Prediction of new-onset atrial fibrillation using routinely available clinical data

Ramesh Nadarajah

Student ID Number: 201198558

Submitted in accordance with the requirements for the degree of Doctor of Philosophy (PhD)

University of Leeds

School of Medicine

August 2023

Principal Supervisor: Professor Chris P Gale

Secondary Supervisor: Professor Jianhua Wu

Intellectual Property Statement

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Publications

Part II of this thesis comprises the following publications by the candidate:

- **Nadarajah R**, Alsaeed E, Hurdus B, Aktaa S, Hogg DC, Bates MGD, Cowan JC, Wu J, Gale CP. Prediction of incident atrial fibrillation in community-based electronic health records: a systematic review with meta-analysis. Heart. 2022 Jul 1;108(13):1020-9.
- **Nadarajah R**, Wu J, Arbel R, Haim M, Zahger D, Benita TR, Rokach L, Cowan JC, Gale CP. Risk of atrial fibrillation and association with other diseases: protocol of the derivation and international external validation of a prediction model using nationwide population-based electronic health records. BMJ Open. 2023 Oct 4.
- **Nadarajah R**, Wu J, Hogg DC, Raveendra K, Nakao YM, Nakao K, Arbel R, Haim M, Zahger D, Parry J, Bates C, Cowan JC, Gale CP. Prediction of short-term atrial fibrillation risk using primary care electronic health records. Heart. 2023 Jul 1;109(14):1072-9.
- Wu J*, **Nadarajah R***, Nakao YM, Nakao K, Hogg DC, Raveendra K, Arbel R, Haim M, Zahger D, Cowan JC, Gale CP. Incident cardiovascular, renal, metabolic diseases and death in individuals identified for risk-guided atrial fibrillation screening: a nationwide cohort study. Open Heart. 2023 Jul 1;10(2):e002357.
- **Nadarajah R**, Wahab A, Reynolds C, Raveendra K, Askham D, Dawson R, Keene J, Shanghavi S, Lip GYH, Hogg DC, Cowan JC, Wu J, Gale CP. Future Innovations in Novel Detection for Atrial Fibrillation (FIND-AF): Pilot study of an electronic health record machine learning algorithm-guided intervention to identify undiagnosed atrial fibrillation. Open Heart. 2023 Sep 1;10(2):e002447.
- Wu J*, **Nadarajah R***, Nakao YM, Nakao K, Arbel R, Haim M, Zahger D, Lip GYH, Cowan JC, Gale CP. Risk calculator for incident atrial fibrillation across a range of prediction horizons. Under review with JAMA Network Open.

*joint first authors

Acknowledgements

I am indebted to my supervisors Professor Chris P Gale and Professor Jianhua Wu. They have been fantastic mentors, and I sincerely hope that they will continue to mentor me for years to come.

Professor Gale is an inspiration as a clinical academic in how he conducts himself and treats others, and he has directed me towards countless exciting new avenues and opportunities. I will always be grateful that I have had the opportunity to work so closely with him over these last few years.

Professor Wu has been a fantastic supervisor and unfailingly generous with his time and expertise. He has an incredible ability to translate complex statistical approaches to more accessible principles and inform of how these can be applied practically. Being able to work with him has been an absolute pleasure and privilege.

Catherine Reynolds has been a tremendous colleague and opened my eyes to the way that studies are planned and executed and I will continue to cherish her company and work ethic.

Katie, Natasha, Dad, Mum, Sat, you always support me through all my adventures (and misadventures). You will always be the most important thing in my life, and everything I do is dependent upon you and done for you.

Abstract

Atrial fibrillation (AF) is common and associated with increased risk of stroke, heart failure and death, yet a fifth of AF disease burden is estimated to be undiagnosed. Screening for AF can increase early detection of AF and associated guideline-directed treatment, but is limited by low yields of newly detected AF. A scalable strategy is required to identify high-risk individuals to make screening for AF more efficient. In the United Kingdom (UK), 98% of the population are registered in primary care with a routinely-collected electronic health record (EHR). The aim of my thesis was to design and evaluate a prediction model that estimates risk of new-onset AF using nationwide routinely-collected primary care EHR data.

A systematic review and meta-analysis was completed to establish the current knowledge base and to inform quantitative analysis. Multivariable prediction models developed and/or validated for incident AF in community-based EHRs were summarised and measures of discrimination performance synthesised. Models eligible for meta-analysis demonstrated only moderate discrimination performance and predicted AF risk over a long prediction horizon, which may be less relevant to guiding AF screening. Models developed with machine learning produced stronger prediction performance for new-onset AF than models developed with traditional regression techniques. Knowledge gaps observed in the systematic review were used to formulate the protocol for developing a novel prediction model for new-onset AF.

Studies were conducted using UK primary care EHRs of 2 081 139 individuals aged 30 years and older without a preceding diagnosis of AF or atrial flutter. A prediction model for incident AF within the next 6 months was developed using a Random Forest classifier (Future Innovations in Novel Detection of Atrial Fibrillation, FIND-AF). FIND-AF could be applied to all EHRs in the dataset and demonstrated excellent discrimination performance on internal validation in the holdout testing dataset (area under the receiver operating characteristic curve [AUROC] 0.824, 95% CI 0.814-0.834). Discrimination performance was robust in both men (AUROC 0.819, 95% CI 0.809-0.829) and women (AUROC 0.821, 95% CI 0.810-0.831), and across different ethnic groups (AUROC, White 0.810, 95% CI 0.799-0.821; Asian 0.796, 95% CI 0.693-0.893; Black, 0.801, 95% CI 0.680-0.973; other non-White ethnic minority, 0.805, 95% CI 0.765-0.845; and ethnicity unrecorded 0.823, 95% CI 0.770-0.875).

The EHRs in the testing dataset were then used to determine the association of higher predicted risk of AF and the occurrence of other cardio-renal-metabolic diseases and

death. Cumulative incidence rates were calculated and Fine and Gray's models fitted at 1, 5, and 10 years for nine diseases and death adjusting for competing risks. Higher predicted risk of AF, compared with lower predicted risk, was associated with higher risk of each of the outcomes (hazard ratio [HR], heart failure 12.54, 95% CI 12.08-13.01; aortic stenosis 9.98, 95% CI 9.16-10.87; stroke/transient ischaemic attack 8.07, 95% CI 7.80-8.34; chronic kidney disease 6.85, 95% CI 6.70-7.00; peripheral vascular disease 6.62, 95% CI 6.28-6.98; valvular heart disease 6.49, 95% CI 6.14-6.85; myocardial infarction 5.02, 95% CI 4.82-5.22; diabetes mellitus 2.05, 95% CI 2.00-2.10; chronic obstructive pulmonary disease 2.02, 95% CI 2.00-2.05; and death 10.45, 95% CI 10.23-10.68), including after adjustment for age, sex, ethnicity, and presence of any of the other outcomes at baseline.

Research grant funding was applied for and awarded to conduct a prospective clinical validation study of the performance of FIND-AF. Ethics approval was achieved and a study protocol formulated to implement the algorithm in the UK primary care setting and establish the yield of new AF across risk estimates when electrocardiogram monitoring is conducted.

Parsimonious regression-based prediction models for new-onset AF were also developed and internally validated for prediction horizons extending from 6 months (AUROC 0.803, 95% CI 0.789-0.821) to 10 years (AUROC 0.780, 95% CI 0.777-0.784), with the aim that these can be applied outside of an EHR setting as a web-based app or risk scoring system, and be used to guide both screening and primary prevention interventions for AF.

In conclusion, my PhD has developed and evaluated novel prediction models for new-onset AF using EHR data routinely recorded in primary care. Such an endeavour addresses an unmet clinical need to efficiently guide AF screening at a population level, in the face of unacceptable morbidity when AF is only diagnosed after the first complication. The results of my PhD will not only provide a means to test the effectiveness of a risk-guided AF screening strategy in clinical studies, but also to further characterise individuals with the machine learning-derived EHR phenotype of higher predicted AF risk to determine if this is an actionable target to further improve patient outcomes.

Table of Contents

List of Tables

Part I

Chapter 1

Part II

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Figures

Part I

Chapter 1

Part II

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Glossary

Abbreviation	Definition
AF	Atrial fibrillation
AFl	Atrial flutter
ALP	Alkaline phosphatase
ALT	Alanine transaminase
ANN	Artificial neural network
API	Application programming interface
AS	Aortic stenosis
AST	Aspartate aminotransferase
AUROC	Area under the receiver operating characteristic
BMI	Body mass index
BP	Blood pressure
CABG	Coronary artery bypass grafting
CAD	Coronary artery disease
CART	Classification and regression tree
CHARGE-AF	Cohorts for Heart and Aging Research in Genomic Epidemiology Atrial Fibrillation
CHARMS	CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies
CHS	Clalit Health Services
CI	Confidence interval
CKD	Chronic kidney disease
cMRI	Cardiac magnetic resonance imaging
COPD	Chronic obstructive pulmonary disease
CPRD	Clinical Practice Research Datalink
CRP	C-reactive protein
DM	Diabetes mellitus
DOAC	Direct oral anticoagulant
DTAC	Digital Technology Assessment Criteria
eFI	Electronic Frailty Index
EHR	Electronic health record
EHRA	European Heart Rhythm Association
EMIS	Egton Medical Information Systems
ESC	European Society of Cardiology
EU	European Union
FACM	Fibrotic atrial cardiomyopathy

FHS	Framingham Heart Study
FHS-AF	Framingham Heart Study score for Atrial Fibrillation
FIND-AF	Future Innovations in Novel Detection of Atrial Fibrillation
GRADE	Grading of Recommendations, Assessment, Development and Evaluation
HES APC	Hospital Episode Statistics Admitted Patient Care
HF	Heart failure
HFpEF	Heart failure with preserved ejection fraction
HFrEF	Heart failure with reduced ejection fraction
HR	Hazard ratio
ICD	International Classification of Diseases and related Health Problems
ILR	Implantable loop recorder
IRR	Incidence rate ratio
LA	Left atrium
LVH	Left ventricular hypertrophy
MESA	Multi-Ethnic Study of Atherosclerosis
MHRA	Medicine and Healthcare products Regulatory Agency
MI	Myocardial infarction
ML	Machine learning
MLR	Multivariable logistic regression
MR	Mitral regurgitation
MRI	Magnetic resonance imaging
NHIS-HEALS	National Health Insurance Service Health screening cohort
NHIS-NSC	National Health Insurance Service-based National Sample Cohort
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health and Care Research

NIVEL-PCD	Netherlands Institute for Health Services Research Primary Care Database
NPV	Negative predictive value
NSAID	Non-steroidal anti-inflammatory drug
NT-proBNP	N-terminal pro-B-terminal natriuretic peptide
OAC	Oral anticoagulant
ONS	Office for National Statistics
OR	Odds ratio
OSA	Obstructive sleep apnoea
PAD	Peripheral arterial disease
PCI	Percutaneous coronary intervention
PI	Prediction interval
PPG	Photoplethysmography
PPI	Patient and Public Involvement
PPV	Positive predictive value
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-analyses
PROBAST	Prediction model Risk of Bias ASsessment Tool
QALY	Quality-adjusted life-years
QOF	Quality Outcomes Framework
QoL	Quality of life
RAAS	Renin-angiotensin-aldosterone system
RCT	Randomised clinical trial
RECORD	Reporting of studies Conducted using Observational Routinely-collected Health Data
RF	Random forests
RR	Relative risk
SBP	Systolic blood pressure
SCAF	Subclinical atrial fibrillation
SD	Standard deviation
SNOMED CT	Systematised Nomenclature of Medicine Clinical Terms
THIN	The Health Improvement Network
TIA	Transient ischaemic attack
TRIPOD	Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

TTR	Time in therapeutic ratio
UK	United Kingdom
USA	United States of America
VHD	Valvular heart disease
VKA	Vitamin K antagonist

Part I

Chapter 1 Introduction

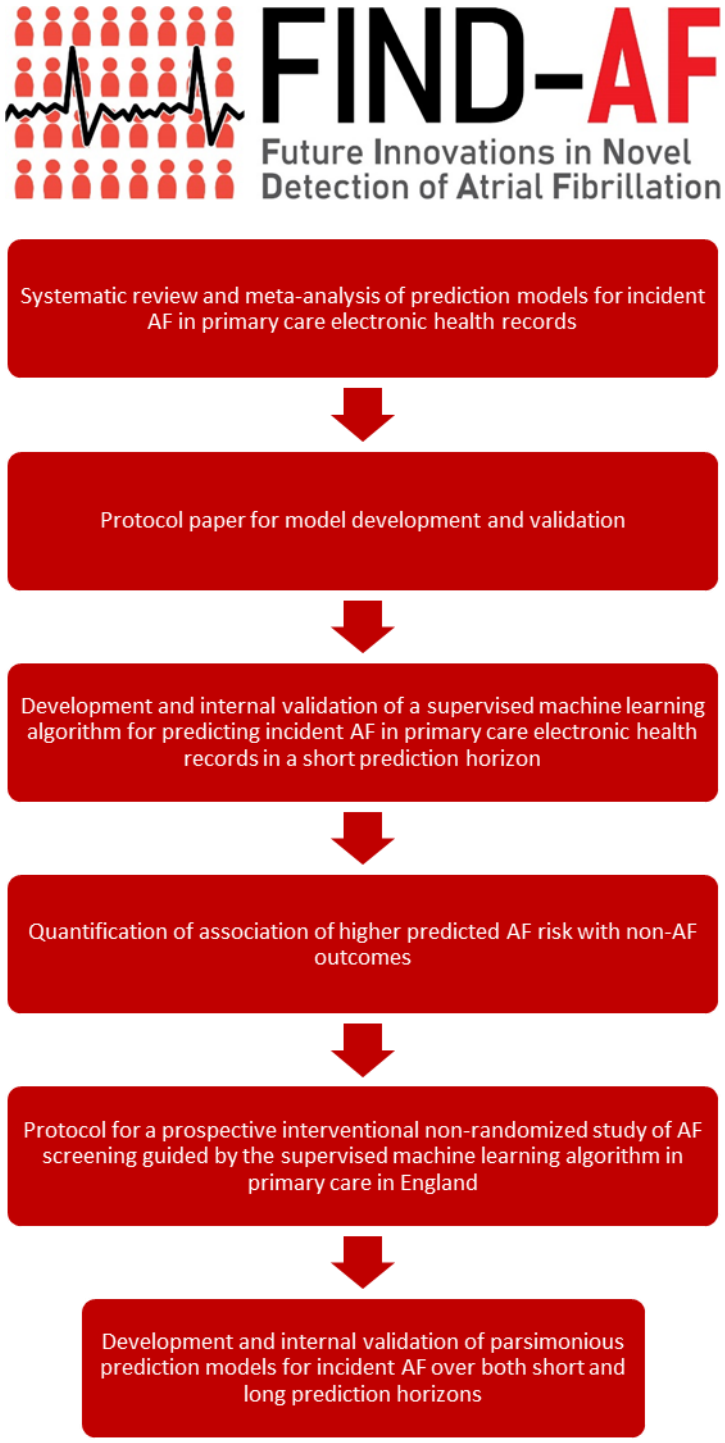
In this thesis I will present the development of prediction models for new-onset atrial fibrillation (AF) using only routinely-collected data available in the community. A systematic review will be conducted to summarise and evaluate prediction models that have previously been developed and/or validated for new-onset AF using data routinely available in the community. This will inform the development of a prediction model through a supervised machine learning technique for use in United Kingdom (UK) primary care electronic health records (EHRs). Furthermore, how risk of AF is associated with occurrence of other diseases and death will be established. Moreover a study will be designed to prospectively clinically validate the prediction model. Such an endeavour aims to inform efforts to improve the clinical and cost-effectiveness of screening for AF. Consideration will also be given to the development of parsimonious models for prediction of new-onset AF over both short and long prediction horizons, which could be utilised outside of an EHR system. The thesis is structured in accordance with the format of an alternative style of doctoral thesis including published material of the University of Leeds.

In Part I, I highlight the epidemiology of AF, complications after it develops, and current guidelines for clinical management. I then summarise the rationale and current evidence for AF screening, the utility of multivariable prediction models for new-onset AF (herein also referred to as incident AF) for guiding AF screening, and the limitations of existing models. Finally, I discuss how AF is associated with the development of a range of diseases and death, which leads to my hypothesis that individuals identified at higher risk of AF may also be at elevated risk of other outcomes.

In Part II, I outline the accomplishments of my PhD studies by presenting papers that have been published, or are under review, with peer-reviewed journals. These papers report: i) the development of prediction models for incident AF, ii) the quantification of the association of predicted AF risk with non-AF outcomes, and iii) the formulation of a protocol for the prospective clinical validation of an AF prediction model.

Part III comprises a critical discussion of the presented material in the context of the literature, with an overview of potential future directions and challenges. Figure 1 provides a central illustration of my PhD studies and accomplishments.

Figure 1 Central illustration of the PhD studies and accomplishments



Abbreviations: AF, atrial fibrillation; EHR, electronic health record

1.1 Atrial Fibrillation

1.1.1 Definition of atrial fibrillation

AF is defined as a supraventricular tachyarrhythmia with uncoordinated atrial electrical activation and consequently ineffective atrial contraction, having the following electrocardiographic characteristics: (i) irregularly irregular R-R intervals (where atrioventricular conduction is not impaired), (ii) absence of distinct repeating P waves, and (iii) irregular atrial activations.[1] Five patterns of AF are distinguished according to the 2020 European Society of Cardiology (ESC) guidelines,[1] based on presentation, duration, and mode of termination of AF episodes (Table 1).[2]

Table 1 Classification of atrial fibrillation according to the 2020 European Society of Cardiology Guidelines

AF pattern	**Definition**
First diagnosed	AF not diagnosed before (also called new-onset AF)
Paroxysmal	AF that terminates spontaneously or with intervention within 7 days of onset
Persistent	AF that is continuously sustained beyond 7 days, including episodes terminated by cardioversion (drugs or electrical cardioversion) after > 7 days
Long-standing persistent	Continuous AF of >12 months, when deciding to adopt a rhythm control strategy.
Permanent	AF that the patient and physician accept, and no further attempts to restore/maintain sinus rhythm will be undertaken

Abbreviations: AF, atrial fibrillation

1.1.2 Pathophysiology and atrial remodelling

AF is characterised with rapid and uncoordinated atrial electrical activity.[2] Electrical re-entry following triggered activity upon a vulnerable substrate precipitates AF.[3] Early afterdepolarisations and delayed afterdepolarisations are the main form of triggered activity.[4] Ectopic beats originating from the pulmonary vein have been identified as the initiating trigger in paroxysmal AF.[5] Anatomic re-entry is mainly due to focal structural changes and fibrosis, while functional re-entry is related to reduction of conduction velocity or reduction of the effective refractory period in the atrial myocardium.[4, 6, 7] Pathological stimuli such as inflammatory and oxidative stress (diabetes mellitus [DM],

obesity and renal failure) as well as volume and pressure overload (hypertension, left ventricular diastolic and systolic dysfunction, and valvular heart disease [VHD]) result in left atrial (LA) structural and functional remodelling promoting AF.[4, 8]

1.1.3 Incidence and prevalence of atrial fibrillation

AF is the most common sustained arrhythmia,[1] and its incidence and prevalence have risen over recent decades. Globally data suggests that the incidence rate has increased by 31% between 1997 and 2017, from 309 new cases per million inhabitants to 403 new cases per million inhabitants. Over this time period the prevalence of AF has also increased by 33%, to over 37 million cases.[9] In the European Union (EU), the number of adults aged over 55 years with AF is projected to rise from 8.8 million in 2010 to 17.9 million in 2060.[10]

The incidence of AF in the United Kingdom is one of the highest globally, and is on the rise. Earlier work that I was involved in, and is outwith of this thesis, demonstrated that that age- and sex-standardised incidence of AF increased by 30% from 1998 to 2017 (322 per 100 000 person-years vs 247 per 100 000 person-years; adjusted incidence rate ratio [IRR] 1·30, 95% CI 1·27 - 1·33), and that crude incidence increased by 47% from 250 per 100 000 people in 1998 to 367 per 100 000 people in 2017.[11] We estimated that the absolute number of yearly new diagnoses of AF had increased by 72% in 2017 compared to 1998 (202 333 vs 117 880). Notably, the total number of new AF cases diagnosed each year in England (202 333) outstripped the combined total number of cases of breast, prostate, lung and bowel cancer in 2021 (199 608).[12] Thus AF is an emerging public health crisis.

