive regimes could be assigned to those with highest risk, while those in lower-risk strata could be assigned to less stringent follow-up, or none at all. It remains, however, insufficiently clear from consensus documents whether other parameters beyond age could be used to differentiate between degrees of AF risk within the community.^{5,12}

We therefore set out to perform a systematic review and meta-analysis with two aims. First, we wished to provide an overview of AF risk models that are applicable to and have been validated in community cohorts. Such models should consist of variables that can be quickly assessed and/or are commonly available from patient records and should not require advanced diagnostic testing. Second, by synthesizing the discriminatory abilities of each included risk model, we aimed to determine which of these may be best suitable for increasing efficiency of future primary AF screening efforts.

Methods

We reported this systematic review and meta-analysis in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.¹³

Data searches

We searched Pubmed, Embase, and Cumulative Index to Nursing and Allied Health Literature (CINAHL) databases from inception through 1 August 2019. We used the keywords 'AF or atrial flutter (AFL)' and 'risk model' as well as related terms. We filtered for studies conducted on humans and written in English. The full search is shown in [Supplementary material online, Table S1](#). We checked the reference list of included studies for additional relevant references.

Study selection

To be eligible for inclusion, studies had: (i) to be original studies in adults (≥ 18 years of age); (ii) to derive, validate, and/or augment a tool for

predicting risk of incident AF/AFL based on multivariable analysis; (iii) to include only patients without a diagnosis of AF/AFL at baseline; and (iv) to incorporate into their risk prediction tool only variables that are applicable and/or commonly available in primary care settings. We included studies with AFL as co-outcome, since AF and AFL have similar clinical relevance.⁵ In light of inclusion criterion iv, we included only studies that used medical history, physical examination, simple laboratory findings, or electrocardiogram (ECG) parameters as variables in the prediction model. We excluded studies that required advanced diagnostic testing [e.g. echocardiography, genetic markers, or specialized (laboratory) tests] for their simple (non-augmented) model. We only included studies written in English. We included studies that diagnosed AF/AFL through medical records, hospitalizations, death certificates, and/or ECG during follow-up examinations. We excluded studies that selected patients for a common disease or risk factor, as such studies would not be generalizable to the community. Moreover, we excluded studies with a mean follow-up duration under 3 months since with shorter follow-up durations there would be an increasing risk of measuring prevalent AF missed at baseline recording, rather than actual incident AF. We uploaded references to a systematic review web application (Covidence, Veritas Health Innovation Ltd, Melbourne, Australia). Three investigators (J.C.L.H., L.V., and R.E.H.) assessed studies for eligibility by screening studies on title and abstract, followed by a full-text screening. Disagreements were resolved by panel discussion (J.C.L.H., L.V., R.E.H., and W.A.M.L.).

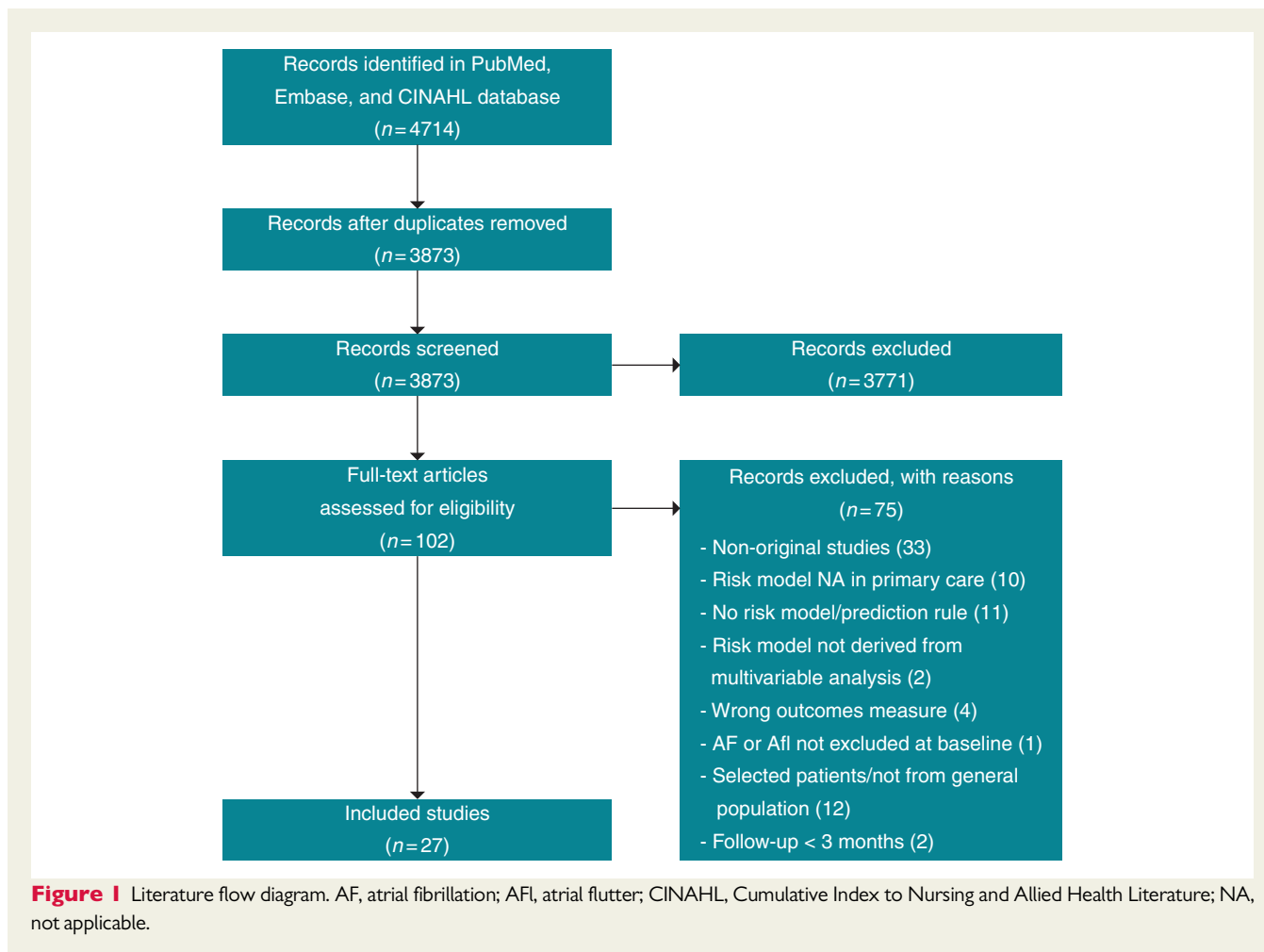
Data extraction and quality assessment

Two investigators (J.C.L.H. and L.V.) extracted data from the included studies regarding study methods, population characteristics, risk prediction model(s), and model performance. For the latter, we extracted the C-statistic and corresponding 95% confidence interval (95% CI) for discrimination, and the *P*-value of a goodness-of-fit test and the ratio of observed and expected AF/AFL cases (O:E ratio) for calibration. When authors did not report an O:E ratio we derived the O:E ratio by analysing calibration plots.¹⁴ When authors performed augmentation of pre-existing models by adding variables with an aim to enhance predictive value of models, we retrieved the net reclassification improvement (NRI) index of the augmented model compared with the original 'simple' model, as well as the augmented model's performance in terms of discrimination and calibration. We included augmentation data only when the augmentation variables were applicable to primary care settings as outlined previously.

