BMJ Open

Validating risk models versus age alone for atrial fibrillation in a young Dutch population cohort: should atrial fibrillation risk prediction be expanded to younger community members?

Jelle C L Himmelreich ,[1] Ralf E Harskamp,[1] Bastiaan Geelhoed,[2] Saverio Virdone,[3] Wim A M Lucassen,[1] Ron T Gansevoort,[4] Michiel Rienstra[2]

For numbered affiliations see end of article.

Correspondence to
Jelle C L Himmelreich;
j.c.himmelreich@amsterdamumc.nl

ABSTRACT

Background Advancing age is the primary selection criterion for community screening for atrial fibrillation (AF), with selection often restricted to those aged ≥65 years. If multivariable models were shown to have considerable additional value over age alone in predicting AF risk among younger individuals, AF screening could be expanded to patients with lower age, but with high AF risk as per a validated risk model.

Methods We validated risk models CHARGE-AF (Cohorts for Heart and Aging Research in Genomic Epidemiology model for AF) and FHS-AF (Framingham Heart Study model for AF), and risk scores CHA_2DS_2-VASc and CHA_2DS_2-VA, and presented their predictive abilities for 5-year and 10-year AF risk versus that of age alone in a young Dutch population cohort (PREVEND) free from AF at baseline. We assessed discrimination by the C-statistic and calibration by the calibration plot and stratified Kaplan-Meier plot using survey-weighted Cox models.

Results During 5-year and 10-year follow-up there were n=98 (2.46/1000 person-years) and n=249 (3.29/1000 person-years) new AF cases, respectively, among 8265 participants with mean age 49±13 years. CHARGE-AF and FHS-AF both showed good discrimination for 5-year and 10-year AF (C-statistic range 0.83–0.86) with accurate calibration for 5-year AF, but overestimation of 10-year AF risk in highest-risk individuals. CHA_2DS_2-VASc and CHA_2DS_2-VA relatively underperformed. Age alone showed similar discrimination to that of CHARGE-AF and FHS-AF both in the overall, young PREVEND cohort and in subgroups for lower age and lower stroke risk.

Conclusion Multivariable models accurately discriminate for 5-year and 10-year AF risk among young European community-dwelling individuals. However, their additional discriminatory value over age alone was limited. Selection strategies for primary AF screening using multivariable models should not be expanded to younger individuals.

Strengths and limitations of this study

► Validation for incident atrial fibrillation of risk models and comparison with age alone, performed specifically in a young cohort.
► Applied appropriate weighting to account for oversampling of patients with microalbuminuria in the PREVEND data set.
► Absence of established methods for use in the presence of survey weights prevented adjustment for the competing risk of death or formal testing for difference in C-statistic.
► Ethnic diversity was minimal in the PREVEND data set; findings mainly extend to a European-white population.

Timely diagnosis of AF is crucial in order to decrease the risk of complications such as heart failure and most notably ischaemic stroke.[2] In this regard, the importance of efficiently identifying patients at high risk of AF, eligible for screening, is increasingly recognised.[3 4]

Previous community AF screening strategies often selected only for advancing age.[5–8] However, multivariable risk models have been shown to more accurately discriminate for elevated AF risk than selecting for high age alone in older community populations.[9] In a recent systematic review and meta-analysis of prediction models validated for incident AF in community settings, three models showed significant summary discrimination: CHARGE-AF (Cohorts for Heart and Aging Research in Genomic Epidemiology model for AF),[10] FHS-AF (Framingham Heart Study model for AF)[11] and CHA_2DS_2-VASc (originally developed to assess stroke risk in individuals with AF, but with demonstrated merits in discriminating for incident AF risk as well).[4 12]

BACKGROUND

Atrial fibrillation (AF) is the most commonly encountered cardiac arrhythmia in the community with a lifetime risk of over 30%.[1]

The systematic review concluded that CHARGE-AF, which was developed to predict patients' absolute 5-year AF risk based on 11 easily obtainable variables, seemed most promising as a risk stratification tool for community AF screening.[4]

An important finding from the systematic review[4] was that validation cohorts of AF prediction models largely constituted older populations, while there was evidence for higher model discrimination among subgroups of younger patients.[13–15] Although younger AF patients would be less likely to reach the threshold for an anti-coagulation indication, it can be argued that early AF detection is still important in these patient to increase long-term health and quality of life through, for example, lifestyle interventions and rate/rhythm control.[2] And while the number needed to screen for AF using single time-point measurements rise steeply with lower age,[16] it remains unanswered whether multivariable models could have increased this efficiency among younger individuals over a primarily age-based cut-off. Further research into the relative value of multivariable models over age alone in predicting AF specifically among younger community-dwelling individuals is thus warranted. Such work could help researchers decide whether to expand their scope to younger patients, or to adhere to the current convention of selecting primarily for advanced age—whether or not followed by triage based on a risk factor[17] or risk model.[18]

In the current study, we aim to assess whether the current convention of selecting primarily for advanced age in AF screening selection could be challenged by using multivariable prediction models or risk scores to discriminate for high AF risk among younger patients as well. We will do so by comparing predictive performance for 5-year and 10-year AF risk of two multivariable models (CHARGE-AF and FHS-AF) and two risk scores (CHA$_2$DS$_2$-VASc and CHA$_2$DS$_2$-VA) with that of age alone, in a young community-based cohort from The Netherlands.

METHODS
Study population
The Prevention of Renal and Vascular End-Stage Disease (PREVEND) study was designed to evaluate the natural course of microalbuminuria and its relation to renal and cardiovascular disease in the general population. Details of this community-based cohort study have been outlined previously.[19–21] In summary, participants in the PREVEND study involved inhabitants of the city of Groningen, The Netherlands, between 28 and 75 years of age who responded to the request to participate in a one-time morning urine albumin screening test. All respondents with urine albumin ≥10 mg/L (n=7768) and a random sample of those with urine albumin <10 mg/L (n=3394) were invited for baseline screening, excluding those who were pregnant or who used insulin. Baseline screening involved assessment of demographic factors, anthropometric measurements, cardiovascular and metabolic risk factors, health behaviour assessment, blood sample

collection and two 24-hours urine samples. All persons who completed baseline screening were invited for follow-up visits at the outpatient clinic at 3-year intervals. A standard 12-lead ECG was performed during each visit.