1.1.4 Complications of atrial fibrillation

1.1.4.1 Stroke and systemic emboli

In the Framingham Heart study (FHS) patients with non-rhuematic AF, compared to those without, have a five-fold higher risk of stroke and systemic embolism.[13] In stroke registries, at least a third of patients with ischemic stroke have either previously known,[14, 15] or newly detected AF at the time of stroke.[16] In more than 25% of AF-related strokes, stroke is the first manifestation of AF.[14]

There is evidence that AF is associated with a state of blood stasis, endothelial dysfunction and clotting activation, thus fulfilling Virchow’s triad of criteria for thrombus

formation.[17, 18] The formation of thrombus in the fibrillating atria leads to the potential of embolism, which may occlude a distal blood vessel. In the brain, this causes cerebral ischaemia, and potentially infarction.[18, 19]

Extracranial systemic embolism is much less common than stroke in patients with AF.[20] In a pooled analysis of four AF antiplatelet and anticoagulation randomised clinical trials (RCTs) including 37 973 patients from more than 40 countries, 221 systemic embolic events occurred during a mean follow-up of 2.4 years, representing 11.5% of total clinically apparent embolic events.[21] Systemic embolic events occurred mostly in the lower limbs (58%) and the mesenteric circulation (22%).[21]

1.1.4.2 Cognitive impairment and dementia

Cognitive decline is strongly linked to AF, and both are associated with advanced age.[22] Compared with controls, the relative risk of cognitive decline for patients with AF is up to 1.8-fold higher.[23-25] AF is associated with a heightened risk of cognitive impairment, dementia, Alzheimer's disease, and vascular dementia, independently of common risk factors for dementia such as age, hypertension, DM, obesity, or stroke.[26] A multi-factoral mechanism underlies this association involving hypoperfusion, activation of the inflammatory and coagulative systems, endothelial injury and circulatory stasis promoting thrombogenicity, resulting in covert thromboembolism, micro-thromboembolism and white matter T2 hyperintense lesions on brain magnetic resonance imaging (MRI).[20, 22]

1.1.4.3 Heart failure

In the FHS the incidence of first-diagnosed heart failure (HF) in patients with AF was 33 per 1000 person-years.[13] AF and HF frequently co-exist, in part due to shared risk factors such as hypertension, DM, coronary artery disease (CAD), and VHD.[20] Individuals with AF are at higher risk of developing both heart failure with preserved and reduced ejection fraction (HFpEF and HFrEF),[27] and the presence of AF is part of the H_2FPEF risk scoring system for the diagnosis of HFpEF.[28] In fact, in the FHS the risk of incident HFpEF (hazard ratio [HR] 2.34, 95% CI 1.48–3.70) was higher than that of HFrEF (HR 1.32, 95% CI 0.83–2.10).[27]

1.1.4.4 Coronary artery disease

In the REGARDS study, AF was associated with a two-fold increased risk of myocardial infarction (MI) (HR 1.96, 95% CI 1.52-2.52),[29] with a greater risk in women (HR 2.16, 95% CI 1.41–3.31) compared with men (HR 1.39, 95% CI 0.91–2.10). AF and MI share similar risk factors, and therefore, common pathophysiologic processes might drive both outcomes. However, there could be AF-specific mechanisms that could lead to MI. For example, it has been demonstrated that AF creates and sustains an inflammatory and prothrombotic environment,[30] with systemic platelet activation, thrombin generation and endothelial dysfunction.[31]

1.1.4.5 Death

In a systematic review and meta-analysis of 64 studies including over a million patients, (149 746 [14.8%] of them having AF), AF increased the risk of death by 46% (pooled relative risk [RR] 1.46, 95% CI 1.39–1.53), with a greater increase in the risk of death from a cardiovascular cause (RR] 2.03, 95% CI: 1.79–2.30).[32]

1.1.5 Management of atrial fibrillation

Clinical management of patients with AF is based on the structured characterisation of AF (the 4S-AF scheme) recommended in the 2020 ESC guidelines.[1] The 4S-AF scheme addresses four specific domains in AF: Stroke risk, Symptom severity, Severity of AF burden, and Substrate severity.

1.1.5.1 Stroke risk and prevention

The main strategies for stroke prevention include oral anticoagulation - with either vitamin K antagonists (VKAs) or direct oral anticoagulants (DOACs) - and percutaneous LA appendage occlusion.[33] Oral anticoagulation can reduce rates of stroke by 64%,[34] with DOACs slightly more effective and much safer pertaining to the bleeding risk, than the VKA warfarin.[35] The ESC and National Institute for Health and Care Excellence (NICE) guidelines recommend that that the decision whether or not to commence an oral anticoagulant (OAC) in people with non-valvular AF should be based upon an objective stroke-risk scoring system, specifically the CHA_2DS_2-VASc score (Table 2).[1, 36] It is recommended that patients with low risk of stroke (CHA_2DS_2-VASc 0 in men, 1 in women) should not receive an OAC for stroke prophylaxis. For men with a CHA_2DS_2-VASc of 1 and women with a CHA_2DS_2-VASc of 2, it is recommended that OACs should be considered, especially if age is the contributing risk factor. In men with a CHA_2DS_2-VASc ≥2 and in woman with a CHA_2DS_2-VASc ≥3 treatment with OAC is recommended.

Table 2 Assessment of stroke risk using CHA_2DS_2-VASc

Criteria	Value	Clarification	Points
Age	<65 years old		0
	65-74 years		1
	≥75 years		2
Sex	Men		0
	Women		1
Congestive heart failure history	Yes/no	Clinical HF, or objective evidence of moderate to severe left ventricular systolic dysfunction, or HCM	1
Hypertension history	Yes/no	or any antihypertensive therapy	1
Stroke/TIA/systemic embolism history	Yes/no		2
Vascular disease history	Yes/no	Angiographically significant CAD, previous MI, PAD, or aortic plaque	1
Diabetes mellitus history	Yes/no	Treatment with oral hypoglycaemic drug and/or insulin or fasting blood glucose >125mg/dl	1

Abbreviations: CAD, coronary artery disease; HCM, hypertrophic cardiomyopathy; HF, heart failure; MI, myocardial infarction; PAD, peripheral arterial disease; TIA, transient ischaemic attack

Both ESC and NICE guidelines recommend that bleeding risk should be assessed, and that risk factors for bleeding should be modified alongside a decision to commence OAC, but that a high bleeding risk should not generally result in withholding OAC.[1, 36] NICE guidelines currently recommend bleeding risk assessment with the HAS-BLED

score (Table 3). In patients that are unable to take an OAC because it is contraindicated or not tolerated, a LA appendage occlusion device is a potential option.

Table 3 Assessment of bleeding risk using HAS-BLED

Criteria	Value	Clarification	Points
Hypertension	Yes/no	SBP>160mmHg	1
Abnormal renal and/or hepatic function	Yes/no	Dialysis, transplant, serum creatinine >200 mmol/L, cirrhosis, bilirubin > 2x upper limit of normal, AST/ALT/ALP >3x upper limit of normal	1 point for each
Stroke	Yes/no	Previous ischaemic or haemorrhagic stroke	1
Bleeding history or predisposition	Yes/no	Previous major haemorrhage or anaemia or severe thrombocytopenia	1
Labile INR	Yes/no	TTR<60% Only relevant in patients receiving VKA	1
Elderly	Yes/no	Age >65 years or extreme frailty	1
Drugs or excessive alcohol drinking	Yes/no	Concomitant use of antiplatelet or NSAID; and/or excessive alcohol per week (>14 units per week)	1

Abbreviations: ALP, alkaline phosphatase; ALT, alanine transaminase; AST, aspartate aminotransferase, NSAID, non-steroidal anti-inflammatory drug; SBP, systolic blood pressure; TTR, time in therapeutic ratio; VKA, vitamin K antagonist

It is worth noting that none of the DOACs are currently recommended for patients with 'valvular AF' (usually considered as those with moderate-severe mitral stenosis or a mechanical heart valve).[1, 36] All individuals with a mechanical heart valve or moderate-severe mitral stenosis are recommended to be offered oral anticoagulation with a VKA irrespective of CHA_2DS_2-VASc score.[1]

1.1.5.2 Symptom severity

AF symptomatology varies greatly. About one-third of patients with AF are asymptomatic, whereas others experience highly symptomatic and disabling symptoms resulting in poor quality of life (QoL).[1] Symptoms can include palpitations, dyspnoea, fatigue, chest pain, poor effort tolerance, dizziness, syncope, and disordered sleeping.[1] Symptom severity is stratified according to the European Heart Rhythm Association (EHRA) symptom score (Table 4).

Table 4 The European Heart Rhythm Association symptom score

Score	Symptoms	Description
1	None	AF does not cause any symptoms
2a	Mild	Normal daily activity not affected by symptoms related to AF
2b	Moderate	Normal daily activities not affected by symptoms related to AF, but patient troubled by symptoms
3	Severe	Normal daily activity affected by symptoms related to AF
4	Disabling	Normal daily activity discontinued

Abbreviations: AF, atrial fibrillation

1.1.5.3 Severity of atrial fibrillation burden

The decision for rate or rhythm control depends on the severity of AF as defined by its temporal pattern and the severity of symptoms.

In the rate control strategy, the target heart rate of less than 80 beats per minute at rest and less than 110 beats per minute at moderate exercise can be achieved according to current recommendations using beta-receptor blockers, non-dihydropyridine calcium channel blockers, digoxin or amiodarone; or atrioventricular node ablation when pharmacological therapy fails.[1]

The rhythm control strategy aims at the restoration and maintenance of sinus rhythm by electrical cardioversion, antiarrhythmic medication or catheter ablation.[1] The most recent ESC guidelines recommend rate control as:[1]

- Background therapy in all patients with AF
- First choice therapy in patients with no or minor symptoms
- Therapy after failure of rhythm control
- Therapy when risks of restoring sinus rhythm outweighs benefits

And rhythm control for:[1]

- Symptom and quality of life improvement in symptomatic patients with AF
- Patients with heart failure
- Patients aged less than 65 years, or those who have daily activities requiring optimal cardiac performance

1.1.5.4 Substrate severity

Assessment of atrial cardiomyopathy, using transthoracic or transoesophageal echocardiography, cardiac magnetic resonance imaging (cMRI) and cardiac computed tomography, is crucial in AF management.[1] Obesity, physical inactivity, obstructive sleep apnea (OSA), DM, hypertension, dyslipidaemia, alcohol abuse, and smoking are considered risk factors for the development and progression of AF.[37] Most of these AF risk factors can potentially be reversed or controlled, and evidence supports that addressing these modifiable risks contribute to secondary AF prevention.[37, 38] Patients with AF who have comprehensively managed their risk factors demonstrate greater reduction in symptoms, reduction in AF burden, and more successful ablations.[39-41] Risk factor management is now integrated into the ESC guidelines as an additional pillar of AF management.[1]

1.1.6 Economic burden of atrial fibrillation

The incremental cost of AF at a national scale in the United States of America (USA), comparing patients with AF and matched controls without AF, is estimated to be $26 billion.[42] In the UK, it is estimated that the direct and proportion of National Health Service (NHS) expenditure for AF in 2030 will be between £2 351 million (1.11%) and £5 562 million (2.63%), with nearly 60% of that cost related to primary admissions.[43]

1.1.7 Summary

- AF is reaching epidemic proportions.
- AF is associated with a range of adverse outcomes, including stroke, heart failure, and death.
- There are effective evidence-based treatments and structured pathways for the management of AF.
- For many patients the first presentation of AF is with a complication.
- Thus the early diagnosis of AF, before the manifestation of the first complication, remains a major public health challenge.

1.2 Aims and Objectives

In this thesis I will develop a prediction model for incident AF using routinely-collected data in primary care EHRs, establish how risk of AF relates to outcomes beyond AF, and establish a protocol for prospective validation of the model in a clinical study.

1.2.1 Objectives

1. Establish whether prediction models for incident AF have been developed and/or validated in primary care EHRs, summarise the techniques used to develop these models, and evaluate the performance of the models.
2. Develop a prediction model for incident AF using a supervised machine learning (ML) approach with routinely-collected data in a primary care EHR database.
3. Establish whether individuals identified as higher risk of AF by the ML algorithm are also at increased risk for other outcomes.
4. Design a study to prospectively clinically validate the ML algorithm for prediction of incident AF.
5. Develop parsimonious prediction models for new-onset AF using traditional regression techniques with routinely-collected data from a primary care EHR database.

1.2.2 Research questions

1. Are there prediction models for incident AF that have been derived or validated in primary care EHRs?
2. What is the reported performance of prediction models for incident AF in primary care EHRs?
3. What prediction model development techniques have been used in EHR databases to predict incident AF?
4. Will a supervised machine learning algorithm demonstrate better prediction performance than a traditional regression technique for incident AF in primary care EHR data? And, if so, will it also outperform previously developed and/or validated prediction models?
5. Is it possible to accurately predict incident AF over a short prediction horizon (6 months) using a primary care EHR dataset?
6. Is it possible to accurately predict incident AF when restricting predictors to age, sex, ethnicity and comorbidities?
7. What is the appropriate clinical study design to prospectively validate a prediction model for incident AF?
8. What is the association between higher risk of AF and occurrence of cardio-renal-metabolic diseases and death?
9. Is it possible to develop parsimonious prediction models for AF that can be optimised for both short and long prediction horizons whilst using the same variables?

1.3 Screening for atrial fibrillation

Diagnosing AF earlier in the disease trajectory could lead to initiation of effective therapy, including OACs to reduce stroke and death.[44] Accordingly screening for AF has been suggested as one strategy to increase AF detection.[45]

1.3.1 Screening for disease

The Commission on Chronic Illness Conference on Preventative Aspects of Chronic Disease, held in 1951, defined screening as “the presumptive identification of unrecognised disease or defect by the application of tests, examinations, or other procedures which can be applied rapidly”.[46] The UK National Screening Committee define screening as “a process of identifying apparently healthy people who may be at risk of a disease or condition…they can be offered information, further tests, appropriate treatment to reduce their risk and/or any complications arising from the disease or condition.[47] Accordingly, screening will involve testing of people who either do not have or have not recognised the signs or symptoms of the condition being tested for (that is, they believe themselves not to have the condition the screening relates to), where the purpose is to reduce risk for that individual of future ill health related to that condition.[47] In the UK NHS several large-scale population screening programmes exist, including abdominal aortic aneurysm screening to men during the year they turn 65, and bowel cancer screening to people aged 60 to 74 every two years.[48] In this Section I summarise the current evidence base for screening for AF.

1.3.2 Burden of undiagnosed atrial fibrillation

Estimates of the prevalence of undiagnosed AF are often based on patient screening studies. In a Spanish multi-centre, population-based, retrospective, cross-sectional, observational study, 1 043 participants over the age of 60 years were randomly selected to undergo an ECG in a pre-arranged appointment.[49] Amongst the participants, with a mean age of 78.9 years, 91 (8.7%) were already known to have AF, but 23 (2.2%) were found to have AF on the ECG that was previously undiagnosed. During the Belgian Heart Rhythm Week, volunteers aged 40 years and older were invited to participate in a free screening programme via flyers and via a media campaign.[50] Of 10 758 participants screened using a one-lead ECG hand-held monitor 228 participants had AF diagnosed at the time of screening, representing a prevalence of 2.2% (95% CI 1.3%-3.0%) in the screened population.

One report from the USA applied a non-parametric back-calculation methodology to estimate the prevalence of undiagnosed AF.[51] Based on measuring the incidence of stroke in a retrospective cohort of health insurance claims data from Medicare between 2004 and 2010, the authors then back-calculated total AF prevalence based on the attributable risk of AF to stroke. Based on this method the authors estimated that the total prevalence (undiagnosed and diagnosed) of AF in the USA was estimated to be 5 331 000 (2.4% of adults), with 698 900 (13.1%) undiagnosed (without a diagnostic code of AF in their claims data). In the UK, a report has estimated that 305 262 individuals in the UK have undiagnosed AF.[52] The authors found that the prevalence of AF in the North West London Whole Systems Integrated Care data warehouse in 2019 was 3.0% (17 800 of 604 135). The authors used the National Cardiovascular Intelligence Network method to calculate that the national total AF prevalence in 2019 was 1 480 221 but the number of patients registered as having an AF diagnosis on Quality Outcomes Framework (QOF) in 2019 was 1 174 949.