Two investigators (J.C.L.H. and L.V.) used the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) checklist to assess the risk of bias and the applicability for our research aims.¹⁵ Our interpretation of each CHARMS domain can be found in the [Supplementary material online, Methods](#). We assessed risk of bias at the cohort level for each of the included studies. We scored each domain as either low, unclear, or high risk of bias. We defined overall risk of bias as: low, when all domains of a cohort within one study were scored as low risk of bias; unclear, when one or more domains of a cohort within one study were scored as unclear risk of bias; and high, when one or more domains of a cohort within one study were scored as high risk of bias. We resolved disagreements by discussion.

Data synthesis and statistical analysis

We reported continuous variables as means \pm standard deviations, and categorical variables as percentages. We evaluated statistical significance in all analyses at the 0.05 level. In individual studies, we assessed the C-statistic of a model, where a 95% CI containing 0.5 indicated insufficient discrimination. Calibration of a model was deemed sufficient when authors reported a *P*-value of a goodness-of-fit test >0.05 and/or an O:E ratio



ranging between 0.95 and 1.05. In assessing augmentation, we defined significant improvement as a positive NRI index with a reported 95% CI that did not contain 0. When a study reported on multiple cohorts, and presented separate data for each cohort, we assessed model performance separately for each cohort within that study.

We performed meta-analysis to assess overall discrimination of included models. The primary expression of associations in meta-analysis was the summary C-statistic and corresponding 95% CI using a random effects inverse variance model with restricted maximum likelihood estimation and Hartung–Knapp corrections.¹⁴ We conducted the meta-analyses in R using the meta and metafor packages (R Foundation for Statistical Computing, version 3.5.1). We performed meta-analysis only when C-statistic data for a prediction model were available for ≥ 3 cohorts.¹⁶ When studies presented a C-statistic without 95% CI, we calculated the 95% CI using methods described previously.¹⁴ In each meta-analysis, we calculated the mean as the summary effect measure, its 95% CI, and the I^2 statistic as an expression of the heterogeneity between studies.¹⁷ When heterogeneity in meta-analysis of C-statistics was high ($I^2 > 30\%$ ¹⁸), we derived a 95% prediction interval (95% PI) using methods described previously.¹⁹ We assessed overall discrimination of models by the summary C-statistic. When the 95% CI (or, in case of high heterogeneity, the 95% PI) of the summary C-statistic included 0.5, we concluded that there was insufficient evidence that the prediction model has significant discriminatory ability for incident AF in such populations as included in the meta-analysis.

We assessed eligibility for inclusion into meta-analysis at the cohort level. Cohorts with low or unclear overall risk of bias were eligible for inclusion into meta-analysis. When studies reported C-statistic data based on the aggregation of multiple cohorts, and one of these cohorts was assessed as having high overall risk of bias, we did not include the aggregate C-statistic data into meta-analysis. When multiple studies reported C-statistic data on the same cohort, we included the first published study into the primary analysis. In the primary meta-analysis of each model, we included cohorts with any follow-up duration.

In our primary analysis, we assessed overall discrimination of all models that had ≥ 3 eligible cohorts with C-statistic data. In the secondary analysis we performed meta-analysis for each risk model that had ≥ 3 eligible cohorts reporting C-statistic data while applying a uniform prediction window, and grouped cohorts according to the applied risk prediction window (e.g. 5 or 10 years) since this is an important methodological considerations when wanting to translate summary risk model performance to clinical settings.¹⁵

We performed a sensitivity analysis in which we restricted the primary and secondary analyses to only those cohorts that had demonstrated sufficient calibration in order to assess overall discrimination among populations where the prediction model had also shown the ability to correctly classify absolute incident AF risk. Finally, we performed a sensitivity analysis in which we replaced primary and secondary meta-analyses data from ‘double’ cohorts (cohorts from the primary meta-analysis that had also been reported on in later studies) with data on that same cohort from

any later study to assess whether later, possibly 'more complete' datasets could be of influence to our conclusions from the primary and secondary analyses.

Results

We found a total of 3873 unique references, 102 of which we subjected to full-text screening. From these, we included 27 studies^{20–46} for our final analysis (see *Figure 1*).

Characteristics of included studies

The 27 included studies were based on 20 different cohorts set in Europe ($n=8$), East Asia ($n=5$), North America ($n=5$), and the Middle East ($n=2$) (see [Supplementary material online, Table S2](#) for characteristics of the included studies and cohorts). Cohort size ranged from 646 to 1 062 073 patients, with a total of 2 978 659 unique participants. Mean age varied from 42 to 76 years, percentage of female participants ranged from 0% to 100%. Mean follow-up of the included cohorts varied from 3 to 20 years, with AF incidence during follow-up ranging from 0.2% to 24.5%. Ten cohorts used AF/Afl as the outcome, and the other 10 cohorts described only AF. Thirteen of the 20 cohorts followed a prospective design, of which 6 cohorts applied prescheduled follow-up examinations to systematically identify AF.

Characteristics of included risk models

The included studies represented data on 21 multivariable prediction models. Ten models had specifically been derived for predicting incident AF (*Table 1*). Of these, nine had been derived in community cohorts^{20–28} and one had been derived in a cohort of outpatients.⁴⁷ Five of the models derived for incident AF had also been externally validated.^{20–24} The intended risk prediction window of models derived for incident AF varied between 5 and 11 years. The FHS-AF (Framingham Heart Study risk score for AF) model had originally been derived for predicting 10-year incident AF risk, but had later been recalibrated and subsequently externally validated for 5-year risk of incident AF.^{33,40}

We identified seven risk models that had originally been derived for predicting other outcomes than incident AF but had been validated for this outcome in community cohorts^{34,48–53} and a further four models that were incidentally employed to predict incident AF but that had not specifically been derived as a prediction model for that outcome^{23,37,44,46} ([Supplementary material online, Table S3](#)).

The number of variables incorporated into each of the included models varied between 5 and 18, with a total of 27 distinguishable variables/variable categories among all included risk models. Age was the only variable used in all models. Other common variables were hypertension history or treatment, heart failure history, sex, and blood pressure, incorporated into 16, 16, 14, and 14 of the 21 included models, respectively.