Diagnosing AF
Incident AF was diagnosed when either AF or atrial flutter was present on the ECG obtained at one of the PREVEND follow-up visits or at an outpatient visit or hospital admission to one of the hospitals in the city of Groningen (University Medical Center Groningen or Martini Hospital). All ECGs were electronically screened for the following criteria: PR interval absence, AF, atrial flutter or atrial ectopy. This method was validated with complete manual screening by two independent observers of all ECGs from the PREVEND baseline visit, and showed 100% sensitivity for AF. All ECGs with suspected AF as determined by electronic screening were manually reviewed by two independent observers. Disagreements or inconsistencies were resolved by two independent cardiologists.[20]

Clinical follow-up data
Follow-up in PREVEND ranged up to 14 years. We calculated follow-up duration as the time between the baseline screening visit and the last contact date in case of loss to follow-up, death, diagnosis of AF or the end of the 5-year or 10-year window of our respective analyses, whichever came first. Information on death was obtained through a national government vital statistics databases.[20]

Models and risk scores
CHARGE-AF is a multivariable Cox model that uses the variables age, race/ethnicity, height, weight systolic blood pressure (SBP), diastolic blood pressure, current smoker status, antihypertensive medication use, diabetes, heart failure and myocardial infarction to derive a person's absolute 5-year risk of AF.[10] The coefficients and calculations used to derive CHARGE-AF as well as the other original models used in our analyses are shown in the online supplemental methods.

FHS-AF is a multivariable Cox model that was originally derived and validated to predict 10-year AF risk, and was later recalibrated and validated to also predict 5-year AF risk.[11 22] Both models require the variables age, sex, body mass index, SBP, hypertension treatment, PR interval and heart failure. The original, 10-year model additionally requires significant cardiac murmur.

CHA$_2$DS$_2$-VASc (Congestive heart failure, Hypertension, Age>75, Stroke/transient ischaemic attack/thromboembolism, Vascular disease, Age 65–74, Sex category) was originally developed to assess stroke risk in individuals with AF, and is commonly used to assess the indication for anticoagulation as stroke prophylaxis in patients with AF.[2 12] The risk score has also been demonstrated to have merits in discriminating for incident AF risk in community cohort.[4]

CHA$_2$DS$_2$-VA is a modified CHA$_2$DS$_2$-VASc omitting sex category as variable. It is recommended by the Cardiac

Society of Australia and New Zealand to assess stroke risk,[23] and is analogous to international recommendations to differentiate between men and women in choosing a CHA_2DS_2-VASc cut-off for stroke risk assessment.[2 24] The model was recently shown to have merits in selecting patients for single time-point AF screening as well as prolonged monitoring for AF, and—like CHA_2DS_2-VASc—has the advantage of not requiring data such as height, weight or blood pressure which are often missing in routine primary care data.[15 18 25] CHA_2DS_2-VA has not yet been validated for incident AF over an extended time period in community cohorts, warranting further assessment.

Extending CHARGE-AF to 10-year follow-up

While CHARGE-AF's discrimination for 10-year AF risk has been validated, the model has never been calibrated to predict absolute AF risk over a 10-year window.[9 13 26–28] We recalibrated CHARGE-AF's coefficients using the PREVEND data set censored at 10-year follow-up in order to enable an assessment of its calibration for 10-year AF, to investigate its use in younger community cohorts and to facilitate a direct comparison of calibration with the FHS-AF model for 10-year AF risk prediction.[11] We used the newly recalibrated 10-year CHARGE-AF model in our analyses on 10-year AF risk.

Multivariable predictors versus age alone

We presented predictive ability for 5-year and 10-year AF risk among risk models and risk scores, followed by a discussion of relative performance between multivariable predictors and age alone. For CHARGE-AF and FHS-AF we were able to assess both discrimination and calibration for 5-year and 10-year AF using their original and recalibrated models. In validating CHA_2DS_2-VASc, CHA_2DS_2-VA and age as predictors for AF risk we were only able to assess discrimination since these models have never been formally calibrated to predict absolute AF risk in the community.

Additional analyses

We performed stratified analyses by age <65 versus ≥65 years of age as well as by CHA_2DS_2-VASc <2 versus ≥2 in order to replicate previous studies that showed differences in performance among these clinically relevant subgroups,[9 13 14] and in order to further assess the merits of multivariable risk models over age alone in younger and lower-risk participants.

In our analysis on the original CHARGE-AF simple model for 5-year AF risk we also assessed model performance when routinely scoring all participants' race as 'Caucasian/white'. We added this analysis because information on race is not routinely scored in Dutch health records, and to validate previous research that used this approach to deal with systematically missing ethnicity data.[15]

Statistical analysis

We presented descriptive statistics as means±SD, median (IQR) or number (percentage). We compared baseline parameters between those who did and did not develop AF during follow-up using the unpaired t-test for continuous variables and Pearson's χ^2 test for categorical variables. We assessed statistical significance at the 0.05 level in all analyses.

In our validation and recalibration analyses, we used a survey-weighted Cox proportional hazards model to account for the oversampling of individuals with microalbuminuria at baseline, and to enable generalisation of results to the general population. Here, we applied a weighing factor of 11.92 to people with urinary albumin excretion <10 mg/L and a weighing factor of 1.66 to people with urinary albumin excretion >10 mg/L. The weighting factors were derived based on the unequal inclusion probabilities.[29]

We assessed discrimination in all analyses using the C-statistic and 95% CI. We assessed calibration in analyses on CHARGE-AF and FHS-AF by the calibration plot as well as the calibration slope and 95% CI where a 95% CI not containing one indicated statistically insufficient calibration.[30] We additionally assessed calibration of CHARGE-AF and FHS-AF by the Kaplan-Meier (KM) plot grouped according to baseline risk in categories <2.5%, 2.5%–5% and >5% for 5-year risk and categories <5%, 5%–10% and >10% for 10-year risk.

We assessed missing data and found that this was <5% for all variables in the CHARGE-AF model, except for use of antihypertensive medication (n=1460, 17.7% missing data). We assumed no antihypertensive medication use in patients who did not have a documented history of hypertension, which left 363 patients with hypertension (4.4%) with missing data on antihypertensive medication use. We assumed absence of significant murmur when data for this variable was missing. The percentage of missing data among ECG was <2.5%. There was no missing data for the outcomes AF or death. To account for missing data in our validation analyses we used single imputation under the missing at random assumption.

We performed statistical analyses using SPSS V.26,[31] and R using the haven, dplyr, rms, survey and survival packages.[32]

Patient and public involvement

No patients were involved in the development of the research question and outcome measures.