1.3.3 Risk of stroke and death in untreated screen-detected atrial fibrillation

No data specifically address the risk of stroke and death in untreated screen-detected AF in the general population. The closest approximation includes cohort studies of individuals with AF detected incidentally in the absence of symptoms. In a cohort from Olmstead County, Minnesota individuals who were asymptomatic at presentation were three times as likely to have had an ischemic stroke before AF diagnosis, and in follow-up they had similar risk of stroke and death as those with symptomatic AF.[53] In a later study individuals with asymptomatic AF at presentation had an increased risk for cardiovascular (HR, 3.12, 95% CI 1.50–6.45) and all-cause mortality (HR 2.96, 95% CI 1.89–4.64) compared to those with typical symptoms after adjustment for CHA_2DS_2-VASc score and age.[54] In 5 555 patients with asymptomatic AF detected incidentally in general practice, the adjusted stroke rate in the 1 460 untreated patients was 4% and all-cause mortality 7% over 1.5 years of follow-up compared with 1% and 2.5%, respectively, in matched controls without AF.[55, 56] Overall this data suggests that ‘screen-detected’ AF, that is, AF found on a single-timepoint ECG incidentally in the absence of symptoms, is not a benign condition and carries a significant risk of adverse outcomes.[57]

1.3.4 Response to treatment of screen-detected atrial fibrillation

Screening for a particular disease implies that an effective therapy improves outcomes. For AF, OACs have a major impact on reducing stroke, systemic embolism, and all-cause mortality.[35] In the cohort study of 5 555 asymptomatic patients with AF detected incidentally in general practice, OAC therapy (n = 2 492) compared with no antithrombotic therapy (n = 1 460) was associated with significantly reduced adjusted risk of stroke from 4% to 1% and death from 7% to 4% in only 1.5 years, suggesting that screen-detected AF may respond similarly.[55, 56]

1.3.5 Current diagnostic pathways for atrial fibrillation in the United Kingdom National Health Service

Recommendations from NICE for diagnosis of AF centre around the situation where symptoms prompt the patient to present to healthcare services.[36] Patients may also present with a complication of AF, such as heart failure or stroke, as AF is commonly asymptomatic.[58] During a clinical examination, palpation of the pulse may reveal an irregularly irregular rhythm, which would raise suspicion for AF, and prompt evaluation with a 12-lead ECG.[36] If there is a suspicion of paroxysmal AF, a more prolonged period of ECG monitoring may be required to detect an episode such as ambulatory ECG monitoring (which records a prolonged surface ECG), an event recorder (which is activated by the patient when symptoms occur), or an implantable loop recorder (ILR).[1, 36]

In the UK NHS there is no specific early identification or screening pathway for patients with asymptomatic undiagnosed AF before a complication. The current NHS recommended practice for these patients is for a healthcare professional to palpate a person's pulse during a NHS health check or during blood pressure checks to try to detect an irregular rhythm. However, this approach has low sensitivity (87%) and specificity (70%) for identifying an individual likely to have AF.[1] Moreover, even if an irregular pulse is noted the person then requires an ECG to confirm the diagnosis – necessitating a further investigation at another appointment when AF may not be apparent.[59] Furthermore, fewer than half of eligible people participate in NHS health checks and many patients with AF do not have hypertension.[60] As such, there are many circumstances where people with AF will fail to be diagnosed and treated to prevent stroke during routine clinical practice.

1.3.6 Screening methods for atrial fibrillation detection

Screening methods for AF include pulse palpation,[61] automated blood pressure monitors,[62, 63] watches,[64, 65] smartphone applications,[66] single-lead ECG recorders,[67] continuous ECG patches,[68] long-term holter monitoring, and wearable belts for ECG recording.[69] Patient-activated ECG recorders (Figure 2) can also be effective in asymptomatic individuals if regular ECG recording is performed at the predefined time (e.g. twice daily in the STROKESTOP RCT).[67, 70]

Figure 2 The Zeincor hand-held patient-activated lead-I electrocardiogram recorder, reproduced with permission from Zenicor®

Continuous ECG devices are available with a recording time from 24 hours to several weeks. Continuous ECG has a higher diagnostic yield than corresponding intermittent ECG, but continuous ECG is limited by the risk of skin irritation which can affect compliance.[71] ILRs are small devices which are inserted subcutaneously on the chest and are used for long-term ECG event recording, even up to several years, and associated with yields of up to 30% amongst high-risk individuals.[72] The use of ILR in AF screening at scale is limited by the invasive procedure needed for implantation, the high cost for devices, and the high workload associated with adjudication of long-term monitoring.[71]

Smartwatches and other ‘wearables’ can passively measure pulse rate from the wrist using an optical sensor for photoplethysmography (PPG) and alerting the individual wearing the device of a pulse irregularity (based on a specific algorithm for AF detection analysing pulse irregularity and variability). The current AF definition omits to

allow a diagnosis of AF based on PPG.[1] Thus the use of these tools in AF screening requires further ECG confirmation in individuals suspected to have AF.

As there is no diagnostic test with 100% sensitivity or specificity (Table 5), the screening process will result in false-positive and false-negative cases. For most individuals with risk factors for stroke, the risk of having AF without OAC treatment is higher than the risk of having OAC treatment without an AF diagnosis.[71] Hence a missed diagnosis constitutes a higher risk than a falsely positive diagnosed AF, and a high sensitivity is very important for the screening test. On the other hand, if the disease prevalence is low in the screened population, the proportion of false positives will be of growing importance in the balance between sensitivity and specificity.[71] For example, with an untreated disease prevalence of 5% and sensitivity and specificity of 95%, a screening of 1 000 individuals will result in 48 true positives, 2 false negatives, 902 true negatives and 48 false positives.[71] Thus the population identified for screening is of critical importance.

Table 5 Sensitivity and specificity of various atrial fibrillation screening tools considering the 12-lead electrocardiogram as the gold standard

	Sensitivity	**Specificity**
Pulse taking[73]	87-97%	70-81%
Automated BP monitors[74, 75]	93-100%	86-92%
Single lead ECG[76, 77]	94-98%	76-95%
Smartphone apps[78, 79]	91.5-98.5%	91.4-100%
Watches[80, 81]	97-99%	83-94%

Abbreviations: AF, atrial fibrillation; BP, blood pressure; ECG, electrocardiogram

1.3.7 Screening strategies for atrial fibrillation

The two strategies used in AF screening are systematic and opportunistic screening. In systematic screening, an entire population or a stratum of a population is targeted for screening. This is the equivalent of NHS breast screening, which is offered every three years to women between the ages of 50 and 71.[82] Opportunistic screening is a strategy in which the participant is offered screening during a healthcare visit not caused by a suspicion of the screened disease.

Systematic screening presents a robust approach to aiming to investigate as much of the population as possible. However it is more expensive and requires a new pathway. It also involves a screening invitation, which has been demonstrated to introduce bias, resulting in lower participation rates amongst individuals with lower socio-economic status, longer distance to the screening site, and more comorbidities.[70]

Opportunistic screening has several advantages.[71] First, it will use the existing structure of the healthcare system and there is no need to organize a separate system for screening examinations. Second, patients with chronic diseases associated with AF development often have regular healthcare contacts, which will give enrichment to the screening process. Third, a participant could have particular confidence for the screening examination offered by their general practitioner, with whom they have regular contact.[83] However there are drawbacks. Individuals never visiting healthcare facilities will not be offered screening in this setting. It is further possible that the pre-existing workload of healthcare professionals will limit the screening capacity, both in terms of performing a test and in terms of handling positive findings. Moreover, accuracy of ECG reading has been shown to be variable in primary care.[84]

Following the advent of heart rhythm recording devices for consumers – mainly smartwatches, smartphones, wearables and handheld units – consumer-initiated AF screening has become increasingly prevalent. The availability of these devices makes it possible for the user to make their own heart rhythm investigation without involving healthcare services. The increased availability could give some advantages such as increased detection, but there are also several drawbacks to this development.[71] First, many of these devices do not record ECG but PPG, and any suspicion of arrhythmia must be confirmed using ECG. Second, none of the automated interpretations algorithms in the devices have a specificity of 100%, and many users will get a false-positive notification of arrhythmias, which could cause unnecessary worries to the user and further investigations consuming healthcare resources. Third, the groups where AF

prevalence will be higher - that is the elderly and those with cardiovascular comorbidities - are less often users of these types of devices. Finally, in publicly financed healthcare systems, there is a risk that consumer-initiated screening to some extent will displace other patient groups in the competition for healthcare resources.[71] In the Apple Heart Study and the Huawei heart study the mean age of the participants was low (35 and 41 years, respectively), and the AF yield was similarly low, at 0.09% and 0.04%, respectively.[64, 85]

1.3.8 Clinical effectiveness of screening for atrial fibrillation

1.3.8.1 Effect of screening for atrial fibrillation on detection rates of atrial fibrillation compared with routine care

1.3.8.1.1 Opportunistic screening

Randomised clinical trial data suggests that opportunistic screening for AF does not increase detection rates of AF compared with routine care in contemporary practice. The SAFE study was the first randomised AF screening trial at scale, starting in 2001,[86] and was to designed to determine the most cost-effective method of screening for AF in the population aged 65 years and over, using a single time-point ECG. It was set in general practice, using 25 practices for intervention and 25 practices for control. In the intervention practices, patients were randomly allocated to systematic (n = 5 000) or opportunistic (n = 5 000) screening. AF screening was performed using pulse taking followed by an ECG recording in cases with irregular pulse. In both systematic and opportunistic arms, AF detection was higher (1.63%) in the screened population compared to the control population (1.04%), and similar proportions of patients with new AF were detected using opportunistic or systematic approach.

The D2AF and VITAL-AF trials investigated opportunistic screening for AF in primary care in the contemporary era. The D2AF trial was a cluster RCT involving 47 intention-to-screen and 49 usual care primary care practices in the Netherlands. In each practice a fixed sample of 200 eligible patients were randomly selected. Opportunistic screening consisted of three index tests: pulse palpation, electronic blood pressure measurements with an AF algorithm, and ECG with a single lead hand-held ECG device. Detection of new AF was not significantly different between the intervention and control arm (144/8 874 patients in intervention arm [1.62%] vs 139/9 102 patients in the control arm [1.53%]; adjusted odds ratio [OR] 1.06, 95% CI 0.84-1.35). In the VITAL-AF trial 16 primary care clinics were randomised 1:1 to AF screening using a handheld single-lead ECG (AliveCor KardiaMobile) during vital sign assessments or usual care.

All confirmatory diagnostic testing and treatment decisions were made by the primary care clinician. New AF diagnoses over one-year follow-up were ascertained electronically and manually adjudicated. Of 30 715 patients without prevalent AF (n=15 393 screening [91% screened], n=15 322 control), 1.72% of individuals in the screening group had new AF diagnosed at one year versus 1.59% in the control group (risk difference 0.13%, 95% CI -0.16–0.42, P=0.38). Overall, because the rate of AF diagnosis in routine care has increased over the last 20 years,[11] these trials suggest that opportunistic screening for AF in primary care may not be useful when applied broadly to patients aged 65 years and older.

1.3.8.1.2 Systematic screening

By contrast, RCTs of systematic screening for AF, using either continuous or intermittent non-invasive ECG monitoring devices, guided by age or stroke risk, have demonstrated increased detection rates for AF compared to routine care (Table 6).[67, 68, 87, 88] Furthermore, a high proportion of individuals diagnosed with AF during systematic AF went on to be treated with OACs (Table 6).

Table 6 Detection rates for new atrial fibrillation in the intervention arm of systematic atrial fibrillation screening randomised clinical trials using non-invasive devices

Study	**Year**	**Inclusion criteria**	**N**	**AF detection protocol**	**Follow-up period**	**New AF detection rate (%)**	**OAC initiation in newly diagnosed AF cases (%)**
STROKESTOP (Sweden)	2015	Age 75-76 years	7 173	12-lead ECG, then 2-week single-lead handheld ECG recorder (Zenicor) twice daily	2 weeks	0.5 (initial assessment) 3.0 (2 week)	93
STROKESTOP II (Sweden)	2016	Age 75-76 with a NT-proBNP >125	3 766	12-lead ECG, then 2-week single-lead handheld ECG recorder (Zenicor) four times daily	2 weeks	0.5 (initial assessment) 4.4 (2 week)	94.5
REHEARSE-AF (UK)	2017	Age >65-years with a CHA_2DS_2-VASc score ≥2	501	Single-lead handheld recorder (AliveCor) twice weekly for 12 months	12 months	3.9	100
mSToPS (USA)	2018	Age ≥75 years, or a man ≥55 years or woman ≥65 years with one or more of the following comorbidities: • Prior CVA • Heart failure	1 366	Single-lead ECG patch (Zio XT) for up to 14 days	4 months	3.9	Not specified per new AF case

		• Diagnosis of both diabetes and hypertension • Mitral valve disease • Left ventricular hypertrophy • COPD requiring home oxygen • Sleep apnoea • History of pulmonary embolism • History of myocardial infarction • Diagnosis of obesity					
SCREEN-AF (Canada and Germany)	2021	Age ≥75 years with hypertension	434	Single-lead ECG patch (Zio XT) for up to 14 days with automated home blood pressure machines with oscillometric AF screening capability to use twice-daily during the ECG monitoring periods.	6 months	5.3	78

Abbreviations: AF, atrial fibrillation; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 years [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74 years, Sex Category; COPD, chronic obstructive pulmonary disease; CVA, cerebrovascular accident; ECG, electrocardiogram; NT-proBNP, N-terminal pro-B-terminal natriuretic peptide; OAC, oral anticoagulant; UK, United Kingdom; USA, United States of America

1.3.8.2 Effect of screening for atrial fibrillation on health outcomes compared to routine care

Health outcomes relevant to AF screening were considered by the USA Preventative Services Taskforce to be all-cause mortality, stroke, stroke-related morbidity and mortality, and quality of life.[89] Two large RCTs have investigated the effect of AF screening on hard clinical endpoints, The LOOP and STROKESTOP studies.

The STROKESTOP study randomized adults aged 75 or 76 years living in 2 regions of Sweden to an invitation to screening (n = 14 387) or to a control group that did not receive an invitation to screening (n = 14 381).[67, 70, 90] At baseline, 12.1% of the intervention group and 12.8%of the control group had known AF.[70] Of those invited to screening, 51.3% participated in the screening intervention, which was 2 weeks of twice-daily intermittent single-lead ECG monitoring with a handheld device for 30 seconds. AF was diagnosed on screening in the presence of at least one 30-second recording with irregular rhythm without p waves or a minimum of 2 similar episodes lasting 10 to 29 seconds during 2 weeks of intermittent recording. The intervention was not masked, and outcome ascertainment was through national health registry data. The primary outcome was originally specified as ischemic stroke but was changed by study investigators in 2017 before any data analysis to a composite endpoint that included ischemic stroke, haemorrhagic stroke, systemic embolism, bleeding leading to hospitalisation, and all-cause mortality. At a median follow-up of 6.9 years, the rate of the composite endpoint events was significantly lower in the invitation-to-screening group (5.45 events/100 person-years) compared with the control group (5.68 events/100 person-years) with an unadjusted HR of 0.96 (95%CI 0.92-1.00, p = 0.045).[70] No significant differences were observed between the invitation-to-screening group and the control group for any of the individual outcomes contributing to the composite end point. No findings were reported for the subgroup of participants without known AF at baseline.

The LOOP study was a RCT conducted in four centres in Denmark. Individuals without known AF aged between 70 and 90 years, with at least one additional stroke risk factor, were randomly assigned 1:3 to ILR monitoring or routine care.[72] Of 6 004 individuals randomly assigned, 1 501 had ILR monitoring and 4 503 made up the control group, with a median follow-up of 5.4 years. An episode of 6 minutes of AF on continuous monitoring was sufficient for diagnosis and consideration of anticoagulation. The LOOP study found no significant reduction in the risk of stroke or systemic

embolism between the ILR group (67 of 1 501 [4.5%]) compared with the control group (251 of 4 503 [5.6%]; HR 0.80, 95% CI 0.61 – 1.05; p = 0.11).[72]

Why do the results of the two RCTs differ? In the first instance, the sample size and the number of events was much smaller in The LOOP study, reflected in the wide confidence intervals for effect size, which may have left the study underpowered. Furthermore in The LOOP study, an episode of 6 minutes of AF on continuous monitoring was sufficient for diagnosis and consideration of anticoagulation. In the ASSERT study individuals with a duration of subclinical AF (SCAF) greater than 24 hours were found to be at increased risk of stroke compared to those without AF but those with SCAF under 24 hours in duration were not found to be at increased risk.[45, 91] It is possible that the AF episodes diagnosed in STROKESTOP were more likely to be of longer duration and hence confer elevated stroke risk and thus had a greater benefit from oral anticoagulation. The threshold of SCAF duration detected on continuous monitoring that would benefit from oral anticoagulation is under evaluation in the ongoing ARTESiA) double-blind RCT that includes participants with stroke risk factors and an episode of SCAF of at least 6 minutes duration. Enrolled patients are randomised 1:1 to aspirin or apixaban, with a composite primary outcome of stroke and systemic embolism and a safety outcome of clinically overt major bleeding.

A meta-analysis of four published RCTs (including the STROKESTOP and The LOOP studies) with a total of 35 836 participants following the intention-to-treat principle demonstrated a modest point estimate in favour of AF screening (RR 0.91, 95% CI 0.84-0.99) but published trials were heterogeneous in their populations, definition of stroke, and screening methodology.[92] The trial sequential analysis in the meta-analysis showed that the cumulative z-score from published data is insufficient to conclude the benefits of screening and calculated an optimal sample size of a total of 103 454 participants randomised.[92] Further trials exploring hard endpoints of AF screening in individuals without known AF are ongoing (Table 7).[71]

Table 7 Ongoing randomised clinical trials investigating the impact of atrial fibrillation screening on health outcomes

Study	**Year**	**Study design**	**Size**	**Follow-up period**	**Outcomes**
SAFER	2017	Age ≥70 years individuals from primary care randomised to receive screening through a single-lead handheld ECG recorder four times daily for three weeks	126 000	5 years	Ischaemic and haemorrhagic stroke
GUARD-AF	2019	Age ≥70 years individuals without known AF or OAC from primary care randomised to receive screening through a continuous ECG patch	52 000	2 years	Stroke leading to hospitalisation and bleeding leading to hospitalisation
HEARTLINE	2020	Age ≥65 years individuals randomised to screening through a smartwatch device and a healthy heart program	150 000	3 years	Composite of cerebrovascular events and all-cause death
STROKESTOP II	2017	Age 75-76 years Stockholm region inhabitants, randomised to receive screening procedure or usual care. Participants randomised to screening were assigned according to NT-proBNP levels to either one-stop screening or intermittent screening four times daily for two weeks with a single-lead handheld ECG recorder	6 868	5 years	Primary outcome: stroke or systemic embolism Secondary outcome: bleeding, stroke, systemic embolism, or all-cause death

Abbreviations: ECG, electrocardiogram; OAC, oral anticoagulant; NT-proBNP, N-terminal pro-B-terminal natriuretic peptide

1.3.8.3 Potential harms for participants of atrial fibrillation screening

In neither of the STROKESTOP or The LOOP studies was screening for AF and treatment of newly-diagnosed AF associated with a statistically higher rate of major bleeding (interventions vs control arm: STROKESTOP hospitalisation for major bleeding, HR 0.98, 95% CI 0.91-1.06, p=0.65; The LOOP study major bleeding: HR 1.26, 95% CI 0.95-1.69, p=0.11).[70, 72]

Population-level screening could lead to significant numbers of false-positive results. These patients might be exposed to unnecessary additional investigations and health anxiety.[45] There are limited data on the psychological effects of AF screening. The SAFE study collected data on anxiety levels and quality of life before and after screening. The screening seemed tolerable to most participants, but anxiety levels were higher amongst those screened positive for AF.[86] A Semi-structured longitudinal interview study of participant engagement in the ongoing SAFER study found that participants were supportive of screening for AF, explaining their participation in screening as a 'good thing to do'.[93] Participants suggested screening could facilitate earlier diagnosis, more effective treatment, and a better future outcome, despite most being unfamiliar with AF. Participating in AF screening helped attenuate participants' concerns about stroke and demonstrated their commitment to self-care and being a 'good patient'. Participants considered engaging in AF screening as low risk, with few perceived harms, if the screening device was non-invasive and they considered themselves unlikely to have the condition.[93]

1.3.9 Cost-effectiveness of atrial fibrillation screening

Economic assessment of AF screening depends on a range of factors:[57, 71]

- Rate of undiagnosed AF in the target population
- Difference in AF detection between the screening intervention and routine practice without screening
- Stroke and mortality risk of the target population
- Expected reduction in stroke and mortality and increase in bleeding risk from OAC
- Cost of the screening methodology
- Country-specific "willingness-to-pay" thresholds to avoid one stroke.