Risk model performance among included cohorts

[Supplementary material online, Table S4](#) shows the results on AF incidence, discrimination, and calibration of the included simple models

among the cohorts in our search. All studies used the C-statistic to assess model discrimination for incident AF within their cohorts. Nine studies assessed calibration by providing both a *P*-value for a goodness-of-fit test and a calibration plot in at least one of the risk models that these studies reported on refs,^{20,21,24,27,29,30,32,36,40} seven studies assessed calibration only by a *P*-value for goodness-of-fit test,^{22,26,35,39,41,43,44} two studies assessed calibration only by a calibration plot,^{25,33} and nine studies reported neither of these calibration parameters.^{23,28,31,34,37,38,42,45,46}

Reported C-statistics for incident AF ranged from 0.58 (95% CI 0.55–0.61)²¹ to 0.842 (95% CI 0.826–0.858).⁴⁴ The highest C-statistic while also showing sufficient calibration was reported in the FHS (Framingham Heart Study) cohort on the in the incidentally used FHS-Lubitz model with a C-statistic of 0.78 (95% CI 0.76–0.80) and *P*-value of the goodness-of-fit test of 0.11.⁴⁴

Augmentation of included risk models

We identified augmentation data applicable to primary care settings for five of the included AF risk models (see [Supplementary material online, Table S5](#)). Significant improvement was demonstrated in CHARGE-AF (Cohorts for Heart and Aging Research in Genomic Epidemiology) with addition of the P-wave axis³⁵ and brain natriuretic peptide (BNP) and/or C-reactive protein (CRP),⁴¹ in the FHS-AF 10-year model with addition of BNP and CRP,³⁹ and in the Seirei model with addition of ECG parameters to the model.²⁷

Risk of bias assessment

[Supplementary material online, Table S6](#) shows the results of the risk of bias assessment for each cohort in the included studies. We assessed the risk of bias of all domains as either low or unclear for all domains except for the participants domain. For this domain, we assessed eight cohorts employed by 10 studies as having a high risk of bias for excluding patients for reasons with a known association with risk of future AF.^{20,26,27,29,33,40,41,43,45,46}

Meta-analyses

Five models were eligible for the primary meta-analysis, as shown in *Figure 2*. Of these, only CHARGE-AF and the FHS-AF for 10-year risk model had originally been derived for incident AF. All primary meta-analyses resulted in high heterogeneity for which we calculated a 95% PI. There were three models that resulted in a summary C-statistic with significant 95% PI in our primary meta-analysis: CHARGE-AF (summary C-statistic 0.71; 95% CI 0.66–0.76; I^2 87%; 95% PI 0.554–0.865; $n=8$ studies; $n=58$ 137 patients), the FHS-AF 10-year model (summary C-statistic 0.70; 95% CI 0.64–0.76; I^2 94%; 95% PI 0.535–0.869; $n=5$ studies; $n=33$ 846 patients), and CHA₂DS₂-VASc (summary C-statistic 0.69; 95% CI 0.64–0.74; I^2 100%; 95% PI 0.540–0.838; $n=5$ studies; $n=2$ 005 813 patients) (see *Figure 3* for a comparison of these three models).

For our secondary analysis, we were able to meta-analyse CHARGE-AF and the FHS-AF 10-year model, each for a 5- and 10-year prediction window (*Figure 4*). Only the meta-analysis of CHARGE-AF with a 5-year prediction window resulted in significant overall discrimination (summary C-statistic 0.72; 95% CI 0.66–0.78; I^2 85%; 95% PI 0.567–0.881; $n=6$ studies; $n=50$ 328 patients).

In our sensitivity analysis of restricting primary and secondary analyses models to cohorts with sufficient calibration, we found no model

Table 1 Characteristics of included risk models developed for incident AF

Model type	ARIC-AF		CHARGE-AF		C ₂ HEST		FHS-AF		Mayo		MHS		PREVEND		Seirei		Suita		WHS	
	Point-based	Cox regression	Point-based	Cox regression	Point-based	Cox regression	Point-based	Cox regression	Point-based	Cox regression	Point-based	Cox regression	Latent class analysis	Point-based	Cox regression	Point-based	Cox regression	Point-based	Cox regression	
Intended prediction window for incident AF (years)	10	5	11	5, 10	10	NS	10	10	10	10	10	10	7	10	10	10	10	10	10	
Model variables ^a																				
Age	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Sex				X	X	X							X	X	X	X	X	X	X	X
Race	X	X											X	X	X	X	X	X	X	X
Body measurements (height, weight, and BMI)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Blood pressure (systolic, diastolic)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Heart rate													X	X	X	X	X	X	X	X
Heart failure history	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Hypertension treatment or history	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Diabetes mellitus history	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Stroke history																				
CHD or MI history	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Vascular disease history																				
Alcohol use																				
Smoking	X	X											X	X	X	X	X	X	X	X
ECG parameters	X	X																		
COPD																				
Autoimmune or inflammatory disease history																				
Significant murmur	X																			
Serum lipids																				
Glomerular filtration rate																				
Urine albumin secretion																				
Thyroid disease			X																	

AF, atrial fibrillation; ARIC-AF, Atherosclerosis Risk In Communities score for Atrial Fibrillation; BMI, body mass index; CHADS₂, Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, prior Stroke or transient ischaemic attack (2 points), CHA₂DS₂-VASc, Congestive HF, Hypertension, Age >75 (2 points), Stroke/transient ischaemic attack/thromboembolism (2 points), Vascular disease, Age 65–74, Sex category: CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; CHD, coronary heart disease; C₂HEST, Coronary artery disease/Chronic obstructive pulmonary disease (2 points), Hypertension, Elderly, Systolic heart failure, Thyroid disease; COPD, chronic obstructive pulmonary disease; ECG, electrocardiogram; FHS-AF, Framingham Heart Study score for Atrial Fibrillation; MHS, Maccabi Healthcare Services; MI, myocardial infarction; NS, not specified; PREVEND, Prevention of Renal and Vascular End-stage Disease; WHS, Women's Health Study.

^aDepicted here are the variables in the simple (non-augmented) models.