RESULTS

Study population

A total of 8592 persons completed baseline screening, from which we excluded participants without an ECG at enrolment (n=248), as well as participants with prevalent AF at the baseline screening (n=79). Among the study population of 8265 participants, mean age at baseline was 49±13 years (range 29–74), the proportion of women was 50.2% and 95% were Caucasian/white. Patients who developed AF at 5-year and 10-year follow-up were significantly older, were more likely to be men and had a higher

burden of cardiovascular risk factors and clinically manifested cardiovascular disease (table 1).

Outcomes and follow-up

Mean follow-up among the total cohort was 11.2 years. Figure 1 illustrates the occurrence of AF and death events and cumulative AF incidence over 10 years. There were n=98 (2.46/1000 person-years) and n=249 (3.29/1000 person-years) incident AF cases at 5 and 10 years of follow-up, respectively. There were n=226 (5.68/1000 person-years) and n=490 (6.48/1000 person-years) deaths at 5 and 10 years of follow-up, respectively.

CHARGE-AF and FHS-AF: validation for 5-year and 10-year AF risk

Validation of CHARGE-AF and FHS-AF for 5-year AF risk both resulted in a C-statistic (95% CI) of 0.86 (0.82 to 0.90). with calibration slopes not significantly deviating from 1 in either analysis (table 2). The calibration plots for both models showed accurate absolute risk prediction throughout deciles of predicted 5-year AF risk (figure 2, *Panels A-B*), with KM plots for both models showing accurate predictions for all predefined risk groups (figure 2, *Panels C-D*).

Recalibration of the CHARGE-AF simple model for 10-year AF risk resulted in the coefficients shown in online supplemental table S1). In predicting 10-year AF risk, the C-statistic (95% CI) for the newly recalibrated CHARGE-AF model was 0.85 (0.82 to 0.88), and that of FHS-AF was 0.83 (0.79 to 0.86). Calibration slopes did not significantly deviate from 1 in either analysis (table 2). Calibration plots of both models showed accurate risk prediction in lower-risk subgroups but overestimation of absolute 10-year AF risk in those with highest predicted risk (online supplemental figure S1, *Panels A-B*). The stratified KM plots showed accurate risk estimation for all predefined risk groups by FHS-AF, while the newly recalibrated CHARGE-AF overestimated absolute risk for the middle (5%–10% 10-year risk) group (online supplemental figure S1, *Panels C-D*).

CHARGE-AF and FHS-AF: additional analyses

In the subgroups analyses for both 5-year and 10-year AF risk, C-statistics of CHARGE-AF and FHS-AF were higher in younger (<65 years) versus higher age (≥65 years), and higher in lower (CHA_2DS_2-VASc<2) versus higher (CHA_2DS_2-VASc≥2) stroke risk (table 2). Calibration slopes did not significantly differ from 1 in the analysed subgroups except for FHS-AF's 5-year analysis in patients with CHA_2DS_2-VASc<2 (table 2). Routinely scoring race as Caucasian/white in the 5-year analysis for CHARGE-AF gave similar results as the overall analysis with a C-statistic of 0.86 (95%CI: 0.82 to 0.90) and a calibration slope of 1.09 (95%CI: 0.86 to 1.32).

CHA_2DS_2VASc and CHA_2DS_2VA: overall and subgroup validation

The C-statistic (95% CI) of CHA_2DS_2VASc and CHA_2DS_2VA in the overall PREVEND cohort were 0.72 (0.63 to 0.81) and 0.80 (0.73 to 0.86), respectively, for 5-year AF, and

0.68 (0.62 to 0.73) and 0.74 (0.70 to 0.79), respectively, for 10-year AF risk (table 2). Like in the overall analysis, CHA_2DS_2VA showed higher C-statistic (95% CI) for AF risk than CHA_2DS_2VASc in all subgroup analyses, with CHA_2DS_2VASc's C-statistic for AF being non-significant in all subgroup analyses except those CHA_2DS_2VASc ≥2 for 5-year and 10-year AF risk.

Multivariable predictors versus age alone

Discrimination of age alone for 5-year and 10-year AF in the overall PREVEND cohort was almost on par with that of CHARGE-AF and FHS-AF, and considerably higher than that of CHA_2DS_2VASc and—to a lesser extent—CHA_2DS_2VA (table 2).

In the subgroup analyses, age alone as predictor shared CHARGE-AF and FHS-AF's pattern of higher C-statistic for AF among younger and lower-CHA_2DS_2VASc subgroups. In most subgroup analyses, including those for participants <65 years of age at baseline (n=7012), C-statistic point estimates of the multivariable models CHARGE-AF and FHS-AF were higher than age alone, but with considerable overlap in confidence intervals. The multivariable models' advantage versus age in terms of C-statistic was most pronounced in older and higher-CHA_2DS_2VASc subgroups in the analyses on 10-year AF risk.

When assessing discrimination of age alone versus that of CHA_2DS_2VASc and CHA_2DS_2VA in the subgroup analyses, age's C-statistic point estimate was higher in all subgroup analyses except when comparing to CHA_2DS_2VA in those with CHA_2DS_2VASc ≥2 (n=1568). As in the overall analysis, however, age alone considerably outperformed CHA_2DS_2VASc in the subgroups analyses, while CHA_2DS_2VA showed intermediate predictive ability.

DISCUSSION

Our study indicates that the discriminatory advantage of multivariable models over age alone assumed from earlier work appears to be limited in a younger, lower-risk cohort. While multivariable models CHARGE-AF and FHS-AF were shown to have good predictive properties for 5-year and 10-year AF risk, their discrimination was only marginally better than age alone as linear predictor in the young PREVEND cohort. The risk scores CHA_2DS_2VASc, and to a lesser extent CHA_2DS_2VA, even underperformed relative to age alone as well as to the multivariable models.