An economic analysis of the SAFE study showed, using probabilistic sensitivity analyses, a 60% likelihood that opportunistic screening was cost-effective in both men and women.[86, 94] Reviews of systematic and opportunistic screening for AF detection indicate that both were more cost-effective than routine practice for those ≥65 years of age, although this outcome depends on method chosen, frequency of screening, and age.[95, 96] The first health-economic study using actual long-term clinical follow-up data from the STROKESTOP study, extrapolated to a Markov model with a life-time perspective, showed that systematic screening for AF was associated with both lower costs and gained quality-adjusted life years (QALYs).[97] The screening strategy was thus dominant compared with non-screening and cost-saving after 3 years based on Swedish cost structure and cost levels. This was mainly explained by a low cost for screening and OAC treatment, in addition to fewer cases of stroke in the screening invitation group.

1.3.10 Recommendations in clinical guidelines

The 2020 ESC guidelines recommend opportunistic screening in individuals aged 65 years and older and suggests taking into consideration systematic screening in individuals aged 75 years and older, or those with stroke risk factors.[1] The NHS Long Team Plan aims for early detection and treatment of AF,[98] but the current UK National Screening Committee policy, based on an external review against programme appraisal criteria in 2019, is that population screening for AF should not be offered by the NHS.[99] Similarly the USA Preventative Services Task Force updated evidence report and systematic review in 2022, including the STROKESTOP study, concluded that although screening can detect more cases of unknown AF, evidence regarding effects of AF screening in primary care populations on health outcomes is limited.[89]

1.3.11 Summary

- Opportunistic AF screening does not appear to increase AF detection in primary care compared with routine care in individuals aged 65 years and older in contemporary care.
- Systematic AF screening is feasible, increases detection of AF compared to routine care, and leads to increased prescription of oral anticoagulation amongst newly diagnosed AF cases.
- Systematic AF screening may reduce the risk of stroke, though more evidence is required, and is not associated with an increased risk of bleeding amongst participants.

- Detection rates for new AF is relatively low when systematic AF screening with non-invasive devices is guided by age or stroke risk.

1.3.12 New approaches to consider for improve the clinical- and cost-effectiveness of atrial fibrillation screening

Detection rates for new AF in RCTs using non-invasive devices have been relatively low (3.0-5.3%), which limits both the clinical- and cost-effectiveness of AF screening.[67, 68, 87, 88] The results of The LOOP study suggest that simply monitoring for longer to achieve a higher detection rate of new AF may not result in an improvement in stroke outcomes.[72] Furthermore, AF is only one of many important risk factors for stroke and rates of ischaemic stroke have been decreasing over recent years.[100] This means that the relative risk reduction for ischaemic stroke from AF screening could be small. Overall for population systematic AF screening to be effective may require i) a high risk population for incident AF to be identified to achieve a sufficient yield through non-invasive AF detection modalities, and ii) to consider whether individuals identified for AF screening could benefit beyond stroke prevention.

1.4 Clinical risk stratification to identify individuals for screening for atrial fibrillation

The eligible population for RCTs of AF screening has often been defined by age. As incidence increases disproportionally in older adults, age is one of the best predictors of AF.[11, 101] However, yields of AF from non-invasive devices, which are the most acceptable AF detection approach to the general public, in an age-based approach are low. Furthermore in the Belgian Heart Rhythm Week, a untargeted voluntary screening programme available to all adults in Belgium organised by the Belgian Heart Rhythm Association one week a year from 2010 to 2014, half of all new AF cases were younger than 65 years of age.[102]

Therefore, for practical reasons, clinical risk stratification tools that better characterise the target population, decrease sample size, and identify subpopulations at risk are needed.[44] ECGs in sinus rhythm can be analysed with deep neural network-developed models to identify individuals at high risk of incident AF,[103, 104] but their practical applicability is limited to only special populations who have undergone an ECG or holter monitoring for other reasons. The use of these models for selecting participants of large population screening programmes seems to be unpractical.[44]

1.4.1 Multivariable prediction models

A prediction model (also sometimes referred to as an algorithm, or risk score, or decision support tool) is a formal combination of multiple predictors from which risks of a specific endpoint can be calculated for individual patients.[72] For an individual with a given state of health (startpoint), a prediction model converts the combination of predictor values to estimate the risk of experiencing a specific endpoint within a specific period, ideally an absolute risk.[105] Prediction models are abundant in the medical literature,[106] but few of the models are implemented or used in clinical practice,[107] and few models are evaluated for their impact on health outcomes.[105]

1.4.1.1 Methods for the development of multivariable prediction models

1.4.1.1.1 Traditional regression

Clinical prediction modelling has historically used regression techniques, which make the assumption that each predictor is related to the outcome in a linear way.[108] The inclusion of variables can be decided by clinical knowledge *a priori*, statistical techniques or through a combination of the two. Predictor selection using statistical techniques is least prone to bias when one starts with a full model that includes all potential variables and then applies a backwards selection approach.[108] Logistic and Cox regression modelling are most often used for short term and long term dichotomous outcomes (whether an individual experiences an event or not), respectively.[105]

1.4.1.1.2 Machine learning

Machine learning (ML) is the scientific discipline of how computers learn from data and, recently, has been facilitated by advances in computing speed and capacity.[109] ML algorithms can be defined as any approach that performs an automated search, either stochastic or deterministic, for the optimal model. In healthcare applications it can be further divided into supervised and unsupervised learning. Prediction of a future clinical diagnosis is a supervised learning (“classification”) task, where large amounts of data are typically annotated (“labelled”) by humans (e.g. presence or absence of diagnosis of AF) and the models then learn from the data which features are important for prediction.[110] In unsupervised learning patterns within and between data are sought by algorithms without any input from the investigator.[110] ML techniques applicable to prediction modelling include regularised logistic regression, support vector machines,

random forests, naïve Bayes and neural networks (Table 8).[109, 111-113] They do not require pre-specification of a model structure but instead search for the optimal fit within certain constraints (specific to the individual algorithm).[114]

Table 8 Descriptions of machine learning techniques applicable to prediction modelling

Machine Learning Technique	**Description**
Regularised logistic regression (LR)	Uses generalised linear models with Least Absolute Shrinkage and Selection Operator (LASSO) regularisation which both reduces the number of features in the model and attenuates the magnitude of their coefficients
Support vector machines (SVM)	Locates a decision boundary (called the hyperplane) based on a subset of data points (support vectors) that maximises the perpendicular distance between the decision boundary and the closest of the data points. The data is transformed with the kernel trick so that classes become linearly separable
Random forests (RF)	Decision trees seek to use variables to discriminate between the two outcomes. At each node in each tree one feature is selected that most effectively achieves this split. Each tree only has access to a subset of training examples and only a subset of features are considered. RF use many decision trees to construct a more robust ensemble output
Naïve Bayes (NB)	Uses Bayes' theorem to predict the probability of an outcome by assuming independence between features
Neural networks (NN)	A number of input neurons representing information taken from each of the features in the dataset, feed through a small number of hidden layers before passing to an output layer where the final decision is presented. As information passes through the neurons it is multiplied by a weight and a non-linear transformation is applied. Weights are adjusted during training based on the discrepancy between output and desired output.

1.4.1.1.3 Comparison of regression and machine learning techniques

Key advantages of machine learning techniques compared with regression techniques can be their handling of non-linearities and heterogeneity of effects, whilst disadvantages include overfitting, interpretability and presentation.[114]

The most basic assumption of regression models is that the relationship between a risk factor and outcome is linear, that is, the effect increases uniformly throughout the range of the predictor. While this may be plausible, for some risk factors, there are many examples that have non-linear relationships. For example, one's change in risk of death moving from age 40 to 50 years is much lower than increasing age from 70 to 80 years.[114] Even though a regression model may approximate the true non-linear relationship well and provide a more parsimonious interpretation, ML methods will capture these non-linearities to a greater extent. Related to non-linearities is heterogeneity of effects. Heterogeneity of effects, sometimes also referred to as interactions, occurs when a variable's relationship with the outcome depends on the level of some other variable.[114] ML models automatically handle non-linearities as they search for the optimal fit, but in a regression model not properly accounting for these interaction effects may degrade the quality of a prediction model.

On the other hand, because ML models produce a more flexible relationship between predictors and the outcome, they have the potential to overfit to the data they are developed on, which may limit their generalisability.[108] Furthermore, making multiple, complex, non-linear transformations, sacrifices interpretability of how risk factors relate to the outcome of interest. Additionally, even though a ML method may show better performance, presentation of the results may be more complicated. For example, many prediction models have been converted into hand calculable scores, and this conversion is usually obtained by rounding regression coefficients into a points-based score for each predictor. However, such a conversion is not obtainable with many ML methods.[114]

1.4.2 Risk factors for atrial fibrillation

A range of factors have been demonstrated to be associated with incident AF,[20] suggesting that a multivariable prediction model for incident AF is feasible. I will discuss them below, divided into non-modifiable risk factors, modifiable risk factors (lifestyle factors) and comorbidities.

1.4.2.1 Non-modifiable risk factors

1.4.2.1.1 Age

Earlier work that I was involved in, and is outwith of this thesis, has demonstrated that the crude incidence of AF increases with age.[11] AF is seen in only a small percentage of individuals aged younger than 55 years (0.1%) but prevalence increases steeply to 9% in individuals older than 80 years.[115] Amongst individuals aged 60 years and older without AF, there is evidence of reductions in atrial voltage and increase in voltage heterogeneity.[116] Furthermore the prevalence of risk factors for AF, including hypertension, heart failure, and coronary artery disease increase with age.[117]

1.4.2.1.2 Sex

In the FHS men were 1.5-fold more likely to develop AF than women, after adjusting for other AF risk factors.[118] The Olmsted County Minnesota Study and the Rotterdam Study reported the age-adjusted AF incidence (per 1 000 person-years) in men to be 4.7 and 11.5, respectively, compared with 2.7 and 8.9 in women.[119, 120] Men often have less favourable risk factor profiles for AF development than women,[11] which may contribute to this sex-dependent difference.

1.4.2.1.3 Ethnicity

Caucasian ethnicity seems to predispose to AF. In the in the Multi-Ethnic Study of Atherosclerosis (MESA), among participants aged 65 years and older, the AF incidence was 46% to 65% lower in Hispanic, Asian, and Black individuals compared with non-Hispanic White individuals.[121, 122] Disentangling the reasons behind the observed differences in incidence rates by ethnicity can be challenging, as they may be contributed to by anatomic characteristics, or socioeconomic and environmental determinants of health. For example, Black individuals have smaller average LA dimensions compared to White individuals (and LA size is an independent predictor of new-onset AF)[123] but also have been shown to have lower access to healthcare resulting in under-diagnosis of AF.[124] Ethnic differences may also be explained by genetic parameters, as discussed below in 1.4.2.1.4.

1.4.2.1.4 Genetics

When genetics were implicated in determining the European ancestry in Black individuals, it was found that for every 10% increase in European ancestry there was a 10% increased risk of incident AF.[125] A genetic study involving three population-based

cohorts in the USA revealed that the single nucleotide polymorphism rs10824026 (chromosome 10: position 73661450) substantially mediated the higher risk for AF in White individuals compared with Black individuals.[122, 126]

Outside of ethnic parameters, a study on monozygotic twins estimated that the heritability of AF to be as high as 62%, indicating a strong genetic component to incidence of AF.[127] In the FHS the adjusted multivariate relative risk of developing AF in individuals where at least one parent had AF was 1.85 (95% CI 1.12-3.06), which increased to 3.17 (95% CI 1.71-5.86) when the sample was further limited to those without antecedent hypertension, heart failure and valvular heart disease.[128] A number of causative mutations for AF have been identified, specifically the ion channel *KCNQ1*, the cardiac peptide *NPPA*, the transcription factor *TBX5*, and a motor protein *MYL4*.[2, 129]

1.4.2.2 Modifiable risk factors

1.4.2.2.1 Physical activity

In the Women's Health Initiative Observational Study of 81 317 post-menopausal women, physical activity was independently associated with lower rates of AF after multivariate adjustment for demographic and clinical risk factors (>9 vs 0 metabolic equivalent task hours per week, HR 0.90, 95% CI 0.85-0.96).[130] By contrast in predominantly health young athletes or middle-aged men with few cardiovascular risk factors, strenuous physical activity is associated with an increased risk of AF.[131] Amongst athletes increased LA size and heightened parasympathetic tone appear to mediate induction of AF.[132, 133]

1.4.2.2.2 Smoking

An analysis of the Rotterdam Study, including 5 668 individuals aged 55 years and older without AF at baseline, demonstrated that both current and former smokers had an approximately 50% increased risk of incident AF compared to never smokers (RR 1.51, 95% CI 1.07-2.12; RR 1.49, 95% CI 1.14-1.97; respectively).[134] Cigarette smoking is associated with heightened C-reactive protein (CRP) levels and thus a pro-inflammatory state,[135] and atrial tissue from smokers has shown that nicotine had profibrotic properties, which correlated with the number of pack years smoking.[136]

1.4.2.2.3 Alcohol

An analysis of over 400 000 middle-aged predominantly White individuals from the UK Biobank showed a J-shaped association between total alcohol consumption and AF, with the lowest risk of AF with fewer than 7 drinks per week.[122, 137] Chronic heavy drinkers demonstrate evidence of alcohol-induced myocardial changes including cardiomegaly, ventricular dilatation, fibrosis, and lipid and inflammatory infiltrates, which can contribute to the AF substrate.[138]

1.4.2.2.4 Obesity

Population-based studies have shown that obesity and elevated body mass index (BMI) increase the risk of AF, independent of other risk factors for AF that are predisposed to by obesity such as hypertension and diabetes.[20] In a large meta-analysis of over 600 000 individuals from 51 studies every 5-unit increase in BMI was associated with a 29% greater excess risk of incident AF.[139] Obesity can induce changes of left ventricular hypertrophy (LVH), diastolic dysfunction, increased LA volume, fibrosis and fat content, reduced LA conduction velocity, and increased AF vulnerability.[37, 140-143] Moreover, obesity is associated with low-grade inflammation and larger epicardial fat, which impair atrial electrophysiology,[143, 144] and activation of the renin-angiotensin-aldosterone system (RAAS).[122]

1.4.2.3 Comorbidities

1.4.2.3.1 Hypertension

Hypertension is the most commonly occurring comorbidity among patients with AF, present in 55% of patients at the time of diagnosis from our analysis of routinely-collected primary care EHRs in England.[11] The relative risk of AF in hypertensive individuals is two-fold for both sexes after adjustment for age and other comorbidities.[145] Hypertension induces LA enlargement,[146] as well as activation of the RAAS,[147] which induces atrial fibrosis.[148] LVH – measured either by ECG or echocardiogram - is a common complication of hypertension and is an additional independent predictor for risk of AF.[118] In the LIFE trial regression of LVH correlated with a 33% reduction in AF incidence.[149] Notably, lower LVH is still associated with a significant reduction in the occurrence of AF even after adjusting for treatment with anti-hypertensives, presence of hypertension, and other risk factors.[150]

1.4.2.3.2 Heart failure

Heart failure was the strongest independent predictor of AF in both men and women in the FHS (men OR 4.5, 95% CI 3.1-6.6; women OR 5.9, 95% CI 4.2-8.4),[151] and it is estimated that between 30 and 40% of patients with heart failure will develop AF.[152] Both HFrEF and HFpEF are associated with an increased risk of AF.[153, 154] The co-existence of AF and HF is partly explained by shared risk factors including hypertension, DM, obesity, OSA, and CAD.[122, 155] In the context of HF experimental animal models have shown evidence of atrial enlargement and histological analysis of human specimens has shown interstitial, atrial and ventricular fibrosis.[156, 157] Furthermore HF also activates the RAAS,[1] and stretching of the atria in HF induces acute electrophysiological arrhythmogenic changes that could trigger AF in a vulnerable substrate.[158]

1.4.2.3.3 Coronary artery disease

The FHS found that antecedent CAD is a significant and independent predictor of AF.[118] AF is a common complication of acute MI, occurring in up to a fifth of patients.[159] There is evidence of LA dilatation, compromised left ventricular function, increased pulmonary capillary wedge pressure and right atrial pressure in patients who develop AF after MI.[159] Furthermore interventions associated with the treatment of MI, such as percutaneous coronary intervention (PCI) or coronary artery bypass grafting (CABG) can also precipitate AF.[160, 161]

1.4.2.3.4 Chronic obstructive pulmonary disease

In a self-controlled case series study using population-using hospital databases from five states in the USA, chronic obstructive pulmonary disease (COPD) patients were found to have a 28% increased AF risk,[162] which further increased with frequent exacerbations and an enlarged LA. COPD-related mechanisms contribute to AF onset. Chronic hypoxaemia modulates the expression of hypoxia-inducible factor 1 and increases systemic inflammation and oxidative stress promoting profibrotic remodelling of the atrial tissue.[163] Long-term advanced COPD is associated with right heart disease,[163] and in rat models of right heart disease a substrate became apparent for AF maintenance prominently involving right atrial fibrosis, conduction abnormalities, and right atrial re-entrant activity.[164]

1.4.2.3.5 Obstructive sleep apnoea

Patients with OSA and without other cardiovascular co-morbidities have an increased risk of AF (HR 1.5, 95% CI 1.17–2.01).[165] OSA is associated with substantial atrial

structural and electrical changes. Intermittent episodes of deoxygenation and reoxygenation induce oxidative stress that, along with chronic neurohormonal activation contribute to atrial fibrosis.[166] Frequent stretching of the atria from recurrent obstructive respiratory episodes also causes myocardial injury and remodeling, as well as local conduction slowing and re-entry.[166] Furthermore, the sympathovagal activation caused by obstructed breathing efforts induces acute electrophysiological arrhythmogenic changes that could trigger AF in a vulnerable substrate.[122, 166]