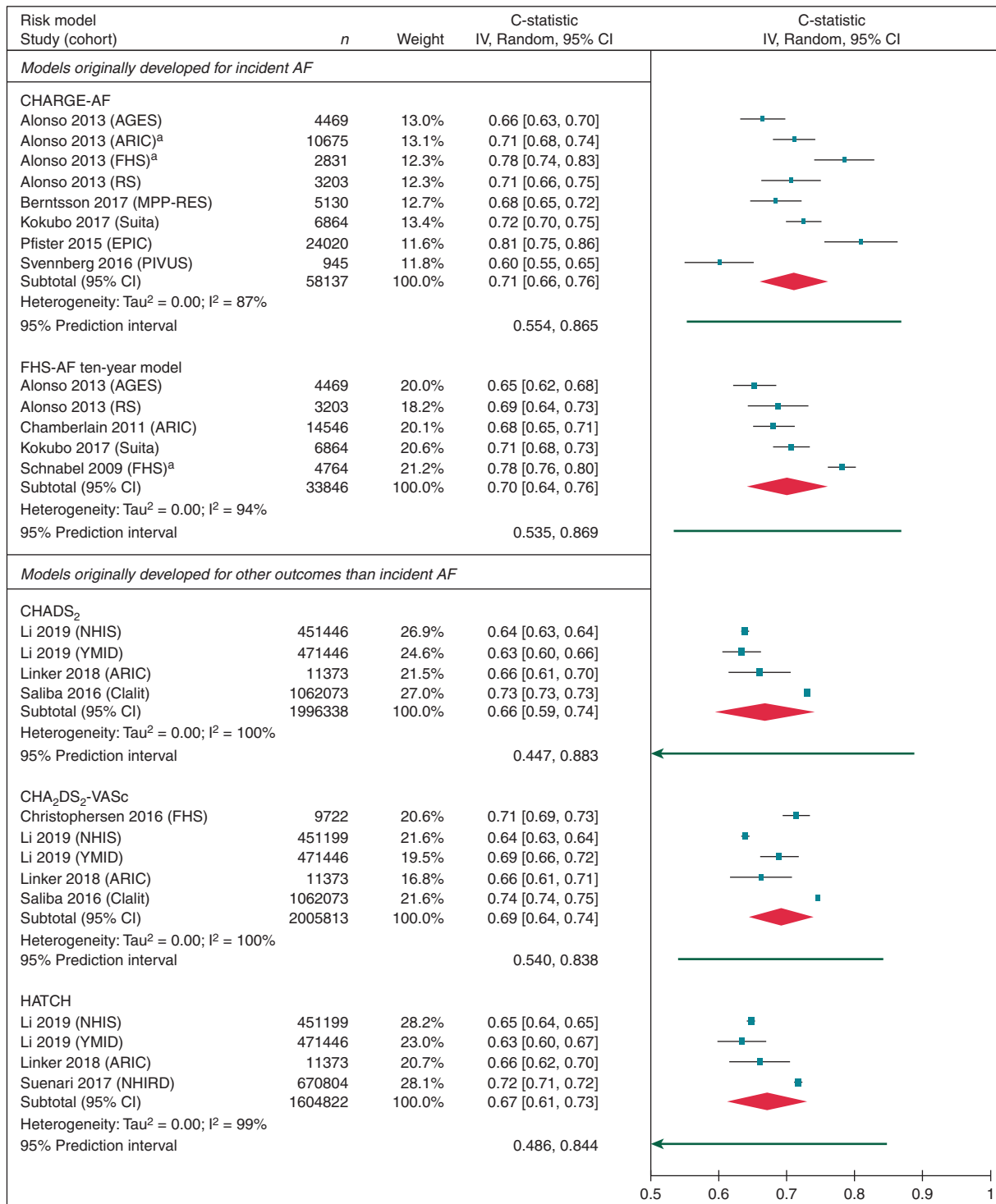


Figure 2 Primary analysis: meta-analysis of C-statistics. 95% CI, 95% confidence interval; AF, atrial fibrillation; AGES, Age, Gene and Environment-Reykjavik Study; ARIC, Atherosclerosis Risk In Communities; CHADS₂, Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, prior Stroke or transient ischaemic attack (2 points); CHA₂DS₂-VASC, Congestive HF, Hypertension, Age >75 (2 points), Stroke/transient ischaemic attack/thromboembolism (2 points), Vascular disease, Age 65–74, Sex category; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; CI, confidence interval; EPIC, European Prospective Investigation into Cancer and Nutrition; FHS, Framingham Heart Study; FHS-AF, Framingham Heart Study score for Atrial Fibrillation; HATCH, Hypertension, Age, stroke or Transient ischaemic attack, Chronic obstructive pulmonary disease, Heart failure; IV, inverse variance; MPP-RES, Malmö Preventive Project Re-examination Study; NHIS, National Health Insurance Service; NHIRD, National Health Insurance Research Database; PIVUS, Prospective Investigation of the Vasculature in Uppsala Seniors; RS, Rotterdam Study; YMID, Yunnan Medical Insurance Database. ^aDerivation cohort.

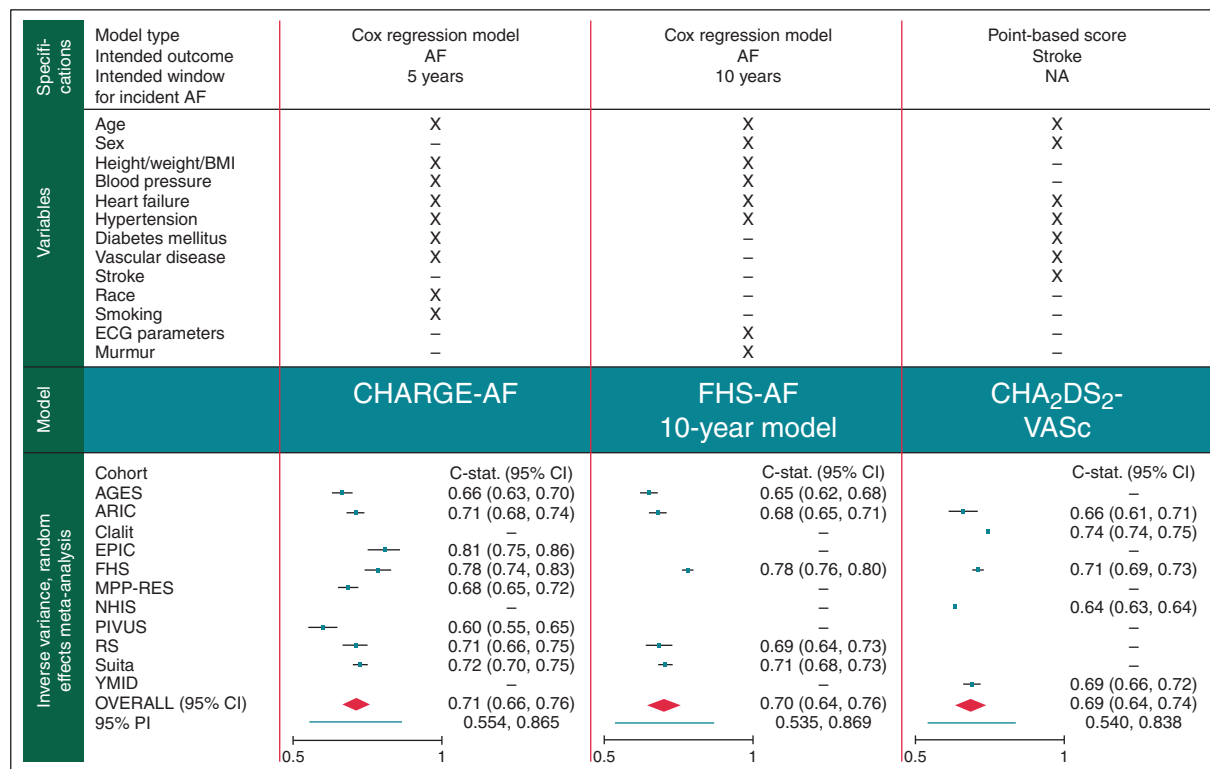


Figure 3 Comparison of the three models that resulted in significant 95% prediction intervals in the primary meta-analysis. 95% CI, 95% confidence interval; 95% PI, 95% prediction interval; AF, atrial fibrillation; AGES, Age, Gene and Environment-Reykjavik Study; ARIC, Atherosclerosis Risk In Communities; BMI, body mass index; CHA₂DS₂-VASc, Congestive HF, Hypertension, Age >75 (2 points), Stroke/transient ischaemic attack/thromboembolism (2 points), Vascular disease, Age 65–74, Sex category; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; C-stat., C-statistic; ECG, electrocardiogram; EPIC, European Prospective Investigation into Cancer and Nutrition; FHS, Framingham Heart Study; FHS-AF, Framingham Heart Study score for Atrial Fibrillation; EPIC, European Prospective Investigation into Cancer and Nutrition; MPP-RES, Malmö Preventive Project Re-examination Study; NA, not applicable; NHIS, National Health Insurance Service; PIVUS, Prospective Investigation of the Vasculature in Uppsala Seniors; RS, Rotterdam Study; YMID, Yunnan Medical Insurance Database. The meta-analyses of C-statistics for the outcome incident AF are grouped by cohort from which C-statistics were reported, allowing for a comparison of multiple models' performance within one cohort insofar as data are available.