Comparison to previous studies

Previous external validation of CHARGE-AF showed that CHARGE-AF has an overall robust performance in older community cohorts.[4] Our study elaborated on these findings by validating CHARGE-AF in a young population with relatively low prevalence of cardiovascular risk factors and relatively low AF incidence. We hereby confirmed that discrimination of multivariable models for new AF is high among younger community members, while showing that the discriminatory advantage of multivariable models

Table 1 Baseline characteristics of study participants in PREVEND cohort stratified by presence of atrial fibrillation at 5 and 10 years of follow-up

Characteristic	All participants (n=8265)	5-year follow-up			10-year follow-up		
		AF (n=98)	No AF (n=8167)	P value*	AF (n=249)	No AF (n=8016)	P value*
Age, years	48 (38–59)	66 (60–71)	48 (38–59)	<0.001	64 (58–69)	47 (38–58)	<0.001
Male	49.8	71.4	49.6	<0.001	70.3	49.2	<0.001
Race, Caucasian/white	95.7	96.9	95.7	0.567	96.7	95.7	0.441
Systolic blood pressure, mm Hg	126 (114–141)	148 (126–161)	125 (114–140)	<0.001	143 (127–156)	125 (113–140)	<0.001
Diastolic blood pressure, mm Hg	73 (67–80)	79 (72–85)	73 (67–80)	<0.001	78 (72–85)	73 (67–80)	<0.001
Height, cm	173 (166–180)	173 (167–180)	173 (166–180)	0.492	175 (169–181)	173 (166–180)	0.003
Weight, kg	77 (68–87)	81 (74–89)	77 (68–87)	0.004	84 (76–93)	77 (68–87)	<0.001
BMI, kg/m²	26 (23–28)	27 (25–29)	26 (23–28)	0.002	27 (25–30)	26 (23–28)	<0.001
Obesity, BMI≥30	15.7	15.5	15.7	0.940	23.2	15.5	0.001
Hypertension	27.8	61.1	27.4	<0.001	57.5	26.9	<0.001
Antihypertensive medication use	13.9	41.9	13.6	<0.001	41.2	13.1	<0.001
Diabetes mellitus	3.8	7.4	3.8	0.070	9.1	3.7	<0.001
Myocardial infarction	3.4	17.5	3.3	<0.001	15.3	3.1	<0.001
Heart failure	0.2	4.1	0.2	<0.001	2.4	0.1	<0.001
Stroke	0.7	6.2	0.6	<0.001	2.9	0.6	<0.001
Current smoker	34.3	23.5	34.4	0.024	28.2	34.4	0.042
Heavy drinker	12.8	12.4	12.8	0.895	11.7	12.9	0.591
Significant murmur	1.1	11.2	0.9	<0.001	8.4	0.8	<0.001
CHA$_2$DS$_2$-VASc score	1 (0–1)	2 (1–3)	1 (0–1)	<0.001	2 (1–2)	1 (0–1)	<0.001
CHA$_2$DS$_2$-VASc≥2	22.8	58.2	22.4	<0.001	55.7	21.9	<0.001
Glucose, mmol/L	4.7 (4.3–5.1)	5.0 (4.6–5.6)	4.7 (4.3–5.1)	0.006	5.0 (4.6–5.6)	4.7 (4.3–5.1)	<0.001
Creatinine, µmol/L	82 (74–92)	90 (79–101)	82 (73–92)	<0.001	88 (77–98)	82 (73–92)	<0.001
eGFR, mL/min	80 (71–90)	73 (65–82)	80 (71–90)	<0.001	75 (67–84)	80 (72–90)	<0.001
Pro-ANP, pmol/L	48 (35–65)	86 (56–129)	47 (34–65)	<0.001	79 (54–110)	47 (34–64)	<0.001
NT-proBNP, ng/L	37 (17–73)	148 (66–371)	37 (17–71)	<0.001	105 (44–251)	36 (16–70)	<0.001
hsCRP, mg/L	1.3 (0.6–2.9)	2.0 (0.7–3.7)	1.3 (0.6–2.9)	0.292	2.1 (0.9–3.5)	1.3 (0.6–2.9)	0.037
hsTnT, pg/mL	0.003 (0.003–0.005)	0.007 (0.004–0.010)	0.003 (0.003–0.005)	0.001	0.006 (0.003–0.009)	0.003 (0.003–0.004)	<0.001
Copeptin, pmol/L	4.6 (2.9–7.5)	5.0 (3.1–9.7)	4.6 (2.9–7.5)	0.451	5.6 (3.4–9.1)	4.6 (2.9–7.5)	0.067
PR interval duration, ms	158 (143–172)	173 (157–197)	158 (143–172)	<0.001	170 (153–188)	158 (143–172)	<0.001
PAC	0.6	3.1	0.5	0.001	1.6	0.5	0.023

Continued

Table 1 Continued

Characteristic	All participants (n=8265)	5-year follow-up			10-year follow-up		
		AF (n=98)	No AF (n=8167)	P value*	AF (n=249)	No AF (n=8016)	P value*
LVH	1.9	8.4	1.8	<0.001	5.7	1.8	<0.001
UAE≥10 mg/L	69.7	76.5	69.6	0.138	76.7	69.5	0.014

Data are presented as percentage or median (IQR).

*P values for difference between participants with and without AF during indicated follow-up period, using unpaired t-test for continuous variables and Pearson's χ^2 test for categorical variables.

AF, atrial fibrillation; BMI, body mass index; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age>75 (2 points), Stroke/transient ischaemic attack/thromboembolism (2 points), Vascular disease, Age 65–74, Sex category; cm, centimetres; eGFR, estimated glomerular filtration ratio; hsCRP, highly sensitive C-reactive protein; hsTnT, highly sensitive troponin T; kg, kilograms; LVH, left ventricular hypertrophy; m, metres; mg/L, milligram per litre; mm Hg, millimetres mercury; ms, milliseconds; ng/L, nanogram per litre; NT-proBNP, N-terminal prohormone of brain natriuretic peptide; PAC, premature atrial contraction; pro-ANP, pro-atrial natriuretic peptide; UAE, urinary albumin excretion.
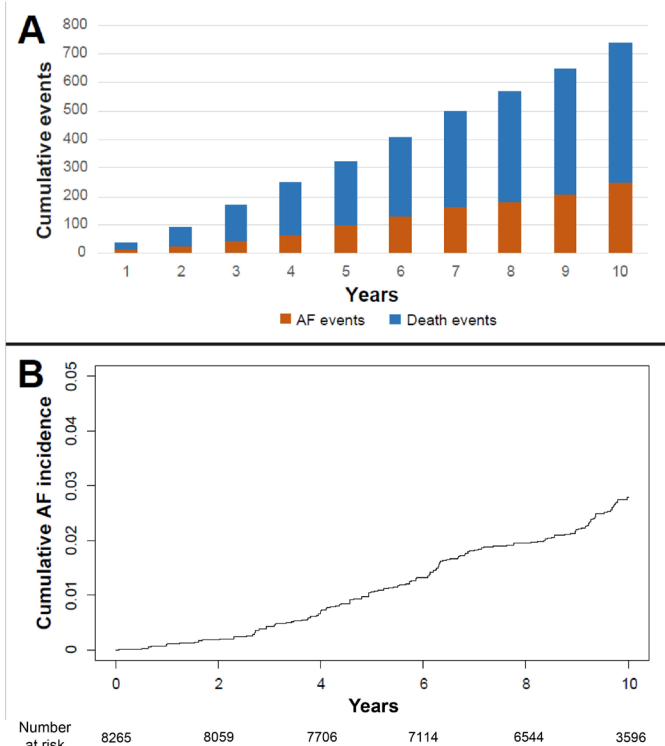