1.4.2.3.6 Diabetes mellitus

In the FHS, DM was associated with a 40% and 60% increased risk of AF in males and females, respectively.[118] A meta-analysis of cohort studies revealed that prediabetes and diabetes both increase the risk of AF, by 20% and 28% respectively.[167] Glucose intolerance and insulin resistance modulate electro-anatomical changes in the atria,[168] and oxidative stress, inflammation, and atrial fibrosis also contribute to the development of the AF substrate.[169, 170]

1.4.2.3.7 Congenital heart disease

Both right and left-sided congenital heart disease increase the risk of atrial tachyarrhythmias, including atrial septal defect, Ebstein's anomaly, tetraology of Fallot, bicuspid aortic valve, ventricular septal defect and patent ductus arteriosus.[171, 172] In patients with left-sided diseases increased LA pressure and volume loading can lead to micro-reentrant cicuits,[173] and right-sided lesions can cause increased atrial pressure, decreased refractory periods, and atrial dilatation in the right atria.[171]

1.4.2.3.8 Valvular heart disease

Valvular heart disease is a strong risk factor for AF globally. In less economically developed countries VHD as a delayed consequence of rheumatic fever (rheumatic heart disease) is a common cause of AF,[174] especially due to mitral stenosis,[175] which leads to LA dilatation, pressure overload and atrial fibrosis.[176] In the absence of rheumatic heart disease, a retrospective cohort study of 940 patients without AF found that, of VHDs, aortic stenosis (AS) has the greatest impact, even in the presence of mitral regurgitation (MR), in increasing the risk of AF significantly after adjustment for age, sex, other VHDs and echocardiographic abnormalities.[177] MR causes structural changes in the left atrium (LA) including dilatation, myofibril hypertrophy and fibrosis due to volume overload.[176] Left ventricular pressure overload in the context of AS causes compensatory concentric LV hypertrophy,[178] which leads to both LA pressure

overload and enlargement.[179] Amongst patients with AS, the magnitude of LA dilatation is much smaller than in patients with MR, suggesting that incident AF is precipitated by pressure overload with subsequent LA structural remodelling, including fibrosis.[177]

1.4.2.3.9 Chronic kidney disease

The adjusted risk of incident AF with chronic kidney disease (CKD) was observed to increase in a stepwise fashion in nearly 17 000 participants from three USA-based community-based cohorts; such that, compared to eGFR>90 as reference, eGFR ranges of 60–89, 45–59, 30–44, and <30 ml/min per 1.73 m^2, demonstrated hazard ratios of 1.09 (95% CI 0.97-1.24), 1.17 (1.00-1.38), 1.59 (1.28-1.98), and 2.03 (1.40-2.96), respectively.[180] Experimental animal models demonstrate that impaired kidney function is linked to myocardial fibrosis,[181] as well as alterations in calcium handling in the cardiomyocyte.[182] Subclinical volume overload may also lead to atrial stretch and contribute to induction of AF in patients with CKD.[180]

1.4.3 Risk prediction models for new-onset atrial fibrillation

Risk prediction models for incident AF prediction applicable in the community were summarised in a 2020 systematic review and meta-analysis by *Himmelreich et al*.[183] To be included studies had to:

- Be original studies in adults (≥18 years of age)
- Derive, validate and/or augment a tool for predicting risk of incident AF (or atrial flutter [AFl]) based on multivariable analysis
- Only include patients without a diagnosis of AF or AFl at baseline
- Incorporate into their risk prediction tool only variables that are applicable and/or commonly available in primary care settings, which the authors defined as medical history, physical examination, simple laboratory findings, or ECG parameters

From 27 included studies they found 21 multivariable prediction models (all regression models), ten of which had been derived for predicting incident AF, and the rest developed for another purpose but validated for incident AF.

The authors extracted data relevant to prediction model performance, discrimination and calibration. Discrimination quantifies the model's ability to distinguish between individuals developing or not developing the outcome, and can be summarised with the

c-statistic (c-statistic = 1 if the model discriminates perfectly, c-statistic = 0.5 if discrimination no better than chance) or area under the receiver operating characteristic (AUROC, same scale as c-statistic) and corresponding 95% CI. Calibration refers to the model's accuracy of predicted probabilities, and can be summarised using the p value of a goodness-of- fit test, the reported ratio for observed to expected events, or the calibration slope. Poor reporting of calibration meant that meta-analysis was limited to 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.[184] Models were only included in meta-analysis when *c*-statistic data was available for 3 or more cohorts. In each meta-analysis, the authors calculated the summary c-statistic, its 95% CI, and the I^2 statistic as an expression of the heterogeneity between studies.[185] When heterogeneity in meta-analysis of c-statistics was high (I^2 > 30%), they derived a 95% prediction interval (95% PI), which indicates a possible range for prediction model performance in a new validation.[186] When the 95% CI (or, in case of high heterogeneity, the 95% PI) of the summary c-statistic included 0.5, they 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.

Five models were eligible for meta-analysis. and their included variables are listed in Table 9. The Cohorts for Heart and Aging Research in Genomic Epidemiology AF (CHARGE-AF) and FHS score for Atrial Fibrillation (FHS-AF) models were developed for incident AF prediction, but the CHA_2DS_2-VASc and $CHADS_2$ scores were originally developed for prediction of stroke risk in patients with AF,[187, 188] and HATCH for prediction of progression from paroxysmal to persistent AF.[189]

Table 9 Variables included in prediction models for incident atrial fibrillation applicable in the community included in primary meta-analysis by *Himmelreich et al*[183]

	Demographics			**Comorbidities**							**Habit**	**Observations**					**O/E**	**ECG**
	Age	Sex	Race	HF	HTN	DM	CVA	CAD / MI	PAD	COPD	Smoking	Height	Weight	BMI	SBP	DBP	Murmur	PR interval
Models derived for incident AF																		
CHARGE-AF	X		X	X	X	X		X			X	X	X		X	X		
FHS-AF	X	X		X	X									X	X		X	X
Models originally derived for another purpose but tested for prediction of incident atrial fibrillation																		
$CHADS_2$	X			X	X	X	X											
CHA_2DS_2-VASc	X	X		X	X	X	X	X	X									
HATCH	X			X	X		X			X								

Abbreviations: BMI, body mass index; $CHADS_2$, Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, prior Stroke or transient ischemic attack [2 points]; CHA_2DS_2-VASc, Congestive HF, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex category; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; CHS, Cardiovascular Health Study; CAD, coronary artery disease; COPD, chronic obstructive pulmonary disease; CVA, cerebrovascular accident; DBP, diastolic blood pressure; DM, diabetes mellitus; ECG, electrocardiogram; FHS-AF, Framingham Heart Study score for Atrial Fibrillation; HATCH, Hypertension, Age,

stroke or Transient ischemic attack, Chronic obstructive pulmonary disease, Heart failure; HF, heart failure; HTN, hypertension/anti-hypertensive medications; MI, myocardial infarction; O/E, on examination; PAD, peripheral arterial disease; SBP, systolic blood pressure

There were three models that resulted in a summary c-statistic with significant 95% PI (that is, did not include 0.5) in the 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 CHA_2DS_2-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). The $CHADS_2$ score (summary c-statistic 0.66; 95% CI 0.59–0.74; I^2 100%; 95% PI 0.447– 0.883; n= 4 studies; n = 1 996 338 patients) and HATCH score (summary c-statistic 0.67; 95% CI 0.61–0.73; I^2 99%; 95% PI 0.486–0.844; n= 4 studies; n = 1 604 822 patients) had 95% PIs that included 0.5.

A secondary analysis was conducted for each risk model that had 3 or more 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 consideration when wanting to translate summary risk model performance to clinical settings.[190] For this analysis only the CHARGE-AF risk score and the FHS-AF score were eligible, each at a 5-year and 10-year prediction window, where only the CHARGE-AF score 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).

The authors found two studies where multivariable prediction models were compared with age alone as the predictor and both found that the multivariable models had significantly higher discrimination for incident AF.[191, 192] They concluded that the use of multivariable risk modes in selecting patients for community AF screening is likely to be more efficient than selecting based on age alone. Of the ample AF risk models found they concluded that the CHARGE-AF model was the most suitable prediction model for incident AF, and likely has merits as a low cost triage test for future primary AF screening efforts.

I consider there to be a number of potential shortcomings with this analysis and the conclusions drawn. First, with regards to the inclusion criteria, models that required ECG parameters were included. However ECG parameters are seldom available in the community, so their inclusion as variables would significantly hamper the ability of a model to be implemented at scale. Second, authors assessed risk of bias using the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) domains,[190] but this has been superseded by the Prediction model Risk Of Bias ASsessment Tool (PROBAST).[193] Compared to PROBAST, the CHARMS

domains have a more lenient criteria with regards to the handling of missing data, model development rigour and reporting of model performance. Accordingly, a lower proportion of studies (24.7%) were considered high risk of bias than would be considered by best practice standards, which gives an unrealistic assessment of the quality of available models. Third, authors used a frequentist method for meta-analysis, but where there are fewer studies or a mixture of study sizes these methods can produce PIs with poor coverage, leading to misleading estimates for the range of performance that could be expected from the model in a new population.[186] Fourth, no models developed with machine learning were included in the analysis, but machine learning multivariable models for AF risk had been published by the time of this review.[194]

With regards to the conclusions drawn, it does appear that integrating risk factors alongside age will more accurately discriminate people likely to develop AF compared with people who will not. However I have doubts that the available evidence suggests CHARGE-AF, or any of the other included models, is suitable to use to triage community AF screening. First, CHARGE-AF was developed and tested for 5- or 10-year prediction horizons but these are not suited to targeting AF screening, as AF that develops more than a few months in the future will not be picked up during screening. Second, prediction performance was not that strong. Summary c-statistics of <0.60, 0.60-0.70, 0.70-0.80, and >0.80 can be considered as inadequate, adequate, acceptable and excellent based on prior publications.[195] The summary predictive performance in meta-analysis for each of the included models was only adequate-to-acceptable (0.67-0.71). Third, CHARGE-AF requires a complete dataset of height, weight, systolic and diastolic blood pressure to calculate AF risk, but these are often missing in routinely-collected records in the community,[196] so the application of this risk score may require additional appointments to obtain this information. This seems impractical when attempting to implement a population screening strategy.

Using the CHA_2DS_2-VASc score may have advantages. Its summary discriminative performance was not that different from CHARGE-AF, and would simultaneously provide an assessment of stroke risk as an indicator of eligibility for anticoagulation. However three of the RCTs of community systematic AF screening (mSTOPS, SCREEN-AF, and REHEARSE-AF)[68, 87, 88] were already guided by elevated stroke risk, and yields of newly detected AF were modest (Table 8).

A 2022 position paper for EHRA on searching for AF, included a section pertaining to assessment of populations at risk of AF.[44] The paper agreed that though several

predictive models for risk stratification of AF have been proposed, most are limited in defining high-risk populations suitable for large-scale screening projects. The main shortcomings noted were that many have been developed to predict the long-term risk (5- or 10-years) and some require biomarkers or imaging data not readily available in the general population. The writing group emphasised the requirement for an improved prediction model – that is, one that is easy to use, applicable to the general population, and based on readily available information such as comorbidities). The writing group specifically mentioned the C_2HEST score as a contrast to other prediction models for AF, in that it only included age and comorbidities.

The C_2HEST score was developed as a simple clinical score for incident AF in an Asian population without structural heart disease (Table 10).[197, 198] The performance of the C_2HEST score in its derivation sample of a regional Chinese insurance database was acceptable (AUROC 0.75, 95% CI 0.73-0.77).[199] However on external validation in a Korean population (AUROC 0.65, 95% CI 0.65-0.66) and older Danish citizens (AUROC: age 65, 0.588, 95% CI 0.585-0.591; age 70, 0.594, 95% CI 0.591-0.597; age 75, 0.593, 95% CI 0.590-0.596) it demonstrated much poorer performance.[200]

Table 10 Assessment of risk of incident atrial fibrillation using C_2HEST

Criteria	**Value**	**Points**
Coronary artery disease	Yes/No	1
COPD	Yes/No	1
Hypertension	Yes/No	1
Elderly	Age>65 years	2
Systolic HF	Yes/No	2
Thyroid disease (hyperthyroidism)	Yes/No	1

Abbreviations: COPD, chronic obstructive pulmonary disease; HF, heart failure

1.4.4 Using primary care electronic health records to identify the population for atrial fibrillation screening

There are over 300 million consultations annually in primary care in the UK.[201] Due to the wide-scale uptake of EHRs in primary care 98% of the UK population are registered with a primary care EHR,[202] and across EU countries 96% of all GPs use an EHR.[203] Primary care EHRs have the advantage of holding comprehensive longitudinal clinical data,[202] meaning they contain healthcare information relating to an individual that can span many years. Accordingly they can be used to characterise an individual's health, and use this information to stratify their risk of AF in the future. Thus, primary care EHRs appears a potential medium through which individuals eligible for AF screening could be identified at scale. In fact, risk prediction models are unlikely to be widely used for targeting AF screening unless they can be incorporated into EHR systems.[204]

Prediction modelling research has historically been conducted in prospective, pre-designed longitudinal cohorts with standardised examinations, investigations and follow-up visits.[183] However the generalisability of these models is called into question by selectivity of the sample population and inclusion of variables that are not routinely available in clinical practice.[183] Furthermore, the performance of prediction models in prospective cohorts may not translate to an EHR setting.[196, 204] For example, the CHARGE-AF model could only be applied to 17.2% of adults aged ≥40 years without known AF in the Netherlands Institute for Health Services Research Primary Care Database (NIVEL-PCD),[196] and when it was applied in a sample of 33 494 patients aged ≥40 years without known AF in the Vanderbilt University Medical Centre outpatient clinic EHRs, its prediction performance was only acceptable (c-statistic 0.708, 95% CI 0.699-0.718),[204] and weaker than its original derivation in a prospective cohort (0.765, 95% CI 0.748-0.781).[205] A systematic review of the performance of models for prediction of incident in primary care EHRs is absent from the literature.

Routinely-collected EHR datasets capture huge sample sizes over many years and can link data from many sources providing an overarching narrative of the patient's health status (Table 11). Thus EHR could provide a fruitful resource for prediction modelling, especially the use of ML techniques.[206] In this project I will consider only structured EHR data as the technology for natural language processing to extract free text is too immature for widespread clinical implementation.[207]

Table 11 Types of data stored in community-based electronic health records

Type of Data	Purpose of data
Structured	
Patient demographic information	Background information including age, ethnicity
Diagnosis codes	Data for a medical diagnosis linked to a medical ontology e.g. International Statistical Classification of Diseases and related Health Problems (ICD-10) or Read codes
Drugs codes	Contains codes for each drug and form
Treatment procedures	Contains types of procedures as linked to an ontology e.g. Current Procedural Terminology codes
Lab tests	Recording all laboratory measurements with linked tables to the type of test and value attached
Observations	Continuous values such as height, weight and BMI.
Referrals	Data reporting referrals from general practice to secondary care
Unstructured	
Clinical notes	Free text inputted by health professionals that could be used to describe any facet of a patient's condition

Abbreviations: BMI, body mass index

In the UK there are four key primary care datasets for research: ResearchOne, Clinical Practice Research Datalink (CPRD), The Health Improvement Network (THIN) and QResearch. Each contains millions of patient records, and can extend back as far as 1987.[202] They have been used to develop prediction models that have then become part of routine practice.[194] For example the Electronic Frailty Index (eFI) was developed in ResearchOne and validated in the THIN database.[208] The eFI uses routinely available primary care EHR data (36 disease states, symptoms/signs, and disabilities) to categorise individuals as being fit, or having mild frailty, moderate frailty or severe frailty. Use of the eFI is supported in NICE guidance,[209] and the model has been integrated into EHR systems SystmOne, EMISWeb ad VISION EHR, reaching 99% of all GPs across the UK, and available for use at no extra cost.[210] The development of eFI also used coding systems that are widely available in other countries, meaning that it is not limited to the UK. This represents a proof of concept for how a prediction model for incident AF may be able to be implemented at scale to target AF screening in the UK and further afield.

1.4.5 Summary

- Multiple risk factors have good evidence for association with new-onset AF.
- Multivariable risk prediction models for incident AF better discriminate individuals likely to develop AF, compared with age alone.
- The utility of currently available prediction models may be limited by moderate discrimination performance, long prediction horizons, and inclusion of variables that may be missing in routinely-collected primary care EHRs.
- The majority of models summarised to date have been developed with traditional regression techniques.
- Primary care EHRs provide an attractive medium to implement a prediction model for incident AF in the general population, but the performance of models in this setting is unknown.

1.5 Association between atrial fibrillation and non-atrial fibrillation outcomes

To date, screening for AF has been targeted at reducing the risk of ischaemic stroke in the screened population by commencing OAC treatment in detected cases. As discussed in Section 1.3.7.2, it is possible a large numbers of patients need to be studied to demonstrate the efficacy of AF screening for stroke prevention. Moreover, AF is only one of the risk factors of stroke,[92] and the rate of ischaemic stroke is decreasing.[100].

Accordingly, it may be prudent to consider what may be the benefits of AF screening beyond prevention of ischaemic stroke. Further possible benefits from AF screening would be lower mortality and a possibility to address undetected structural heart disease and untreated cardiovascular risk factors such as hypertension, obesity, alcohol consumption and OSA.[71]

AF frequently develops as a result of, and in parallel with, other diseases. The GLORIA-AF international cohort of 21 241 participants found that 71.2% had at least two concomitant, chronic, comorbid conditions, with nine in ten having hypertension, one in three HF, one in three CKD, and one in four CAD.[211] A systematic review comprising 9 686 513 participants with and without AF (587 867 [6.1%] with AF) found that AF was associated with an increased risk of a range of outcomes: all-cause

mortality (RR 1.46, 95% CI 1.39-1.54) cardiovascular mortality (RR 2.03, 95% CI 1.79-2.30) major cardiovascular events (RR 1.96, 95% CI 1.53-2.51), stroke (RR 2.42, 95% CI 2.17-2.71), ischaemic stroke (RR 2.33, 95% CI 1.84-2.94), ischaemic heart disease (RR 1.61, 1.38-1.87), HF (RR 4.99, 95% CI 3.04-8.22), CKD (RR 1.64, 95% CI 1.41-1.91) and peripheral arterial disease (RR 1.31, 95% CI 1.19-1.45).[32] The highest absolute risk increase with AF was for incident HF (11.1 events/1000 participant years, 95% CI 5.7-20.0). Furthermore, in a retrospective cohort study of Medicare beneficiaries, investigators showed that HF was the most common non-fatal cardiovascular event among adults with AF, and hospitalisation for HF was almost twice as common as hospitalisation for stroke.[212]

The mechanism by which AF is associated with an increased occurrence of a range of different cardiovascular diseases seems to extend beyond the arrhythmia. In Section 1.2 and 1.1.4 there appears to be a bidirectional relationship between AF and other comorbidities, that is, they are both a risk factor and complication of AF. Age, smoking, obesity, inflammatory diseases and hypertension are shared risk factors between AF, vascular disease, aortic stenosis, HF, DM and CKD.[213-215] It seems likely that AF could be acting as a marker for shared underlying risk factors and cardiovascular disease.[32]

The minority of patients with AF die as a result of stroke,[216] and so it is argued that interventions aimed at reducing outcomes beyond stroke are warranted in patients with AF. As discussed in Section 1.1.5.4, management of comorbidities and risk factors is now a central pillar of management for patients with AF.[1] That is, a reduction of the burden of non-stroke events in individuals with AF may be actionable through a focus on the management of cardiovascular risk factors, optimisation of established cardiovascular disease, and the identification of undetected cardio-renal-metabolic disease. It may be that individuals deemed eligible for risk-based AF screening also have risk factors, and both detected and undetected comorbidities that could be optimised to reduce the risk of events beyond stroke, irrespective of whether AF is detected during screening.