with significant overall discrimination due to high heterogeneity (Supplementary material online, Figures S1 and S2). Our sensitivity analysis on double cohorts in the primary and secondary analyses did not lead to different conclusions on overall discriminatory ability of meta-analysed models in all but one comparison (see Supplementary material online, Table S7).

Discussion

In this systematic review and meta-analysis, we provided an overview of prediction models for incident AF risk that are applicable in and had been derived, validated, and/or augmented in community cohorts. We identified 21 risk models that met these criteria, of which 10 had specifically been derived for predicting AF incidence in the community. In meta-analysis of C-statistics, three models showed significant overall discrimination for AF incidence at any follow-up duration and with any calibration despite high heterogeneity. Two of those models were derived specifically for incident AF risk prediction:

CHARGE-AF and the FHS-AF 10-year model. Only CHARGE-AF showed significant overall discrimination among cohorts with a uniform prediction window (the model's originally intended 5-year window).

Clinical relevance

The outcomes of this systematic review and meta-analysis are highly relevant for the field of primary AF screening. Previous AF screening programmes showed only moderate efficiency in selecting at-risk patients from the community, with an estimated number needed to screen of 111 among 23 studies that had screened community cohorts for incident AF by various methods.¹² Patients were often selected for screening based only on age.^{7-9,54} The age criterion in selecting patients for AF screening has its clinical merits since oral anticoagulation in AF patients is indicated in all women ≥65 and all men ≥75 years of age and should be considered in men aged ≥65 years in the absence of other risk factors.^{5,55} Age as a criterion, however, should not be considered absolute in selecting patients for primary AF screening. Half of all AF cases detected in the Belgian

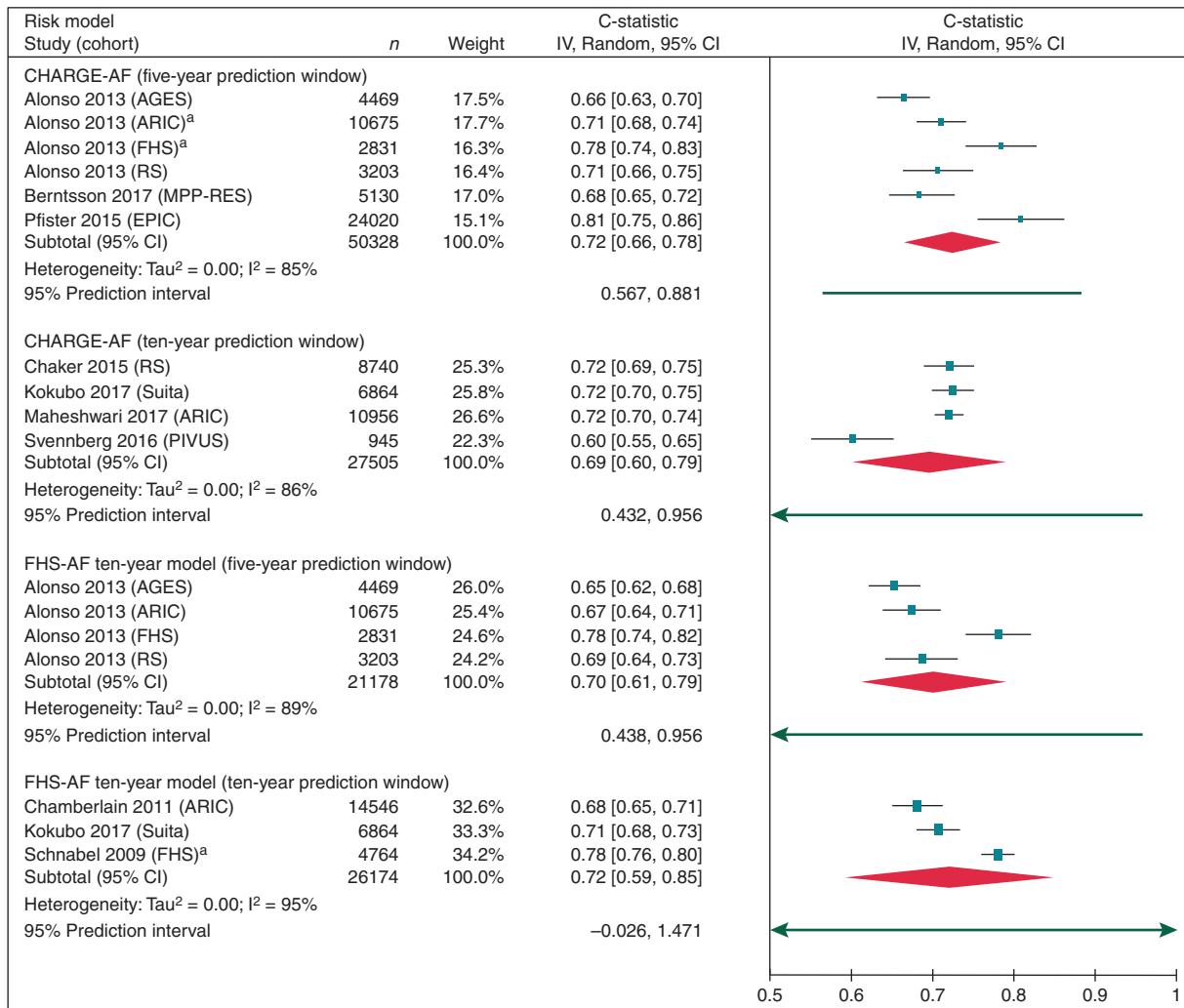


Figure 4 Secondary analysis: meta-analysis of C-statistics grouped according to application of a uniform prediction window within a model. AGES, Age, Gene and Environment-Reykjavik Study; ARIC, Atherosclerosis Risk In Communities; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; CI, confidence interval; EPIC, European Prospective Investigation into Cancer and Nutrition; FHS, Framingham Heart Study; FHS-AF, Framingham Heart Study score for Atrial Fibrillation; IV, inverse variance; MPP-RES, Malmö Preventive Project Re-examination Study; PIVUS, Prospective Investigation of the Vasculature in Uppsala Seniors; RS, Rotterdam Study; 95% CI, 95% confidence interval. ^aDerivation cohort.