Figure 1 (A) cumulative number of events over 10-year follow-up; (B) cumulative AF incidence during 10-year follow-up. AF, atrial fibrillation.

versus age alone assumed from earlier work appears to be limited in a younger, lower-risk cohort.[15]

A recent study validating risk models for prevalent AF on 12-lead ECG confirmed the high potential of risk models in efficient patient selection for AF screening.[33] As in our work, CHARGE-AF was among the top performing models. In their cohort with mean age 64.8 years (72.9 among AF-positive cases), the authors concluded that diagnostic accuracy of multivariable models was superior to the age cut-off ≥65 years. The authors did not provide an analysis comparing multivariable models to age as continuous linear predictor for patients under 65 years of age, as we did in the current work. Although generalisability to true community screening was limited by the nature of their sample (self-referred, self-funded attendees to a private vascular screening clinic), it was a first large-scale indication that multivariable models could be used for AF screening selection.[33]

The PREVEND cohort's extensive follow-up and rich data set including ECG parameters allowed us to recalibrate CHARGE-AF for 10-year AF risk, and to directly compare the model with the 10-year FHS-AF model. Our finding that discrimination for new AF was similar for CHARGE-AF and FHS-AF, while both performed better than CHA_2DS_2-VASc, concurred with previous work.[4 9]

We were to our knowledge the first to validate CHA_2DS_2-VA for 5-year and 10-year AF risk in the community. The higher C-statistics seen using CHA_2DS_2-VA compared with its counterpart, CHA_2DS_2-VASc, are not surprising given the omission of female sex category,

Table 2 Validation of CHARGE-AF, FHS-AF, CHA₂DS-VASc₂, CHA₂DS-VA and age alone for 5-year and 10-year AF risk

	CHARGE-AF*	FHS-AF†	CHA$_2$DS$_2$-VASc	CHA$_2$DS$_2$-VA	Age
5-year risk prediction					
All (n=8265; 98 events)					
C-statistic (95% CI)	0.86 (0.82 to 0.90)	0.86 (0.82 to 0.90)	0.72 (0.63 to 0.81)	0.80 (0.73 to 0.86)	0.84 (0.78 to 0.89)
Calibration slope (95% CI)	1.09 (0.86 to 1.31)	1.18 (0.98 to 1.38)	–	–	–
Age <65 (n=7012; 45 events)					
C-statistic (95% CI)	0.84 (0.79 to 0.89)	0.84 (0.77 to 0.91)	0.58 (0.42 to 0.74)	0.68 (0.57 to 0.80)	0.78 (0.71 to 0.86)
Calibration slope (95% CI)	1.17 (0.86 to 1.47)	1.17 (0.95 to 1.39)	–	–	–
Age ≥65 (n=1253; 53 events)					
C-statistic (95% CI)	0.69 (0.59 to 0.78)	0.69 (0.59 to 0.79)	0.56 (0.45 to 0.67)	0.64 (0.55 to 0.73)	0.65 (0.55 to 0.75)
Calibration slope (95% CI)	1.10 (0.52 to 1.68)	1.06 (0.53 to 1.59)	–	–	–
CHA$_2$DS$_2$-VASc <2 (n=6697; 41 events)					
C-statistic (95% CI)	0.85 (0.79 to 0.91)	0.88 (0.82 to 0.93)	0.57 (0.45 to 0.68)	0.66 (0.54 to 0.77)	0.82 (0.74 to 0.90)
Calibration slope (95% CI)	1.12 (0.77 to 1.47)	1.37 (1.01 to 1.73)	–	–	–
CHA$_2$DS$_2$-VASc ≥2 (n=1568; 57 events)					
C-statistic (95% CI)	0.70 (0.60 to 0.79)	0.69 (0.60 to 0.79)	0.61 (0.52 to 0.70)	0.67 (0.58 to 0.76)	0.65 (0.55 to 0.76)
Calibration slope (95% CI)	0.80 (0.34 to 1.25)	0.77 (0.38 to 1.16)	–	–	–
10-year risk prediction					
All (n=8265; 249 events)					
C-statistic (95% CI)	0.85 (0.82 to 0.88)	0.83 (0.79 to 0.86)	0.68 (0.62 to 0.73)	0.74 (0.70 to 0.79)	0.81 (0.78 to 0.85)
Calibration slope (95% CI)	1.00 (0.88 to 1.12)	0.93 (0.79 to 1.08)	–	–	–
Age <65 (n=7012; 139 events)					
C-statistic (95% CI)	0.84 (0.80 to 0.87)	0.80 (0.75 to 0.85)	0.57 (0.49 to 0.65)	0.64 (0.58 to 0.70)	0.79 (0.74 to 0.83)
Calibration slope (95% CI)	1.10 (0.91 to 1.28)	0.90 (0.73 to 1.08)	–	–	–
Age ≥65 (n=1253; 110 events)					
C-statistic (95% CI)	0.69 (0.63 to 0.75)	0.69 (0.62 to 0.76)	0.54 (0.46 to 0.62)	0.62 (0.55 to 0.69)	0.58 (0.51 to 0.66)
Calibration slope (95% CI)	0.93 (0.64 to 1.21)	1.02 (0.61 to 1.43)	–	–	–
CHA$_2$DS$_2$-VASc <2 (n=6697; 122 events)					
C-statistic (95% CI)	0.84 (0.80 to 0.88)	0.82 (0.77 to 0.87)	0.52 (0.45 to 0.58)	0.63 (0.57 to 0.69)	0.81 (0.76 to 0.86)
Calibration slope (95% CI)	1.07 (0.88 to 1.26)	0.96 (0.74 to 1.18)	–	–	–
CHA$_2$DS$_2$-VASc ≥2 (n=1568; 127 events)					
C-statistic (95% CI)	0.72 (0.65 to 0.78)	0.70 (0.63 to 0.77)	0.58 (0.52 to 0.64)	0.64 (0.58 to 0.71)	0.62 (0.54 to 0.69)
Calibration slope (95% CI)	0.81 (0.57 to 1.06)	0.69 (0.41 to 0.98)	–	–	–