1.5.1 Summary

- AF is associated with the development of a range of diseases and death, which appears to extend beyond the effect of the arrhythmia.
- Individuals with AF have a high comorbidity burden at point of diagnosis and shared risk factors and pathological pathways with other diseases.

- Outcomes for patients with AF can be improved by management of comorbidities and risk factors.
- It is possible that individuals identified by a multivariable prediction model as at higher risk of AF may also be at elevated risk of other outcomes.

Part II

Chapter 2 Prediction of incident atrial fibrillation in community-based electronic health records: a systematic review with meta-analysis

Ramesh Nadarajah, Eman Alsaeed, Ben Hurdus, Suleman Aktaa, David Hogg, Matthew G D Bates, Campbell Cowan, Jianhua Wu, Chris P Gale

2.1 Summary of the publication

- This paper presents the systematic review and meta-analysis used to inform the knowledge gap to be addressed by a prediction model for incident AF developed in primary care EHR data.
- The study found that prediction models for incident AF have been developed and/or validated in community-based EHRs but have a series of shortfalls:
    - Prediction models eligible for meta-analysis show only moderate discrimination performance.
    - Many prediction models have prediction horizons that are often 5- or 10-years, making it difficult to judge the merits of investigating individuals in the short-term.
    - Many studies do not report calibration performance and inadequately handle missing data, which places their performance at risk of bias.
    - Prediction models derived using machine learning can show improved prediction performance in the development dataset compared to models developed using traditional regression techniques.

2.2 Publication status

- Published 1 July 2022
- Heart. 2022 Jul 1;108(13):1020-9.

2.3 Abstract

2.3.1 Objective

Atrial fibrillation (AF) is common and associated with an increased risk of stroke. We aimed to systematically review and meta-analyse multivariable prediction models

derived and/or validated in electronic health records (EHR) and/or administrative claims databases for the prediction of incident AF in the community.

2.3.2 Methods

Ovid Medline and Ovid Embase were searched for records from inception to 23 March 2021. Measures of discrimination were extracted and pooled by Bayesian meta-analysis, with heterogeneity assessed through a 95% prediction interval (PI). Risk of bias was assessed using PROBAST (Prediction model Risk of Bias ASsessment Tool) and certainty in effect estimates by GRADE (The Grading of Recommendations, Assessment, Development and Evaluation).

2.3.3 Results

Eleven studies met inclusion criteria, describing nine prediction models, with four eligible for meta-analysis including 9,289,959 patients. $CHADS_2$ (summary c-statistic 0.674; 95% CI 0.610 – 0.732; 95% PI 0.526 – 0.815), CHA_2DS_2-VASc (summary c-statistic 0.679; 95% CI 0.620 – 0.736; 95% PI 0.531 – 0.811) and HATCH (summary c-statistic 0.669; 95% CI 0.600 – 0.732; 95% PI 0.513 – 0.803), resulted in a c-statistic with a statistically significant 95% PI and moderate discriminative performance. No model met eligibility for inclusion in meta-analysis if studies at high risk of bias were excluded and certainty of effect estimates was ‘low’. Models derived by machine learning demonstrated strong discriminative performance, but lacked rigorous external validation.

2.3.4 Conclusions

Models externally validated for prediction of incident AF in community-based EHR demonstrate moderate predictive ability and high risk of bias. Novel methods may provide stronger discriminative performance.

2.3.5 Systematic Review Registration

PROSPERO CRD42021245093

2.4 Introduction

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia and is associated with a five-fold increased risk of stroke.[1, 13] This risk can be reduced by two-thirds by a number of effective oral anticoagulants,[35, 217] but it is estimated that 30% of patients living with AF are undiagnosed and its first manifestation is stroke in more than 10% of patients.[16, 218]

International guidelines recommend opportunistic rather than systematic screening in asymptomatic patients, using age over 65 years as the only risk predictor. [1, 219] In many European countries, a large proportion of the population are registered in primary care with a routinely-collected electronic health record (EHR).[196, 202] A multivariable prediction model that utilises this data source to give a more discriminative assessment of risk could allow far-reaching, cost-effective targeted screening.

There are several prediction models for incident AF in community-dwelling individuals but they have predominantly been tested in prospective cohorts and their performance may not translate to EHR data. To show utility for targeting screening in the general population using real-world EHR, a model would need to have been tested in EHR or administrative claims databases relevant to the general population or primary care (herein referred to as community-based EHR).[204]

We performed a systematic review and meta-analysis with a number of aims. First, to identify prediction models for incident AF derived or validated in community-based EHR. Second, to summarise the performance of individual prediction models to understand if any would be suitable for use in targeted screening. Third, to summarise the methods by which prediction models have been developed in EHR to inform future research within the field.

2.5 Methods

We reported this systematic review and meta-analysis in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines.[220]

2.5.1 Search strategy and inclusion criteria

The research question was framed using the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS; Table 1).[190]

Table 1 Formulation of research question using CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS)

CHARMS key items to guide framing of review, search strategy and study inclusion and exclusion criteria	**Comments for this systematic review**
Prognostic versus diagnostic prediction model	Prognostic prediction model
Intended scope of the review	Models to inform referral for diagnostic testing
Types of prediction modelling studies	Prediction model development without external validation in independent data, prediction model development with external validation in independent data, external model validation, possibly with model updating
Target population to whom the prediction model applies	Adults in the general population who have a primary care or community electronic health record
Outcome to be predicted	Specific future event, diagnosis of atrial fibrillation
Time span of prediction	Any time interval
Intended moment of using the model	Models to be used in adults in primary care using electronic health records to predict risk of development of atrial fibrillation in the future, and inform targeted screening

We searched the Medline and Embase databases through the Ovid platform from inception through 23 March 2021. We used a combination of keywords and subject headings related to AF, prediction models and EHR based on previous literature.[183, 221, 222] The search was limited to the English language and to human studies. The full search strategy is provided in Table 2. We manually searched the reference lists of

included studies and previous systematic reviews.[183, 221] Duplicates were removed using Endnote's duplicate identification strategy and then manually.

Table 2 Search terms and search strategy with full results

Database(s): **Ovid MEDLINE(R) ALL** 1946 to March 23, 2021

#	**Searches**	**Results**
1	atrial fibrillation/ or atrial flutter/	61138
2	(atrial fibrillation or atrial flutter).ti,ab.	77059
3	1 or 2	90196
4	ROC Curve/ or (stratification or discrimination or discriminate or c-statistic or c statistic or Area under the curve or Calibration or Indices or Algorithm or Multivariable).ti,ab.	852681
5	Mass screening/ or Screen*.ti,ab.	823683
6	Prevalence/ or prevalenc*.ti,ab. or incidence/ or incidenc*.ti,ab.	1568534
7	population/ or population*.ti,ab.	1832889
8	5 or 6 or 7	3645965
9	(communit* or data*).ti,ab.	4711739
10	(general adj3 population).ti,ab.	122216
11	database/ or dataset/	1216
12	(Electronic Health Record* or electronic medical record* or electronic personal record* or electronic patient record* or personal health record* or personal medical record* or computer health record* or computer medical record* or computer patient record* or ehr? or phr? or ephr? or emr? or paehr?).ti,ab.	47628
13	Electronic Health Records/ or exp medical records systems computerized/ or exp health records personal/	43724
14	Primary Health Care/ or (primary care or general practic*).ti,ab.	184646
15	9 or 10 or 11 or 12 or 13 or 14	4935714
16	3 and 4 and 8 and 15	1342
17	limit 16 to (english language and humans)	1072

Database(s): **Embase Classic+Embase** 1947 to 2021 March 23

#	Searches	Results
1	exp heart atrium fibrillation/ or exp atrial fibrillation/ or exp heart atrium flutter/	93727
2	(atrial fibrillation or atrial flutter).ti,ab,kw.	143413
3	1 or 2	166612
4	predict.ti.	78253
5	(validat* or rule*).ti,ab.	1059357
6	(predict* and (outcome* or risk* or model*)).ti,ab.	1313986
7	((history or variable* or criteria or scor* or characteristic* or finding* or factor*) and (predict* or model* or decision* or identif* or prognos*)).ti,ab.	4704335
8	decision*.ti,ab. and statistical model/	7315
9	(decision* and (model* or clinical*)).ti,ab.	276542
10	(prognostic and (history or variable* or criteria or scor* or characteristic* or finding* or factor* or model*)).ti,ab.	370949
11	(stratification or discrimination or discriminate or c statistic or "area under the curve" or auc or calibration or indices or algorithm or multivariable).ti,ab.	1183028
12	receiver operating characteristic/	141783
13	4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12	6613249
14	exp mass screening/	262462
15	Screening.ab,ti,kw.	811498
16	exp prevalence/	814190
17	Prevalence.ab,ti,kw.	959326
18	exp incidence/	533091
19	Incidence.ab,ti,kw.	1207644
20	14 or 15 or 16 or 17 or 18 or 19	3136647

21	(communit* or data*).ti,ab.	6555772
22	(general adj3 population).ti,ab.	184176
23	database/ or dataset/	450631
24	Electronic Health Records/ or (electronic health record* or electronic medical record* or electronic personal record* or electronic patient record* or personal health record* or personal medical record* or computer health record* or computer medical record* or computer patient record*).ti,ab. or (ehr? or phr? or ephr? or emr? or paehr?).ti,ab.	94165
25	21 or 22 or 23 or 24	6816789
26	3 and 13 and 20 and 25	7804
27	letter.pt. or letter/	1179167
28	note.pt.	848283
29	conference abstract.pt.	4066914
30	editorial.pt.	690770
31	case report/ or case study/	2780774
32	(letter or comment*).ti.	217544
33	27 or 28 or 29 or 30 or 31 or 32	9080295
34	animal/ not human/	1523407
35	nonhuman/	6523962
36	exp animal experiment/	2699975
37	exp experimental animal/	749370
38	animal model/	1434020
39	exp rodent/	4130266
40	(rat or rats or mouse or mice).ti.	1709912
41	34 or 35 or 36 or 37 or 38 or 39 or 40	9318330
42	33 or 41	17366453
43	26 not 42	3796
44	limit 43 to english language	3636

To be eligible for inclusion a study had to:

- Be an original study in human adults (≥ 18 years of age).
- Develop and/or validate a prediction model(s) for incident AF or atrial flutter (AFl) based on multivariable analysis in a community-based EHR. We included AFl as a co-outcome because it has a similar indication for anticoagulation.[1]
- Be written in English.

Articles were excluded if they:

- Included patients with AF or AFl at baseline.
- Only reported measures of association between risk factors and incident AF rather than a full prediction model.
- Studied only a subset of the general population, for example individuals diagnosed with a particular morbidity.
- Incorporated variables that would not be routinely available in community-based EHR (e.g. ECG parameters).

In this review we were interested in models that could use structured 'coded' data in community-based electronic health records or administrative claims databases. To make screening in the community for AF more cost-effective and feasible the model would use variables that are available, calculate the risk automatically, and require minimal additional visits for baseline risk stratification. We only considered the use of structured 'coded' data as the technology for natural language processing to extract free text into 'coded' data is too immature for widespread clinical use. We used examples of primary care or population-based health information databases across the world to define the variables most likely to be coded or extractable, accepting that there will be some variation.

The information that was considered likely to be available in community-based data sources ± linkages (depending on whether the original purpose of the database was documentation of clinical care, epidemiological surveillance, or health system planning) were:

- Sociodemographic variables including but not limited to age, sex, ethnicity and indices of multiple deprivation.

- Disease conditions and procedures including but not limited to other cardiovascular diseases, diabetes mellitus, chronic lung disease, renal disease, inflammatory disease, cancer, hypothyroidism and surgical procedures.
- Clinical assessments including but not limited to heart rate, systolic and diastolic blood pressure, height, weight and body mass index.
- Medications prescribed including but not limited to antihypertensives, statins, antidepressants, anxiolytics/hypnotics and antipsychotics.
- Lifestyle factors including but not limited to smoking status and alcohol consumption.
- Simple laboratory tests and biomarkers including but not limited to total, high-density lipoprotein and low-density lipoprotein cholesterol, triglycerides, creatinine, c-reactive protein, erythrocyte sedimentation rate.
- Referrals

We excluded the following types of variables that are either not routinely available as structured codes, or are very rarely tested for in clinical practice and so are not generalizable:

- Analysis of electrocardiograph (ECG) parameters (e.g. PR interval, QRS duration, p-wave duration).
- Analysis of advanced diagnostic testing such as echocardiography parameters (e.g. LA dimensions, left ventricular end-diastolic diameter).
- Genetic markers and specialised (laboratory) tests (e.g. midregional sequence of pro-atrial natriuretic peptide).

We uploaded records to a systematic review web application (Rayyan, Qatar Computing Research Institute).[223] Four investigators (RN, EA, BH and SA) independently screened them for inclusion by title, abstract and full text and supplemental materials. Disagreements were resolved by consultation with a fifth investigator (JW).

2.5.2 Data extraction and quality assessment

Two investigators (RN and EA) independently extracted the data from the included studies based on CHARMS. This included the following domains: data source, participants, outcome(s), candidate predictors, sample size, missing data, and model

development, performance and evaluation. Discrepancies were resolved with a third investigator (JW). All data came from the primary reference, unless otherwise stated.

To allow quantitative synthesis of the predictive performance of the models we extracted measures of discrimination and calibration.[184] Discrimination quantifies the model's ability to distinguish between individuals developing or not developing the outcome. We extracted data on the c-statistic (c-statistic = 1 if the model discriminates perfectly, c-statistic = 0.5 if discrimination no better than chance) or area under the receiver operating characteristic (AUROC) and corresponding 95% confidence interval (95% CI). When the 95% CI was not reported we calculated it using methods described by *Debray et al*.[184] Calibration refers to the model's accuracy of predicted probabilities; we extracted data on the p-value of a goodness-of-fit test and the reported ratio for observed to expected (O:E) events or calibration slope.

Two investigators (RN and JW) assessed each model in each study for risk of bias and applicability to our review question using the Prediction model Risk Of Bias ASsessment Tool (PROBAST).[193] Discrepancies were resolved with a third investigator (CPG). Each model was assessed for risk of bias as either 'high', 'unclear' or 'low' in four domains (participants, predictors, outcomes and analysis) through a range of signalling questions. Applicability to our review question was assessed for each model in three domains (participants, predictors and outcomes) using the same scale.[193]

2.5.3 Data synthesis and statistical analysis

We reported continuous variables as means ± standard deviation and categorical variables as percentages. Calibration was infrequently reported, so we restricted meta-analysis to discrimination through a summary measure of c-statistic and corresponding 95% CI. In our primary analysis we assessed overall discrimination for models that had ≥ 3 EHR cohorts with c-statistic data. When multiple c-statistic data for a model were reported in a single cohort by different studies we only included the first published study.

We calculated the 95% prediction interval (PI) to depict the extent of between-study heterogeneity and to indicate a possible range for prediction model performance in a new validation.[186] When the 95% CI or PI of the summary c-statistic included 0.5 we concluded that there was insufficient evidence that the prediction model has statistically significant discriminatory ability.[183, 224] We used a Bayesian approach throughout as

frequentist methods, where there are fewer studies or a mixture of study sizes, have produced prediction intervals with poor coverage.[186]

All Bayesian meta-analysis models assume random effects by default. Results are based on the posterior median. Prediction intervals are directly obtained from the corresponding posterior quartiles. The standard model for random effects meta-analysis assumes that the 'true' performance is normally distributed within and across studies.[225] Within-study normality of performance estimates can be justified with this selection of included studies because they are all large. *Snell et al.* showed that the between-study distribution of the c-statistic on the original scale is not normally distributed when there is variability in the predictor effect across studies (which is likely in this selection of studies as they include different populations, and adopt slightly different definitions for predictors).[225] They found that the logit scale is more appropriate for the estimation of prediction interval. Consequently we used the "valmeta" function of the "metamisc" package in R software which applies a logit transformation to the c-statistic prior to calculation of summary c-statistic and prediction interval.[226]

For appropriate prior distributions we borrowed from earlier work by *Debray et al.* which recommended a half Student-*t* distribution with location *m*, scale σ, and *v* degrees of freedom where we set $m = 0$ and define σ equal to the largest empirical value of $\hat{\tau}$ (to allow for more extreme values of heterogeneity).[186] These hyperparameter values allow to penalise the extent of between-study heterogeneity when the number of included validation studies is low.[186] Further we also used $v = 3$ to ensure that the variance $\sigma^2 v/(v-2)$ exists and samples of τ were truncated above 10 to rule out unreasonable values. Thus the resulting priors are given as $\tau_{discr} \sim \text{Student}-t(0, 0.5^2, 3)T[0.10]$ which has been shown to allow for large but realistic values for between-study heterogeneity.[186]

We conducted meta-analyses in R using the metafor and metamisc package (R foundation for Statistical Computing 3.6.3).[226-228]

We performed a number of sensitivity analyses:

- To only include studies where the participants domain in PROBAST assessment was 'low' or 'unclear' risk of bias.
- To only include studies where the overall PROBAST assessment was 'low' or 'unclear' risk of bias.

- Where a cohort had been reported multiple times we replaced the meta-analysis data with the data on the same cohort from any later study.
- We excluded data from one of the Korean National Health Insurance Service Health screening cohort (NHIS-HEALS) and Korean National Health Insurance Service-based National Sample cohort (NHIS-NSC) because they originated from the same EHR database.

The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach was used to assess the certainty of the evidence.[229] The certainty of the evidence was graded as 'high' (further research is very unlikely to change our confidence in the effect estimate), 'moderate' (further research is likely to have an important impact on our confidence in the effect estimate), 'low' (further research is very likely to have an important impact on our confidence in the effect estimate and is likely to change the estimate) or 'very low' (any estimate of effect is very uncertain).

The initial certainty level of the included prediction modelling studies was set at 'high' because the association between the predictors and outcomes was considered irrespective of any causal connection.[230] Eight criteria were considered to further downgrade or upgrade the certainty of the evidence; five criteria which might downgrade the overall certainty of the evidence (methodological limitations of the study, indirectness, imprecision, inconsistency and likelihood of publication bias) and three which might potentially upgrade the overall certainty of the evidence (large effect, dose-response relation in the effect, and opposing plausible residual bias or confounding).