Heart Rhythm Week were younger than 65 years of age.⁵⁶ Moreover, there is evidence that CHARGE-AF has higher discrimination among younger patients, although calibration here was lower due to lower absolute AF risk in this younger subgroup.³⁶ Finally, the two studies within our search that compared multivariable models with age alone as the predictor both found that the multivariable models had significantly higher C-statistics for incident AF.^{26,34} We conclude, therefore, that the use of multivariable risk models in selecting patients for community AF screening is likely to result in more efficient screening than selecting based on age alone. Given that there is adequate stroke prevention therapy available once AF is detected, it is likely that the use of such models in AF screening will result in more efficient stroke prevention.⁵ More work on the implementation of multivariable risk models in AF screening as well as on

long-term follow-up of screening-detected AF cases, however, is necessary to test these hypotheses.

Whether an immediate start of anticoagulation therapy is warranted when AF is detected in younger patients with risk factors other than high age will subsequently depend on the number and nature of these other risk factors. However, as shown in Table 1, most AF risk prediction models include a multitude of the variables in the CHA₂DS₂-VASc score used to assess anticoagulant treatment indication.^{5,55} In a younger patient selected for AF screening based on a multivariable prediction model due to presence of other risk factors than high age, an AF diagnosis is therefore likely to still be relevant in terms of the need for anticoagulant therapy, if not for prevention of other pathology associated with AF such as heart failure.⁵⁷

One of the aims of this work was to determine which model may be best suitable for increasing efficiency of future primary AF screening efforts. Our work showed that there are ample AF risk models to choose from, however with one model that currently stands out between the others: CHARGE-AF. Despite heterogeneity in included cohorts, CHARGE-AF showed significant summary discrimination over a relatively short (5-year) risk prediction window. The model contains variables that are generally easy to extract from health records, and requires only body measurements that are easily obtainable (height, weight, and blood pressure). The FHS-AF model, in contrast, though performing nearly as well in overall discrimination, requires variables 'significant murmur' and ECG variables which are less easy to acquire or interpret for many care professionals. Concluding, CHARGE-AF currently seems the most suitable prediction model for incident AF, and likely has merits as a low-cost triage test for future primary AF screening efforts.

Derivation, validation, and augmentation

In risk models derived for incident AF in community cohorts, there was a trend that the derivation cohort had the highest C-statistic compared with external validation cohorts. The only exception was CHARGE-AF, where Pfister *et al.*³⁶ reported a C-statistic of 0.808. Calibration of CHARGE-AF in their cohort, however, was insufficient (*P*-value for goodness-of-fit test <0.001 and O:E ratio 0.47) due to a systematic overestimation of 5-year AF risk in all risk deciles. One explanation lies in the differences in demographics, as Pfister's cohort was younger and had lower baseline prevalence of diabetes mellitus than the CHARGE-AF derivation cohorts. Depending on whether one's aim is to distinguish high from low-risk patients, or to predict absolute 5-year incident AF risk, a researcher may use this knowledge to decide whether or not to recalibrate a model for his own target population. In augmentation studies, we saw that addition of BNP and CRP to a model seemed most promising in terms of improving risk classification. We note, however, that the significance of an added value of BNP and CRP to CHARGE-AF was not consistent, and that the augmentation studies provided no information on the added costs of augmentation parameters relative to those of acquiring the simple model risk score.²⁹

Previous work

Previous systematic reviews have focused on individual predictors for AF,^{58,59} on AF as a risk factor for other outcomes,^{60,61} or on risk models for adverse outcomes in AF patients.^{62,63} However, to our knowledge, this is the first systematic review and meta-analysis on performance of incident AF risk prediction models, and the first with a focus on such risk models validated in and applicable to the community.

Future work

Future studies could focus on finding optimal cut-offs for the more promising AF prediction models, and to find the most cost-effective use of multivariable models within various screening schemes. Researchers may opt here, e.g. for either a dichotomization into patients with higher and lower risk or assigning patients to one of multiple risk strata. Patients at higher risk could be offered a more intensive, sensitive screening scheme (e.g. multiple-point screening or Holter monitoring) when compared with

patients at low risk (single point or no screening). Further research could also assess whether implementation of multivariable models in AF screening could be aided by software that automatically extracts patient data from health records, informs the physician of a patient's current risk category, and suggests parameters that should be updated for a more accurate current AF risk stratification.

Strengths and limitations

This study has a number of strengths. First, we included only studies performed in community cohorts, which contributed to the value of our results for primary care AF screening. Second, we included any risk model that was used to predict risk of incident AF. This enabled us to expand our scope to models that had originally not been intended for predicting incident AF, but that may have merits in predicting this outcome. Third, we attributed high bias to studies that excluded or over-represented patients based on factors that are likely to be associated with risk of incident AF, further contributing to the generalizability of our results to the community. Fourth, we included only the C-statistic of raw, non-bootstrapped data into meta-analysis in order to not bias the meta-analysis with potentially overly narrow confidence intervals. Finally, we refrained from meta-regression or subgroup meta-analysis based on e.g. a subdivision of cohorts' mean age, AF incidence or region, to explain the heterogeneity in our results. Such analyses from aggregate data are known to have a high risk of especially ecological bias and are inferior to subgroup results derived from individual participant data (IPD).¹⁴ An IPD meta-analysis, however, was not the scope of the current study.

The primary limitation of our study is the high heterogeneity of included studies. We attempted to cope with this limitation by performing sensitivity analyses and by calculating a 95% PI in our meta-analyses with high heterogeneity. The outcomes of our meta-analyses with significant 95% PI can be considered generalizable to such populations as included into those meta-analyses, despite high heterogeneity. As a second limitation, we did not provide a meta-analysis on model calibration since such analyses are often challenging due to a lack of calibration measures reported among studies.¹⁴ Indeed, we found that meta-analysable data on calibration was poorly reported on among included studies ([Supplementary material online, Table S4](#)). Moreover, summarizing O:E data would have automatically excluded those models that were not originally intended for incident AF, since expected incident AF rates would never have been defined for such models. We addressed calibration by performing a sensitivity analysis among cohorts which had demonstrated sufficient calibration by their applied risk model(s). A third limitation is that we included both prospective and retrospective cohort studies. This may have introduced bias as AF is not always symptomatic⁶⁴ and asymptomatic patients are less likely to undergo rhythm evaluation when left to their physicians' discretion than when ECG is performed in the context of a prescheduled follow-up. The restriction of our search to studies written in English which we applied for quality-related as well as practical reasons, finally, has been found not to lead to significant bias.⁶⁵

Conclusion

We provided an overview of prediction models for incident AF risk that are applicable in and have been derived, validated, and/or augmented in community cohorts. We identified 21 risk models that met these criteria. Of these, CHARGE-AF seemed the most robust in terms of performance as well as applicability in the community.

Supplementary material

Supplementary material is available at *Europace* online.