*Using the coefficients of the original CHARGE-AF model (Alonso 2013) in our analyses on 5-year risk prediction, and using the coefficients of the recalibrated CHARGE-AF model for 10-year risk prediction (current work) in our analyses on 10-year risk prediction.
†Using the coefficients of the recalibrated FHS-AF model (Schnabel 2010) in our analyses on 5-year risk prediction, and using the coefficients of the original FHS-AF model (Schnabel 2009) in our analyses on 10-year risk prediction.
AF, atrial fibrillation; CHA$_2$DS$_2$-VASc, Congestive heart failure, Hypertension, Age, Diabetes and previous Stroke or Transient Ischaemic Attack, Vascular disease and female Sex category; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology model for AF; FHS-AF, Framingham Heart Study model for AF.

which is known to be negatively associated with incident AF.[2]

The association of individual determinants and biomarkers with AF in the PREVEND data set outside the context of risk prediction models can be found in previous work.[20] Addition of biomarkers to CHARGE-AF as done by previous authors, whether through CHARGE-AF's established augmented model or by addition of individual biomarkers,[9 10] showed no meaningful discriminatory improvement in our sample (data not shown).

Clinical implications and future directions

Our findings are relevant for patient selection for AF screening, an aim for which risk models are now increasingly looked at.[3] The similarity in discrimination between multivariable models and age alone, both in the overall
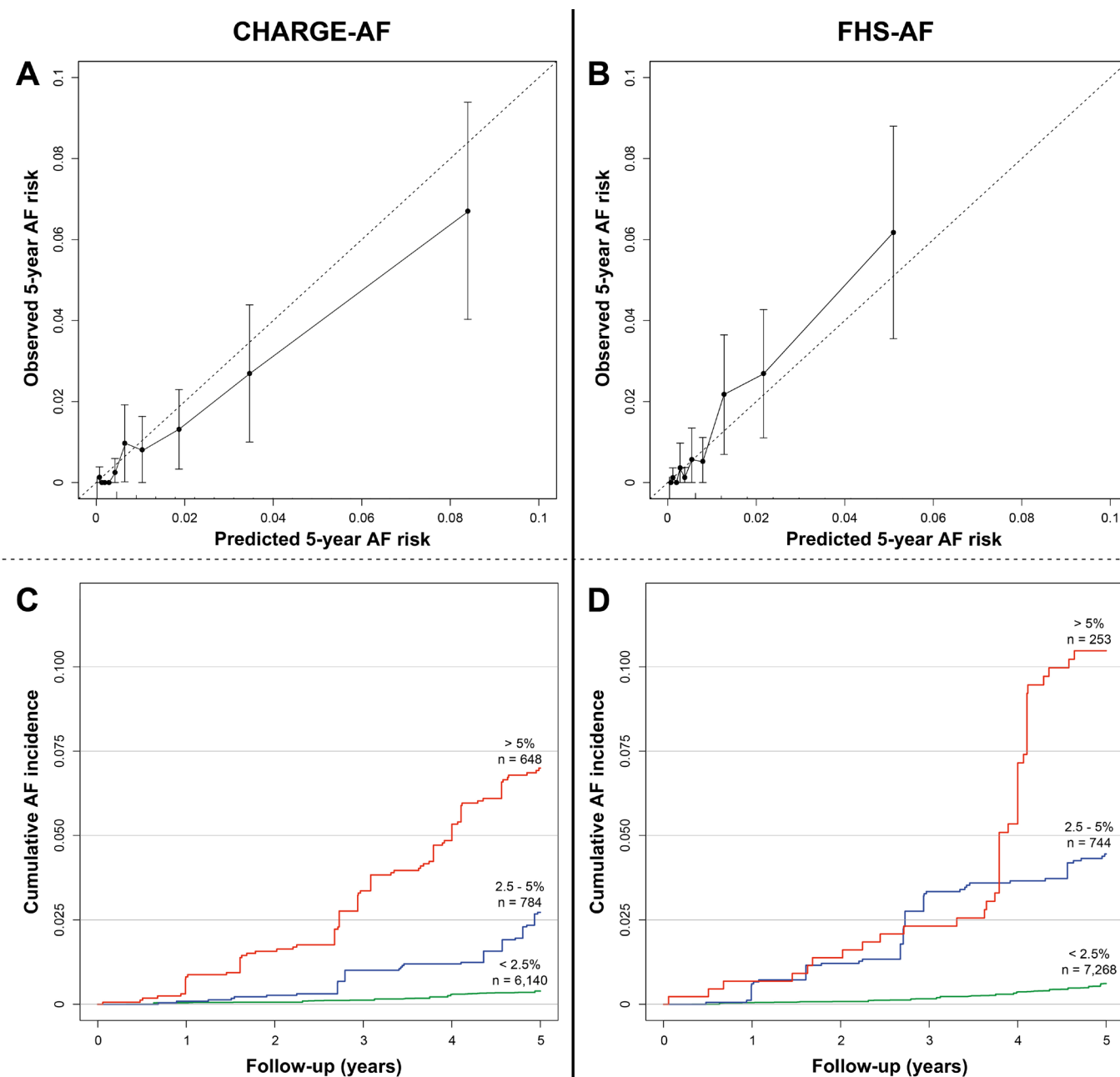
7

CHARGE-AF

FHS-AF

Figure 2 (A and B) Calibration plots for 5-year AF risk of the CHARGE-AF and FHS-AF models, respectively, in the PREVEND cohort; (C and D) Kaplan-Meier plots or 5-year AF risk according to subgroups <2.5% (green line), 2.5%–5% (blue line) and >5% (red line) predicted 5-year risk for 5-year AF risk of the CHARGE-AF and FHS-AF models, respectively, in the PREVEND cohort. Calibration plot interpretation: Points with brackets indicate intersects of mean observed and mean CHARGE-AF predicted AF risk per decile of predicted CHARGE-AF risk, with 95% CI of observed risk per decile. Data exceeding the neutral (dotted) line indicates underestimation of absolute AF risk by the CHARGE-AF simple model in our sample, while data below the neutral line indicates overestimation of absolute AF risk by the model. Spikes on the x-axis represent the distribution of AF-free survivors by baseline predicted risk. AF, atrial fibrillation; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology model for AF; FHS-AF, Framingham Heart Study model for AF.