The criteria that might downgrade the overall certainty of evidence were considered as follows:

- Methodological limitations of the studies were assessed by considering the overall risk of bias judgement across studies based on the overall PROBAST risk of bias assessment. Indirectness was assessed by making a global judgement on how dissimilar the research evidence is to the research question at hand (in terms of population and outcomes across studies).
- Indirectness was assessed through concerns regarding the applicability of each included study from PROBAST (i.e. when the populations, predictors or outcomes of the study differ from the research question) and an overall judgement across studies was made.

- Imprecision was assessed by considering the optimal total number of events across all studies. A minimum threshold of 10 events per variable was considered as the minimum required in regression modelling development studies, and 100 when machine learning methods had been used.[206, 231] For external validation studies a minimum sample size of at least 200 events was less concerning for imprecision.[232] Results may also be imprecise when the 95% confidence intervals of c-statistic of all studies or of the largest studies include insufficient discrimination performance (0.5).
- A global judgement on inconsistency was evaluated through the consistency of the model discrimination performance and the range of the 95% PI as a statistical measure of heterogeneity. Widely differing estimates of the c-statistic indicated inconsistency or if the 95% PI of the summary c-statistic was wide and included 0.5.
- Publication bias was suspected when the body of evidence consisted of only positive studies from small sample sizes or all studies were funded by industry.

The criteria that might upgrade the overall certainty of evidence were considered as follows:

- A large magnitude of effect (i.e. highly discriminatory predictive performance) was considered if the c-statistic exceeded 0.7 in the majority of studies.[233]
- Dose-response relation in this effect was not applicable here since this review was not focused on drugs or pharmaceutical agents.
- Whether all plausible confounders and biases were accounted for is not applicable here as we only included studies that described a multivariable prediction model.

One investigator (RN) rated the certainty of the evidence for the primary outcome and this was checked by a second investigator (JW). The criteria used are discussed below.

2.5.4 Patient and public involvement

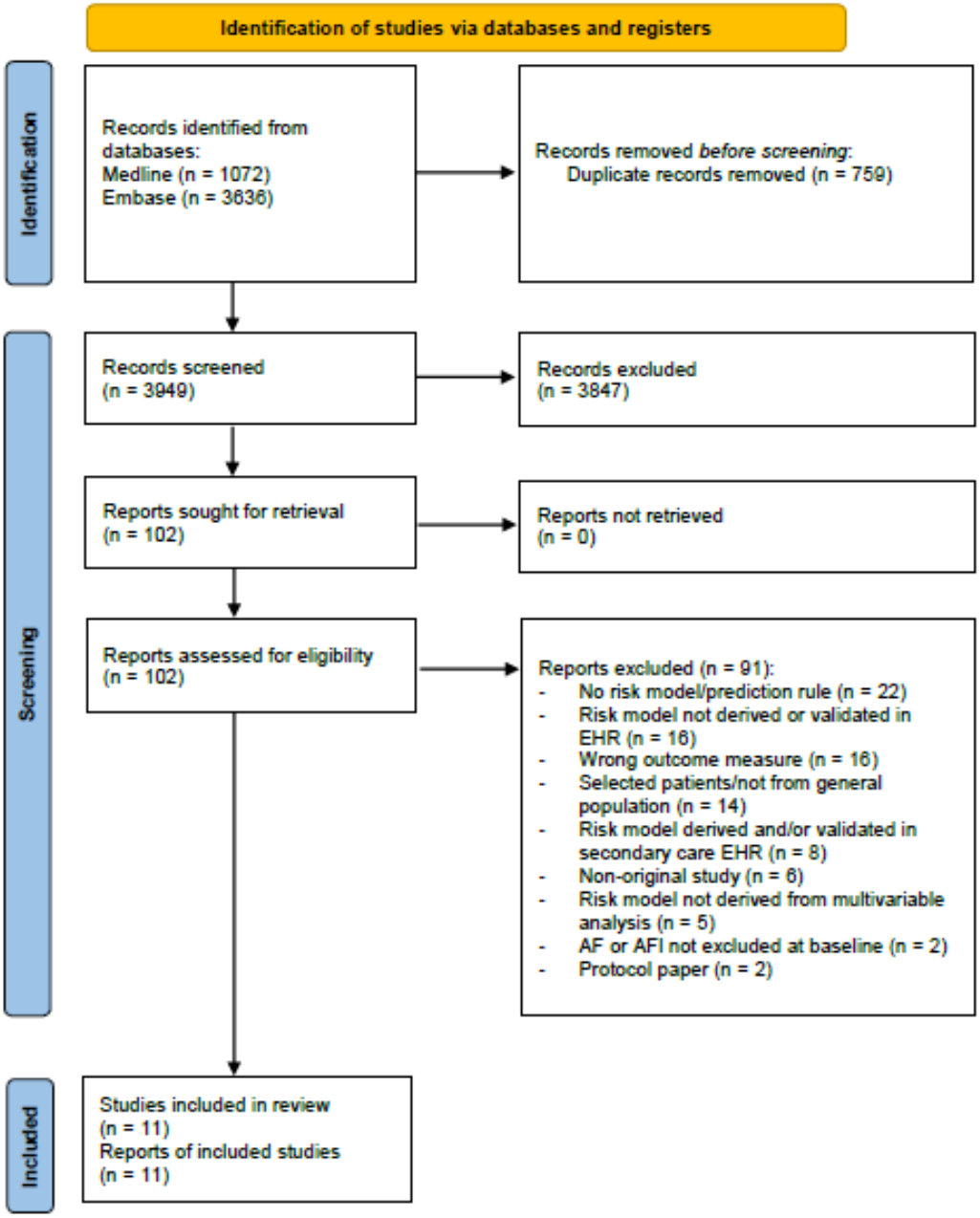
Patients or the public were not involved in the design, conduct, reporting or dissemination plans of our research**.**

2.6 Results

2.6.1 Study selection

The study selection process is described in Figure 1. We identified 3949 unique records, reviewed 102 full-text reports and included 11 studies.

Figure 1 Flow diagram of literature search.



Abbreviations: AF, atrial fibrillation; AFl, atrial flutter; EHR, electronic health record

2.6.2 Characteristics of included studies

The 11 included studies were based on nine cohorts from eight EHR databases, located in Asia Pacific (n =3), Europe (n = 3) and the Middle East (n = 2) (Table 3-4).[194, 196, 199, 234-241] The number of times a prediction model had been derived or validated in EHR was skewed to Asia Pacific (n = 17) compared with Europe (n = 5) and the Middle East (n = 3) (Table 5).

The total number of participants in the included studies was 17,889,536. Cohort size ranged from 96,778 to 2,994,837 (Table 4). The mean age varied from 41.3 to 65.7 years and the proportion of female participants ranged from 47.3% to 54.7%. The mean follow-up ranged from 2.9 years to 10.9 years (Table 3-4). The incidence of AF during follow-up ranged from 0.2% to 5.8% (Table 4).

Table 3 Characteristics of cohorts in included studies

Study	**Cohort (Country)**	**EHR description**	**Age (mean ± SD)**	**Women (%)**	**BMI (mean ± SD)**	**Diabetes (%)**	**Hyper-tension (%)**	**Heart failure (%)**
Aronson 2018[234]	MHS (IL)	Ambulatory clinics	62.0 ± 9.0	53.7	28.2 ± 5.1	13.5	34.3	1.00
Chao 2013[235]	NHIRD (TW)	National health insurance	41.3 ± 16.4	49.1	N/A	3.1	5.2	0.40
Hill 2019[194]	CPRD (UK)	Nationwide primary care	56.0 ± 14.5	53.4	27.6 ± 6.0	6.9	25.0	0.70
Himmelreich 2020[196]	Nivel-PCD (NL)	Nationwide primary care	65.5 ± 11.4	52.5	N/A	42.7	66.5	4.20
Hu-WS 2019[236]**	NHIRD (TW)	National health insurance	41.3	49.3	N/A	2.1	15.1	0.80
Kim 2020[241] **	NHIS-NSC (KR)	National Health Insurance	47.7	50.5	23.7	6.3	21.2	2.40
Li 2019[199] **	YMID (CN)	Regional Medical insurance	47.0	47.3	N/A	4.0	9.7	0.15
	NHIS-HEALS (KR)	National health examination program	56.1 ± 9.3	46.0	N/A	8.3	31.7	1.20
Saliba 2016[238]	ClalitHS (IL)	State-mandated health services	65.7 ± 11.2	54.7	N/A	25.3	48.9	4.30
Sekelj 2020[239]	Discover (UK)	Regional primary care	52.2 ± 13.3	51.0	27.0 ± 6.1	23.2	17.9	0.50

Suenari 2017[240]	NHIRD (TW)	National health insurance	42.4 ± 16.0	49.1	N/A	3.2	5.5	0.40

Abbreviations: BMI, body mass index; CN, China; ClalitHS, Clalit Health Services; CPRD, Clinical Practice Research Datalink; EHR, electronic health records; IL, Israel; KR, Republic of Korea; MHS, Maccabi Healthcare Services; NHIRD, National Health Insurance Research Database; NHIS-HEALS, National Health Insurance Service - Health screening Cohort; NHIS-NSC, National Health Insurance Service-based National Sample Cohort; Nivel-PCD, Netherlands Institute for Health Services Research Primary Care Database; NL, Netherlands; N/A, not available; SD, standard deviation; TW, Taiwan; UK, United Kingdom; YMID, Yunnan Medical Insurance Database

N.B. ** In Kim 2020, Li 2019 and Hu-WS 2019 the percentage of patients related to sex, diabetes, hypertension, heart failure was calculated from reported values categorised by incident AF or not.

Table 4 Characteristics of outcomes in included studies

Study	**Study aim**	**AF cases (n) / total patients (n) (%)**	**Outcome definition**	**Outcome coding**	**Enrolment period (mean F/U in years)**	**Exclusion criteria**
Aronson 2018[234]	D	5 660 / 96 778 (5.80)	AF, AFI	ICD codes	2005 – 2015 (10.0)	Past history of AF, incomplete follow-up
Chao 2013[235]	EV	9 187 / 702 502 (1.30)	AF*	ICD codes*	2000 – 2009 (9.0)	Age < 18 years, past history of cardiac arrhythmia, rheumatic heart disease
Hill 2019[194]	D, EV	95 607 /	AF, AFI	Read codes	2006 – 2016	Age < 30 years, past history of AF

		2 994 837 (3.19)			(N/A)	
Himmelreich 2020[196]	EV	5 264 / 111 475 (4.72)	AF, AFl	ICPC-1 codes	2013 – 2018 (N/A)	Age < 40 years, past history of AF
Hu-WS 2019[236]	EV	12 051 / 692 691 (1.74)	AF	ICD codes	1996 – 2013 (10.9)	Age < 18 years, past history of AF, incomplete data
Kim 2020[241]	D, EV	5 824 / 432 587 (1.35)	AF, AFl	ICD codes	2009 – 2013 (N/A)	Age < 18 years, past history of AF, mitral valve stenosis or prosthetic valve disease, missing data for smoking or alcohol, change in residence
Li 2019[199]	D, EV	921 / 471 446 (0.20)	AF	ICD codes	2001 – 2012 (4.1)	Past history of AF, incomplete data, readmission
	EV	12 143 / 451 199 (2.69)	AF	ICD codes	2002 – 2013 (7.3)	Past history of AF, mitral stenosis, prosthetic heart valves, valve replacement or valvuloplasty, or cardiomyopathy
Saliba 2016[238]	EV	23 223 / 1 062 073 (2.19)	AF	ICD codes	2012 – 2014 (2.9)	Age < 50 years, past history of AF

Sekelj 2020[239]	EV	17 880 / 604 135 (2.96)	AF, AFl	Read codes	2006 – 2013 (N/A)	Age < 30 years, past history of AF, incomplete data for height, weight, BMI, systolic BP and diastolic BP
Suenari 2017[240]	EV	9 174 / 670 804 (1.40)	AF	ICD codes	2000 – 2011 (9.0)	Age < 20 years, past history of cardiac arrhythmia

Abbreviations: AF, atrial fibrillation; AFl, atrial flutter; BMI, body mass index; BP, blood pressure; D, derivation; ECG, electrocardiogram; EV, external validation; F/U, follow-up; ICD, International Classification of Diseases; ICPC-1, International classification of Primary care version 1 diagnostic codes; N/A, not available.

N.B. *In Chao 2013 it is not reported how outcome was defined or measured but given the authors were using the same database as Suenari 2017, we have assumed outcomes were measured in the same way;

2.6.3 Characteristics of included prediction models

The included studies reported data on nine multivariable prediction models (Table 5). Three models had originally been derived for a purpose other than incident AF prediction.[187-189] Five models had been derived in community-based EHR; three using machine learning techniques.[194, 236, 241] In two of these studies, a range of machine learning techniques had been investigated with the optimum technique chosen by discriminative performance (Table 6).[194, 241] Amongst machine learning techniques, random forests were investigated in all three studies[194, 236, 241] and neural networks were considered in two.[194, 241]

Table 5 Characteristics of included prediction models

Model	**Study**	**Predicted outcome**	**Number of predictors**	**Derivation EHR cohort (country)**	**External validation EHR cohort (country)**
Models originally derived for another purpose but tested for prediction of incident AF					
$CHADS_2$	Gage 2001[187]	Stroke risk	5	-	ClalitHS (IL) NHIRD (TW) NHIS-HEALS (KR) NHIS-NSC (KR) YMID (CN)
CHA_2DS_2-VASc	Lip 2010[188]	Stroke risk	7	-	ClalitHS (IL) Nivel-PCD (NL) NHIS-HEALS (KR) NHIS-NSC (KR) YMID (CN)
HATCH	de Vos 2010[189]	Progression to persistent AF	5	-	NHIRD (TW) NHIS-HEALS (KR) NHIS-NSC (KR)

					YMID (CN)
Regression model derived in a prospective cohort design					
CHARGE-AF	Alonso 2013[205]	Incident AF or AFl	11	-	CPRD (UK) Nivel-PCD (NL)
Regression models derived in EHR					
C_2HEST	Li 2019[199]	Incident AF	6	YMID (CN)	NHIRD (TW) NHIS-HEALS (KR)
MHS	Aronson 2018[234]	Incident AF or AFl	10	MHS (IL)	N/A
Machine learning models derived in EHR					
CPRD	Hill 2019[194]	Incident AF or AFl	100	CPRD (UK)	Discover (UK)
NHIRD	Hu-WS 2019[236]	Incident AF	19	NHIRD (TW)	N/A
NHIS-NSC	Kim 2020[241]	Incident AF or AFl	22	NHIS-NSC (KR)	N/A

Abbreviations: AF, atrial fibrillation; AFl, atrial flutter; $CHADS_2$, Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, prior Stroke or transient ischemic attack [2 points]; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); ClalitHS, Clalit Health Services; CN, China; CPRD, Clinical Practice Research Datalink; EHR, electronic health records; HATCH, Hypertension, Age, stroke or Transient ischemic attack, Chronic obstructive pulmonary disease, Heart failure; IL, Israel; KR, Republic of Korea; MHS, Maccabi Healthcare Services; NHIRD, National Health Insurance Research Database; NHIS-HEALS, National Health Insurance Service - Health screening Cohort; NHIS-NSC, National Health Insurance Service-based National Sample Cohort; Nivel-PCD, Netherlands Institute for Health Services Research Primary Care Database; NL, Netherlands; TW, Taiwan; UK, United Kingdom; YMID, Yunnan Medical Insurance Database

Table 6 Performance of machine learning and traditional regression techniques during model development

Technique	Discrimination		Calibration	
	c-statistic	**95%CI**	**p-value of GOF test**	**O:E ratio**
Hill 2019 (CPRD)				
Neural network	0.818*	0.817 - 0.819	N/A	N/A
Random forest	0.812*	0.811 - 0.813	N/A	N/A
Support vector machine	0.811*	0.810 - 0.812	N/A	N/A
Logistic LASSO	0.811*	0.810 - 0.812	N/A	N/A
Traditional regression	0.797*	0.796 - 0.798	N/A	N/A
Hu-WS 2019 (NHIRD)				
Random forest	0.948	0.947 - 0.949	N/A	N/A
Kim 2020 (NHIS-NSC)				
Extreme gradient boosting	0.845	0.837 - 0.853	N/A	N/A
Random forest	0.838	0.830 - 0.846	N/A	N/A
Naïve Bayes	0.833	0.825 - 0.841	N/A	N/A
Deep neural network	0.813	0.800 - 0.826	N/A	N/A
Decision tree	0.801	0.787 - 0.815	N/A	N/A
Support vector machine	0.766	0.757 - 0.775	N/A	N/A
Traditional Regression	0.684	0.675 - 0.693	N/A	N/A

Abbreviations: CI, Confidence Interval; CPRD, Clinical Practice Research Datalink; GOF, goodness-of-fit; LASSO, least absolute shrinkage and selection operator; N/A, not available; NHIRD, National Health Insurance Research Database; NHIS-NSC,

National Health Insurance Service-based National Sample Cohort; O:E, observed versus expected events

N.B. * 95% CI for c-statistic not reported in article, so estimated from the reported c-statistic according to methods described by Debray *et al.* 2017

All studies reported a measure of discrimination (either c-statistic or AUROC), but only two studies provided a measure of calibration.[196, 234] Three prediction models – CPRD (Clinical Practice Research Datalink), C_2HEST and HATCH - showed a c-statistic greater than 0.75 in an external validation study (Table 7).[237, 239]

Table 7 Outcomes of studies reporting on prediction models

Prediction Model	Study aim	Study (cohort)	Observed AF/total population (%)	Discrimination		Calibration		Follow up duration (years)	RoB Partici-pants domain	RoB Overall
				c-statistic	95%CI	p-value of GOF test	O:E ratio			
Models originally derived for another purpose but tested for prediction of incident atrial fibrillation										
$CHADS_2$	EV	Chao 2013 (NHIRD)	9,187 / 702,502 (1.30)	0.713	0.707 - 0.719	N/A	N/A	10	L	H
	EV	Saliba 2016 (ClalitHS)	23,223 / 1,062,073 (2.19)	0.728	0.711 - 0.731†	N/A	N/A	3	U	H
	EV	Li 2019 (YMID)	921 / 471,446 (0.20)	0.632	0.604 - 0.660	N/A	N/A	11	L	H
	EV	Li 2019 (NHIS-HEALS)	12,143 / 451,199 (2.69)	0.637	0.632 - 0.642	N/A	N/A	11	H	H
	EV	Kim 2020 (NHIS-NSC)	5,824 / 432,587 (1.35)	0.652	0.646 - 0.657	N/A	N/A	5	H	H