Funding

This work was supported by the Netherlands Organisation for Health Research and Development (ZonMw) [80-83910-98-13046]. The authors had full autonomy in design, conduct, and reporting of the manuscript. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 648131) and German Center for Cardiovascular Research (DZHK e.V.) (81Z1710103). Salary support for R.E.H. was provided by a Rubicon Fellowship grant by the Netherlands Organisation for Scientific Research (NWO) (grant No. 452173116).

Conflict of interest: none declared.

References

- Chugh SS, Roth GA, Gillum RF, Mensah GA. Global burden of atrial fibrillation in developed and developing nations. *Glob Heart* 2014;**9**:113–9.
- Magnussen C, Niiranen TJ, Ojeda FM, Gianfagna F, Blankenberg S, Njolstad I et al. Sex differences and similarities in atrial fibrillation epidemiology, risk factors, and mortality in community cohorts: results from the BiomarcARE Consortium (Biomarker for Cardiovascular Risk Assessment in Europe). *Circulation* 2017;**136**:1588–97.
- Schnabel RB, Yin X, Gona P, Larson MG, Beiser AS, McManus DD et al. 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham Heart Study: a cohort study. *Lancet* 2015;**386**:154–62.
- Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke* 1991;**22**:983–8.
- Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B et al. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Europace* 2016;**18**:1609–78.
- Freedman B, Camm J, Calkins H, Healey JS, Rosenqvist M, Wang JG et al. Screening for atrial fibrillation: a report of the AF-SCREEN International Collaboration. *Circulation* 2017;**135**:1851–67.
- Lowres N, Neubeck L, Salkeld G, Krass I, McLachlan AJ, Redfern J et al. Feasibility and cost-effectiveness of stroke prevention through community screening for atrial fibrillation using iPhone ECG in pharmacies. The SEARCH-AF study. *Thromb Haemost* 2014;**111**:1167–76.
- Jacobs MS, Kaasenbrood F, Postma MJ, van Hulst M, Tieleman RG. Cost-effectiveness of screening for atrial fibrillation in primary care with a handheld, single-lead electrocardiogram device in the Netherlands. *Europace* 2018;**20**:12–8.
- Aronsson M, Svennberg E, Rosenqvist M, Engdahl J, Al-Khalili F, Friberg L et al. Cost-effectiveness of mass screening for untreated atrial fibrillation using intermittent ECG recording. *Europace* 2015;**17**:1023–9.
- Lowres N, Neubeck L, Redfern J, Freedman SB. Screening to identify unknown atrial fibrillation. A systematic review. *Thromb Haemost* 2013;**110**:213–22.
- Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012;**98**:683–90.
- Mairesse GH, Moran P, Van Gelder IC, Elsner C, Rosenqvist M, Mant J et al. Screening for atrial fibrillation: a European Heart Rhythm Association (EHRA) consensus document endorsed by the Heart Rhythm Society (HRS), Asia Pacific Heart Rhythm Society (APHRS), and Sociedad Latinoamericana de Estimulacion Cardiaca y Electrofisiologia (SOLAEC). *Europace* 2017;**19**:1589–623.
- Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;**62**:1006–12.
- 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.
- 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.
- Partlett C, Riley RD. Random effects meta-analysis: coverage performance of 95% confidence and prediction intervals following REML estimation. *Stat Med* 2017;**36**:301–17.
- Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011;**342**:d549.
- Higgins JPT, Green S. (eds). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated March 2011]. The Cochrane Collaboration, 2011. www.handbook.cochrane.org.
- Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* 2009;**172**:137–59.
- Alonso A, Krijthe BP, Aspelund T, Stepas KA, Pencina MJ, Moser CB et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc* 2013;**2**:e000102.
- Chambertain AM, Agarwal SK, Folsom AR, Soliman EZ, Chambless LE, Crow R et al. A clinical risk score for atrial fibrillation in a biracial prospective cohort (from the Atherosclerosis Risk in Communities [ARIC] study). *Am J Cardiol* 2011;**107**:85–91.
- Li YG, Pastori D, Farcomeni A, Yang PS, Jang E, Joung B et al. A Simple Clinical Risk Score (C₂H₂EST) for predicting incident atrial fibrillation in Asian subjects: derivation in 471,446 Chinese subjects, with internal validation and external application in 451,199 Korean subjects. *Chest* 2019;**155**:510–8.
- Rienstra M, Geelhoed B, Yin XY, Siland JE, Vermond RA, Mulder BA et al. Cluster individuals based on phenotype and determine the risk for atrial fibrillation in the PREVENT and Framingham Heart Study Populations. *PLoS One* 2016;**11**:e0165828.
- Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB Sr et al. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *Lancet* 2009;**373**:739–45.
- Aronson D, Shalev V, Katz R, Chodick G, Mutlak D. Risk score for prediction of 10-year atrial fibrillation: a community-based study. *Thromb Haemost* 2018;**118**:1556–63.
- Everett BM, Cook NR, Conen D, Chasman DI, Ridker PM, Albert CM. Novel genetic markers improve measures of atrial fibrillation risk prediction. *Eur Heart J* 2013;**34**:2243–51.
- Hamada R, Muto S. Simple risk model and score for predicting of incident atrial fibrillation in Japanese. *J Cardiol* 2019;**73**:65–72.
- Kokubo Y, Watanabe M, Higashiyama A, Nakao YM, Kusano K, Miyamoto Y. Development of a basic risk score for incident atrial fibrillation in a Japanese general population—the Suita study. *Circ J* 2017;**81**:1580–8.
- Alonso A, Roetker NS, Soliman EZ, Chen LY, Greenland P, Heckbert SR. Prediction of atrial fibrillation in a racially diverse cohort: the Multi-Ethnic Study of Atherosclerosis (MESA). *J Am Heart Assoc* 2016;**5**:1–8.
- Berntsson J, Smith JG, Nilsson PM, Hedblad B, Melander O, Engstrom G. Pro-atrial natriuretic peptide and prediction of atrial fibrillation and stroke: the Malmo Preventive Project. *Eur J Prev Cardiol* 2017;**24**:788–95.
- Chaker L, Heeringa J, Dehghan A, Medici M, Visser WE, Baumgartner C et al. Normal thyroid function and the risk of atrial fibrillation: the Rotterdam study. *J Clin Endocrinol Metab* 2015;**100**:3718–24.
- Christophersen IE, Yin XY, Larson MG, Lubitz SA, Magnani JW, McManus DD et al. A comparison of the CHARGE-AF and the CHA(2)DS(2)-VASc risk scores for prediction of atrial fibrillation in the Framingham Heart Study. *Am Heart J* 2016;**178**:45–54.
- Kumarathurai P, Mouridsen MR, Mattsson N, Larsen BS, Nielsen O, Gerdts TA et al. Atrial ectopy and N-terminal pro-B-type natriuretic peptide as predictors of atrial fibrillation: a population-based cohort study. *Europace* 2017;**19**:364–70.
- Linker DT, Murphy TB, Mokdad AH. Selective screening for atrial fibrillation using multivariable risk models. *Heart* 2018;**104**:1492–9.
- Maheshwari A, Norby FL, Soliman EZ, Koene R, Rooney M, O'Neal WT et al. Refining prediction of atrial fibrillation risk in the general population with analysis of P-wave axis (from the Atherosclerosis Risk in Communities Study). *Am J Cardiol* 2017;**120**:1980–4.
- Pfister R, Bragelmann J, Michels G, Wareham NJ, Luben R, Khaw KT. Performance of the CHARGE-AF risk model for incident atrial fibrillation in the EPIC Norfolk cohort. *Eur J Prev Cardiol* 2015;**22**:932–9.
- Rienstra M, Yin XY, Larson MG, Fontes JD, Magnani JW, McManus DD et al. Relation between soluble ST2, growth differentiation factor-15, and high-sensitivity troponin I and incident atrial fibrillation. *Am Heart J* 2014;**167**:109–15.e2.