young PREVEND cohort and in the subgroup analyses of patients <65 years of age, suggest limited additional value of multivariable models for AF screening triage among younger community-dwelling adults. The current analysis therefore supports the current convention of applying patient selection algorithms for AF screening primarily to those with advanced age.[7 17]

For implementation in community-based medicine, CHARGE-AF holds an edge over other existing risk models derived for AF in that it does not require ECG or laboratory data.[11 28 34] However, recent work has indicated that even for CHARGE-AF to be used as a remote, low-cost triage test for AF screening selection, there are still considerable limitations from incomplete routine care

data.[15] Until such limitations are overcome, for example, by valid imputation techniques applicable in routine primary care data, researchers may need to look for more universally applicable risk models to assess risk of AF. However both models in our analysis that do not require laboratory, ECG or physical measurements (CHA_2DS_2-VA and CHA_2DS_2-VASc) were still outperformed by age as linear predictor for AF risk in the young PREVEND cohort. Recent work has indeed indicated that age ≥65 years is almost on par with CHA_2DS_2-VA in sensitivity for AF on single time-point screening.[25] Future work on AF risk prediction using routine care data, rather than prospective cohorts with high data completeness, could further work towards the optimal selection strategy for AF screening in settings with a high degree of incomplete data. We advise to incorporate age as single predictor in such analyses as well.

A future direction from this work's recalibration of CHARGE-AF for 10-year AF risk is to investigate integration of the model with established cardiovascular risk calculators such as SCORE in Europe, QRISK in the UK or ASCVD in North America.[35–37] Many of the variables used in these tools (eg, age, blood pressure, smoking status) as well as treatment recommendations (eg, anti-hypertensive medication use) are variables that can also be used to calculate the CHARGE-AF model. Physicians can use these data to discuss working on modifiable risk factors that may not only positively affect the risk of (cardio)vascular disease, but perhaps also the risk of AF and associated complications, such as heart failure and stroke.[2] We note that further external validation of the newly recalibrated CHARGE-AF model for 10-year AF risk is warranted before researching integration with other cardiovascular risk calculators.

Strengths and limitations

Primary strength of this community-based validation study is that it is the first to specifically validate risk models for incident AF in a young cohort. With 85% of its participants being below the age of 65, the PREVEND cohort allowed us to be the first to thoroughly test whether current conventions in AF screening selection could be challenged by the use of multivariable models in younger patients, to which the answer seems to be negative. Further strengths are the cohort's detailed clinical assessment, the availability of ECG measurements, the extensive follow-up period and robust validation of AF by 3 yearly study visit including ECG and linkage of data from local hospitals. This allowed for validation of multiple models in one data set as well as for recalibration of the original CHARGE-AF model for 10-year data. Where there was missing data, we resolved this by imputation, as recommended over complete case analysis in risk model validation.[38] We note that CHARGE-AF validation in complete cases was similar to the imputed data set (data not shown). Finally, the analysis on routinely scoring race as Caucasian/white allowed us to validate this approach to deal with structural missingness of this variable in a predominantly Caucasian/white cohort, as has been done before in routine care data.[15]

There are a number of limitations to consider. First, the PREVEND cohort's oversampling of patients with micro-albuminuria impairs the cohort's generalisability to the community at large and to cohorts previously used in risk model validation studies. We ameliorated this limitation by using a weighting scheme based on the known oversampling rate in the PREVEND data set, as has been used previously.[29] In the absence of established methods for use in the presence of survey weights, we were unable to adjust for the competing risk of death or perform a formal test for difference in C-statistic. Third, while a significant effort was made to capture AF by performing 3 yearly ECG and by linkage of data from local hospitals, PREVEND did not include continuous ECG recordings and may thus have overlooked asymptomatic episodic/paroxysmal forms of AF. Also, AF cases diagnosed by general practitioners may not have been included when the AF diagnosis had not been made in one of the local hospitals. Fourth, ethnic diversity was minimal in PREVEND, and as such our findings mainly extend to a European-white population. While CHARGE-AF has been assessed in African-American, Hispanic and East-Asian populations,[10] the performance of CHARGE-AF in other ethnic groups should be an area of future research. Fifth, due to the demographic make-up of the PREVEND cohort, the sample of participants aged 65 years or older was limited in size. While this resulted in relatively wide C-statistic confidence intervals for this subgroup, we note that our results coincide with aggregated results from previous validation studies performed in older community cohorts.[4] Sixth, we were unable to externally validate our recalibrated CHARGE-AF coefficients for predicting 10-year AF risk. We chose not to perform split-data set validation as this has been advised against,[39] and as such, external validation has to be awaited before definitive conclusions can be drawn as of the performance of the recalibrated 10-year CHARGE-AF model.

CONCLUSION

In this validation study of multivariable predictors and age alone for 5-year and 10-year AF, we showed that there is little additional prognostic value of multivariable models over age alone in younger individuals. Discrimination of age alone for new AF was on par with that of two multivariable models (CHARGE-AF and FHS-AF), and exceeded that of two multivariable risk scores (CHA_2DS_2-VASc and CHA_2DS_2-VA). Our work suggests that selection strategies for primary AF screening using multivariable models should not be expanded to younger individuals.