38. Saliba W, Gronich N, Barnett-Griness O, Rennett G. Usefulness of CHADS₂ and CHA₂DS₂-VASc scores in the prediction of new-onset atrial fibrillation: a population-based study. *Am J Med* 2016;**129**:843–9.
39. Schnabel RB, Larson MG, Yamamoto JF, Sullivan LM, Pencina MJ, Meigs JB et al. Relations of biomarkers of distinct pathophysiological pathways and atrial fibrillation incidence in the community. *Circulation* 2010;**121**:200–7.
40. Schnabel RB, Aspelund T, Li G, Sullivan LM, Suchy-Dicey A, Harris TB et al. Validation of an atrial fibrillation risk algorithm in whites and African Americans. *Arch Intern Med* 2010;**170**:1909–17.
41. Sinner MF, Stepas KA, Moser CB, Krijthe BP, Aspelund T, Sotoodehnia N et al. B-type natriuretic peptide and C-reactive protein in the prediction of atrial fibrillation risk: the CHARGE-AF Consortium of community-based cohort studies. *Europace* 2014;**16**:1426–33.
42. Suenari K, Chao TF, Liu CJ, Kihara Y, Chen TJ, Chen SA. Usefulness of HATCH score in the prediction of new-onset atrial fibrillation for Asians. *Medicine (Baltimore)* 2017;**96**:e5597.
43. Svennberg E, Lindahl B, Berglund L, Eggers KM, Venge P, Zethelius B et al. NT-proBNP is a powerful predictor for incident atrial fibrillation—validation of a multimarker approach. *Int J Cardiol* 2016;**223**:74–81.
44. Lubitz SA, Yin XY, Fontes JD, Magnani JW, Rienstra M, Pai M et al. Association between familial atrial fibrillation and risk of new-onset atrial fibrillation. *JAMA* 2010;**304**:2263–9.
45. Rosenberg MA, Patton KK, Sotoodehnia N, Karas MG, Kizer JR, Zimetbaum PJ et al. The impact of height on the risk of atrial fibrillation: the Cardiovascular Health Study. *Eur Heart J* 2012;**33**:2709–17.
46. Magnani JW, Zhu L, Lopez F, Pencina MJ, Agarwal SK, Soliman EZ et al. P-wave indices and atrial fibrillation: cross-cohort assessments from the Framingham Heart Study (FHS) and Atherosclerosis Risk in Communities (ARIC) study. *Am Heart J* 2015;**169**:53–61.e1.
47. Brunner KJ, Bunch TJ, Mullin CM, May HT, Bair TL, Elliot DW et al. Clinical predictors of risk for atrial fibrillation: implications for diagnosis and monitoring. *Mayo Clin Proc* 2014;**89**:1498–505.
48. Chambless LE, Folsom AR, Sharrett AR, Sorlie P, Couper D, Szklo M et al. Coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC) study. *J Clin Epidemiol* 2003;**56**:880–90.
49. Gage BF, Waterman AD, Shannon W, Boehler 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.
50. Lip GY, Nieuwlaet 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.
51. Wilson PWF, 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.
52. D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P; CHD Risk Prediction Group. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA* 2001;**286**:180–7.
53. de Vos CB, Pisters R, Nieuwlaet R, Prins MH, Tieleman RG, Coelen R-JS et al. Progression from paroxysmal to persistent atrial fibrillation clinical correlates and prognosis. *J Am Coll Cardiol* 2010;**55**:725–31.
54. Fitzmaurice DA, Hobbs FD, Jowett S, Mant J, Murray ET, Holder R et al. Screening versus routine practice in detection of atrial fibrillation in patients aged 65 or over: cluster randomised controlled trial. *BMJ* 2007;**335**:383.
55. Jones C, Pollit V, Fitzmaurice D, Cowan C, Grp GD. The management of atrial fibrillation: summary of updated NICE guidance. *BMJ* 2014;348.
56. Proietti M, Mairesse GH, Goethals P, Scavee C, Vijgen J, Blankoff I et al.; Belgian Heart Rhythm Week Investigators. A population screening programme for atrial fibrillation: a report from the Belgian Heart Rhythm Week screening programme. *Europace* 2016;**18**:1779–86.
57. Singh JP, Thomas SS. Atrial fibrillation and heart failure prevention: do we need a risk score? *JACC Heart Fail* 2017;**5**:53–5.
58. Jones NR, Taylor KS, Taylor CJ, Aveyard P. Weight change and the risk of incident atrial fibrillation: a systematic review and meta-analysis. *Heart* 2019;**105**:1799–805.
59. Himmelreich JCL, Lucassen WAM, Heugens M, Bossuyt PMM, Tan HL, Harskamp RE et al. Frequent premature atrial contractions are associated with atrial fibrillation, brain ischaemia, and mortality: a systematic review and meta-analysis. *Europace* 2019;**21**:698–707.
60. He WQ, Chu YJ. Atrial fibrillation as a prognostic indicator of myocardial infarction and cardiovascular death: a systematic review and meta-analysis. *Sci Rep* 2017;**7**:1–13.
61. Ruddox V, Sandven I, Munkhaugen J, Skattebu J, Edvardsen T, Otterstad JE. Atrial fibrillation and the risk for myocardial infarction, all-cause mortality and heart failure: a systematic review and meta-analysis. *Eur J Prev Cardiol* 2017;**24**:1555–66.
62. Proietti M, Farcomeni A, Romiti GF, Di Rocco A, Placentino F, Diemberger I et al. Association between clinical risk scores and mortality in atrial fibrillation: systematic review and network meta-regression of 669,000 patients. *Eur J Prev Cardiol* 2018;doi:10.1177/2047487318817662.
63. Caldeira D, Costa J, Fernandes RM, Pinto FJ, Ferreira JJ. Performance of the HAS-BLED high bleeding-risk category, compared to ATRIA and HEMORR2HAGES in patients with atrial fibrillation: a systematic review and meta-analysis. *J Interv Card Electrophysiol* 2014;**40**:277–84.
64. Frykman V, Frick M, Jensen-Urstad M, Ostergren J, Rosenqvist M. Asymptomatic versus symptomatic persistent atrial fibrillation: clinical and noninvasive characteristics. *J Intern Med* 2001;**250**:390–7.
65. Morrison A, Polisen J, Huserreau D, Moulton K, Clark M, Fiander M et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *Int J Technol Assess Health Care* 2012;**28**:138–44.