Author affiliations
[1]Department of General Practice, Amsterdam Public Health, Amsterdam UMC Locatie AMC, Amsterdam, The Netherlands
[2]Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
[3]Department of Statistics, Thrombosis Research Institute, London, UK

[4]Department of Nephrology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

ORCID iD
Jelle C L Himmelreich http://orcid.org/0000-0003-0430-1583

REFERENCES

1. Magnussen C, Niiranen TJ, Ojeda FM, 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.
2. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European association for Cardio-Thoracic surgery (EACTS): the task force for the diagnosis and management of atrial fibrillation of the European Society of cardiology (ESC) developed with the special contribution of the European heart rhythm association (EHRA) of the ESC. *Eur Heart J* 2021;42:373–98.
3. Freedman B, Camm J, Calkins H, et al. Screening for atrial fibrillation: a report of the AF-SCREEN international collaboration. *Circulation* 2017;135:1851.
4. Himmelreich JCL, Veelers L, Lucassen WAM, et al. Prediction models for atrial fibrillation applicable in the community: a systematic review and meta-analysis. *Europace* 2020;22:684–94.
5. Lowres N, Neubeck L, Salkeld G, 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.
6. Aronsson M, Svennberg E, Rosenqvist M, et al. Cost-Effectiveness of mass screening for untreated atrial fibrillation using intermittent ECG recording. *Europace* 2015;17:1023–9.
7. Uittenbogaart SB, Verbiest-van Gurp N, Lucassen WAM, et al. Opportunistic screening versus usual care for detection of atrial fibrillation in primary care: cluster randomised controlled trial. *BMJ* 2020;370:m3208.
8. Kaasenbrood F, Hollander M, de Bruijn SH, et al. Opportunistic screening versus usual care for diagnosing atrial fibrillation in general practice: a cluster randomised controlled trial. *Br J Gen Pract* 2020;70:e427–33.
9. Linker DT, Murphy TB, Mokdad AH. Selective screening for atrial fibrillation using multivariable risk models. *Heart* 2018;104:1492–9.
10. Alonso A, Krijthe BP, Aspelund T, 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.
11. Schnabel RB, Sullivan LM, Levy D, et al. Development of a risk score for atrial fibrillation (Framingham heart study): a community-based cohort study. *Lancet* 2009;373:739–45.
12. Lip GYH, Nieuwlaat R, Pisters R, et al. 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.
13. Chaker L, Heeringa J, Dehghan A, et al. Normal thyroid function and the risk of atrial fibrillation: the Rotterdam study. *J Clin Endocrinol Metab* 2015;100:3718–24.
14. Pfister R, Brägelmann J, Michels G, et al. Performance of the CHARGE-AF risk model for incident atrial fibrillation in the EPIC Norfolk cohort. *Eur J Prev Cardiol* 2015;22:932–9.
15. Himmelreich JCL, Lucassen WAM, Harskamp RE, et al. CHARGE-AF in a national routine primary care electronic health records database in the Netherlands: validation for 5-year risk of atrial fibrillation and implications for patient selection in atrial fibrillation screening. *Open Heart* 2021;8:e001459.
16. Lowres N, Olivier J, Chao T-F, et al. Estimated stroke risk, yield, and number needed to screen for atrial fibrillation detected through single time screening: a multicountry patient-level meta-analysis of 141,220 screened individuals. *PLoS Med* 2019;16:e1002903.
17. Gladstone DJ, Wachter R, Schmalstieg-Bahr K, et al. Screening for atrial fibrillation in the older population: a randomized clinical trial. *JAMA Cardiol* 2021;6:558-567.
18. Pessoa-Amorim G, Casadei B, Jones NR. Active monitoring for atrial fibrillation (AMALFI): protocol and pilot from a mail-based randomized trial of screening for subclinical atrial fibrillation in high-risk individuals. *ESC Heart & Stroke* 2020;9.
19. Hillege HL, Fidler V, Diercks GFH, et al. Urinary albumin excretion predicts cardiovascular and noncardiovascular mortality in general population. *Circulation* 2002;106:1777–82.
20. Vermond RA, Geelhoed B, Verweij N, et al. Incidence of atrial fibrillation and relationship with cardiovascular events, heart failure, and mortality: a community-based study from the Netherlands. *J Am Coll Cardiol* 2015;66:1000–7.
21. Diercks GF, van Boven AJ, Hillege HL, et al. Microalbuminuria is independently associated with ischaemic electrocardiographic abnormalities in a large non-diabetic population. The PREVEND (prevention of renal and vascular endstage disease) study. *Eur Heart J* 2000;21:1922–7.
22. Schnabel RB, Aspelund T, Li G, et al. Validation of an atrial fibrillation risk algorithm in whites and African Americans. *Arch Intern Med* 2010;170:1909–17.
23. Brieger D, Amerena J, Attia JR, et al. National heart Foundation of Australia and cardiac Society of Australia and New Zealand: Australian clinical guidelines for the diagnosis and management of atrial fibrillation 2018. *Med J Aust* 2018;209:356–62.
24. January CT, Wann LS, Calkins H, 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. *J Am Coll Cardiol* 2019;74:104–32.
25. Boriani G, Palmisano P, Malavasi VL, et al. Clinical factors associated with atrial fibrillation detection on Single-Time point screening using a hand-held Single-Lead ECG device. *J Clin Med* 2021;10:729.
26. Maheshwari A, Norby FL, Soliman EZ, 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.
27. Svennberg E, Lindahl B, Berglund L, et al. NT-proBNP is a powerful predictor for incident atrial fibrillation - Validation of a multimarker approach. *Int J Cardiol* 2016;223:74–81.

28 Kokubo Y, Watanabe M, Higashiyama A, *et al*. Development of a Basic Risk Score for Incident Atrial Fibrillation in a Japanese General Population - The Suita Study. *Circ J* 2017;81:1580–8.

29 Marcos EG, Geelhoed B, Van Der Harst P, *et al*. Relation of renal dysfunction with incident atrial fibrillation and cardiovascular morbidity and mortality: the PREVEND study. *Europace* 2017;19:1930–6.

30 Steyerberg EW, Vickers AJ, Cook NR, *et al*. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–38.

31 IBM Corp. *Ibm SPSS statistics for windows*. Armonk, NY: IBM Corp, 2017.

32 R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2019URL. https://www.R-project.org/

33 Poorthuis MHF, Jones NR, Sherliker P, *et al*. Utility of risk prediction models to detect atrial fibrillation in screened participants. *Eur J Prev Cardiol* 2021;28:586–95.

34 Chamberlain AM, Agarwal SK, Folsom AR, *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.

35 Conroy RM, Pyörälä K, Fitzgerald AP, *et al*. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the score project. *Eur Heart J* 2003;24:987–1003.

36 Lloyd-Jones DM, Huffman MD, Karmali KN, *et al*. Estimating longitudinal risks and benefits from cardiovascular preventive therapies among Medicare patients: the million hearts longitudinal ASCVD risk assessment tool: a special report from the American heart association and American College of cardiology. *J Am Coll Cardiol* 2017;69:1617–36.

37 Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017;357:j2099.

38 Collins GS, de Groot JA, Dutton S, *et al*. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* 2014;14:40.

39 Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* 2016;69:245–7.