CHA_2DS_2-VASc	EV	Saliba 2016 (ClalitHS)	23,223 / 1,062,073 (2.19)	0.744	0.741 - 0.747	N/A	N/A	3	U	H
	EV	Li 2019 (YMID)	921 / 471,446 (0.20)	0.687	0.659 - 0.716	N/A	N/A	11	L	H
	EV	Li 2019 (NHIS-HEALS)	12,143 / 451,199 (2.69)	0.637	0.632 - 0.642	N/A	N/A	11	H	H
	EV	Himmelreich 2020 (Nivel-PCD)	5,264 / 111,475 (4.72)	0.669	0.661 - 0.677	N/A	N/A	5	L	H
	EV	Kim 2020 (NHIS-NSC)	5,824 / 432,587 (1.35)	0.654	0.646 - 0.661	N/A	N/A	5	H	H
HATCH	EV	Suenari 2017 (NHIRD)	9,174 / 670,804 (1.40)	0.716	0.710 - 0.723	N/A	N/A	9	L	U
	EV	Li 2019 (YMID)	921 / 471,446 (0.20)	0.633	0.598 - 0.667	N/A	N/A	11	L	H
	EV	Li 2019 (NHIS-HEALS)	12,143 / 451,199 (2.69)	0.646	0.641 - 0.651	N/A	N/A	11	H	H

	EV	Kim 2020 (NHIS-NSC)	5,824 / 432,587 (1.35)	0.669	0.661 - 0.676	N/A	N/A	5	H	H
	EV	Hu-WS 2020 (NHIRD)	12,051 / 692,691 (1.74)	0.771*	0.767 - 0.775	N/A	N/A	14	L	H
Machine Learning models										
CPRD	D	Hill 2019 (CPRD)	95,607 / 2,994,837 (3.19)[+]	0.827*	0.826 - 0.828	N/A	N/A	11	L	H
	EV	Sekelj 2020 (Discover)	17,880 / 604,135 (2.96)	0.870*	0.867 - 0.873	N/A	N/A	8	L	H
NHIRD	D[#]	Hu-WS 2019 (NHIRD)	14,212 / 682,237 (2.08)	0.948	0.947 - 0.949	N/A	N/A	14	L	H
NHIS-NSC[§]	D	Kim 2020 (NHIS-NSC)	5,824 / 432,587 (1.35)	0.845	0.837 - 0.853	N/A	N/A	5	H	H
Regression Models derived in electronic health records										
C_2HEST	D	Li 2019 (YMID)	921 / 471,446 (0.20)	0.750	0.730 - 0.770	N/A	N/A	11	L	H
	EV	Li 2019 (NHIS-HEALS)	12,143 / 451,199 (2.69)	0.654	0.649 - 0.659	N/A	N/A	11	H	H

	EV	Hu-WS 2020 (NHIRD)	12,051 / 692,691 (1.74)	0.790*	0.785 - 0.793	N/A	N/A	14	L	H
MHS	D	Aronson 2018 (MHS)	5,660 / 96,778 (5.80)	0.743	0.737 - 0.749	N/A	0.970**	10	L	H
Regression model derived in a prospective cohort design										
CHARGE-AF	EV	Hill 2019 (CPRD)	95,607 / 2,994,837 (3.19)$^{+}$	0.725*	0.723 - 0.727	N/A	N/A	11	L	H
	EV	Himmelreich 2020 (Nivel-PCD)	5,264 / 111,475 (4.72)	0.736	0.727 - 0.744	0.001	0.69	5	L	H

Abbreviations: AF, Atrial Fibrillation; $CHADS_2$, Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, prior Stroke or transient ischemic attack [2 points]; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); CI, Confidence Interval; ClalitHS, Clalit Health Services; CPRD, Clinical Practice Research Datalink; D, derivation; EHR, electronic health records; EV, external validation; GOF, goodness-of-fit; H, high; HATCH, Hypertension, Age, stroke or Transient ischemic attack, Chronic obstructive pulmonary disease, Heart failure; L, low; MHS, Maccabi Healthcare Services; NHIRD, National Health Insurance Research Database; NHIS-HEALS, National Health Insurance Service - Health screening Cohort; NHIS-NSC, National Health Insurance Service-based National Sample Cohort; Nivel-PCD, Netherlands Institute for Health Services Research Primary Care Database; N/A, not available; O:E, observed versus expected events; ROB, risk of bias; U, unclear; YMID, Yunnan Medical Insurance Database

N.B. * 95% CI for c-statistic not reported in article, so estimated from the reported c-statistic according to methods described by Debray *et al.* 2017; ** For Aronson 2018 the reported O:E was extracted by Himmelreich *et al.* 2020; [#] In Hu-WS 2019 the authors do an EV but in a subset of the NHIRD dataset pertaining to secondary care inpatients, preventing us from including this data into this review; † In Saliba 2016 the 95% upper CI for c-statistic is reported as 0.725 but this is less than the stated c-statistic of 0.728, so the 95% upper CI has been estimated from the reported c-statistic according to methods described by Debray *et al.* 2017; [+] In Hill 2019 a total of 2,994,837 patients were included in the baseline model with 167,672 included in the time-varying model. The number of events are not differentiated between baseline and time-varying model. This dataset was divided between training (1,996,788) and holdout (998,049) for testing but number of events in each are not reported. For the EV of CHARGE-AF it is not specified which subset of the data is used for validation; [§] In Kim 2020 prediction model development using machine learning was completed both with and without the predictor $PM_{2.5}$ - which is fine particular matter air pollution. In this analysis we have only included the model without $PM_{2.5}$ as it is judged not to be a predictor that would be routinely available in primary care or population EHR.

Table 8 and 9 summarise the variables used. The ten most frequently included variables are summarised in Figure 2. Age and chronic heart failure were the only variables included in every model. The number of variables incorporated into machine learning models was far greater than traditional regression models (Table 5). The CPRD model was unique in incorporating time-varying variables (e.g. change in body mass index (BMI) between the last two quarters of the year).[194]

Table 8 Baseline variables used in prediction models

Model	**Predictors**				
	Patient characteristics	**Medical History**	**Physical measurements**	**Investigations**	**Other**
Models originally derived for another purpose but tested for prediction of incident atrial fibrillation					
$CHADS_2$	Age	Hypertension, CHF, diabetes mellitus, CVA			
CHA_2DS_2-VASc	Age, sex	Hypertension, CHF, stroke/TIA/thromboembolism, vascular disease			
HATCH	Age	Hypertension, CHF, stroke/TIA, COPD			
Machine Learning models					
CPRD	Age, sex, race, smoking status	Hypertension, anti-hypertensive medication, CHF, congenital heart disease, MI, LVH, type 1 DM, type 2 DM	Height, weight, BMI, SBP, DBP		
NHIRD	Age (years), age group, sex	Hypertension, CHF, COPD, rheumatological disease, dyslipidaemia, DM, CVA or TIA, sleep disorder, cancer,			Follow-up duration (years), mean CHA_2DS_2-VASC score

		hyperthyroidism, vascular disease, gout, CKD or ESRD, anaemia			
NHIS-NSC*	Age, sex, smoking (pack-year), alcohol	Hypertension, CHF, MI, vascular disease, stroke/TIA, COPD	BMI, SBP	Triglycerides, total cholesterol, HDL cholesterol, LDL cholesterol, eGFR, GGT, fasting blood glucose, Haemoglobin, AST	Socioeconomic status
Regression Models derived in electronic health records					
C_2HEST	Age	Hypertension, ischaemic heart disease, CHF, COPD, thyroid disease			
MHS	Age, sex	Anti-hypertensive medication, MI, CHF, peripheral vascular disease, inflammatory disease in a female, COPD	BMI, SBP		
Regression model derived in a prospective cohort design					
CHARGE-AF	Age, race, smoking status	Anti-hypertensive medication, MI, CHF, DM	Height, weight, SBP, DBP		

Abbreviations: AST, aspartate aminotransferase; BMI, body mass index; $CHADS_2$, Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, prior Stroke or transient ischemic attack [2 points]; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points],

Stroke/transient ischemic attack/thromboembolism [2 points]; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); CHF, chronic heart failure; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; CPRD, Clinical Practice Research Datalink; CVA, cerebrovascular accident; DBP, diastolic blood pressure; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; GGT, gamma glutamyl transferase; HATCH, Hypertension, Age, stroke or Transient ischemic attack, Chronic obstructive pulmonary disease, Heart failure; HDL, high density lipoprotein; L, low; LDL, low density lipoprotein; LVH, left ventricular hypertrophy; MHS, Maccabi Healthcare Services; MI, myocardial infarction; NHIRD, National Health Insurance Research Database; NHIS-HEALS, National Health Insurance Service - Health screening Cohort; NHIS-NSC, National Health Insurance Service-based National Sample Cohort; SBP, systolic blood pressure; TIA, transient ischaemic attack

N.B. * In Kim 2020 prediction model development using machine learning was completed both with and without the predictor $PM_{2.5}$ - which is fine particular matter air pollution. In this analysis we have only included the model without $PM_{2.5}$ as it is judged not to be a predictor that would be routinely available in primary care or population EHR.

Table 9 Time-varying variables in CPRD model of *Hill et al*[194]

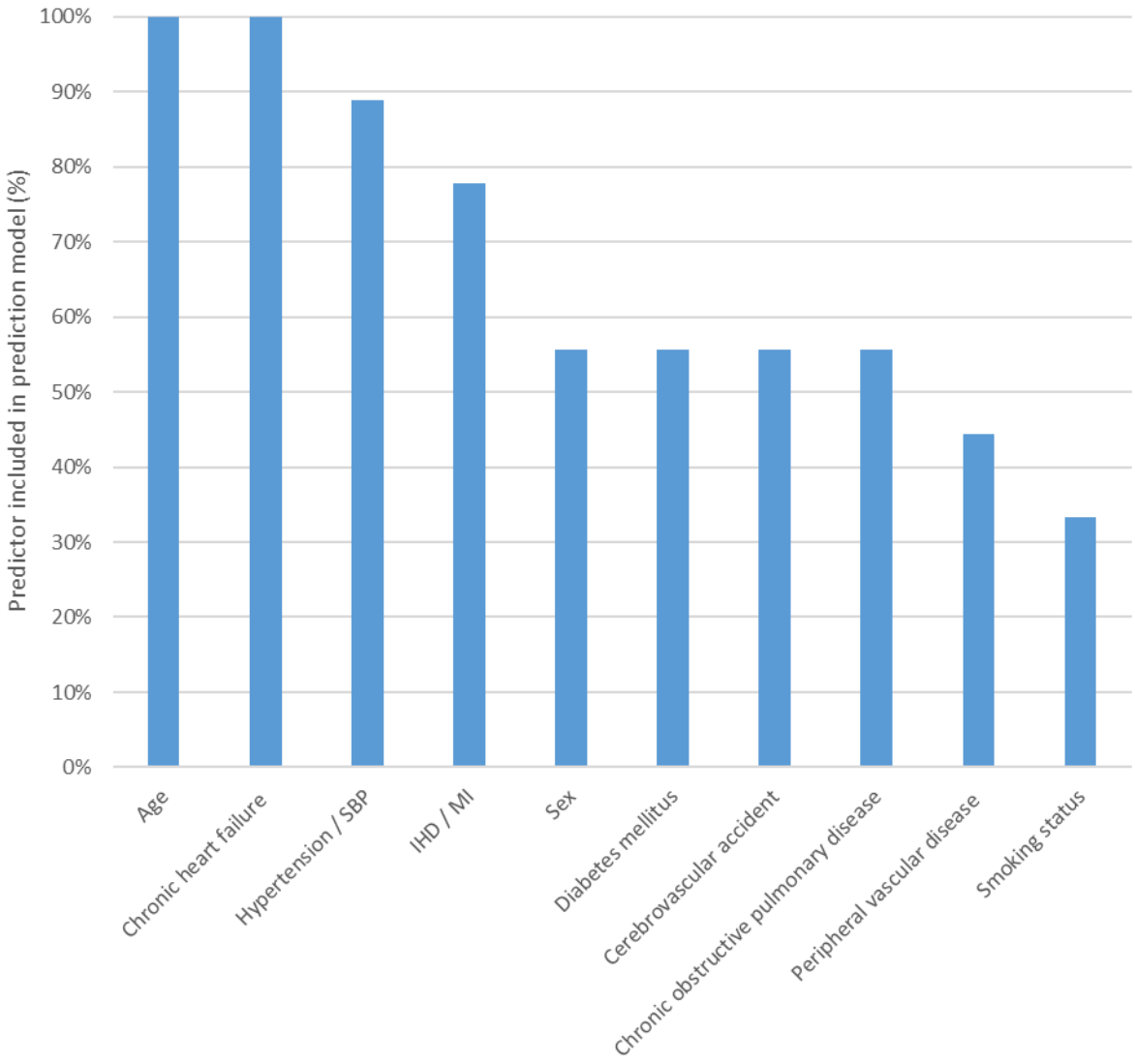
Variable	**Description of time-varying component**
Patient Characteristics	
Age	Age in years at start of each 91-day quarter
Sex	Male or female
Race	Known white or other

Smoking status	Known current smoker or other
Height	Latest recorded value
Weight	A new set of predictors was derived using clinical measurements over the year prior to AF date (or equivalent for matched non-AF patients): • latest value recorded in each quarter • difference between latest and earliest values recorded in total • difference between min and max values in each quarter • difference between min and max values across successive quarters • difference between min and max values recorded in total • number of measurements recorded in each quarter • number of measurements recorded in total
BMI	
DBP	
SBP	
Medical History	
Hypertension	For each comorbidity, a new set of predictors was derived to indicate whether an event was observed in each quarter over the year prior to AF diagnosis (or equivalent for matched non-AF patients), or at any time prior to this
Anti-hypertensive medication	
CHF	
Ischaemic heart disease	
Congenital heart disease	

MI	
LVH	
Type 1 DM	
Type 2 DM	

Abbreviations: AF, Atrial Fibrillation; AST, aspartate aminotransferase; BMI, body mass index; $CHADS_2$, Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, prior Stroke or transient ischemic attack [2 points]; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points]; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); CHF, chronic heart failure; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; CPRD, Clinical Practice Research Datalink; CVA, cerebrovascular accident; DBP, diastolic blood pressure; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; GGT, gamma glutamyl transferase; HATCH, Hypertension, Age, stroke or Transient ischemic attack, Chronic obstructive pulmonary disease, Heart failure; HDL, high density lipoprotein; L, low; LDL, low density lipoprotein; LVH, left ventricular hypertrophy; MHS, Maccabi Healthcare Services; MI, myocardial infarction; NHIRD, National Health Insurance Research Database; NHIS-HEALS, National Health Insurance Service - Health screening Cohort; NHIS-NSC, National Health Insurance Service-based National Sample Cohort; SBP, systolic blood pressure; TIA, transient ischaemic attack

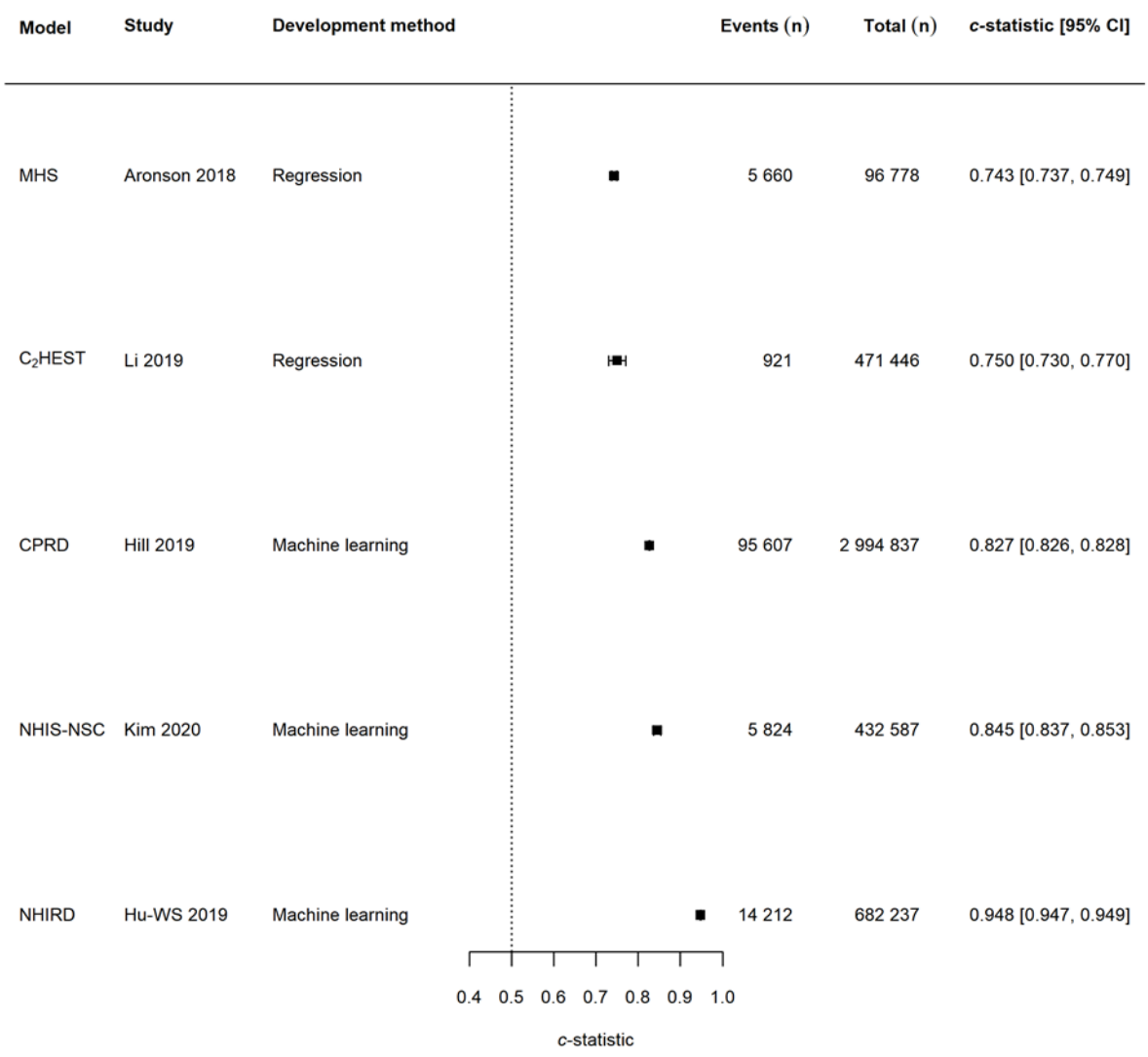
Figure 2 An overview of the ten predictors most frequently incorporated in the prediction models in this study.



Abbreviations: IHD, ischaemic heart disease; MI, myocardial infarction; SBP, systolic blood pressure

Figure 3 plots the performance of traditional regression and machine learning models in the development population of each study. Table 6 summarises the performance of traditional regression and machine learning techniques during model development in the CPRD and NHIS-NSC datasets. In each case, machine learning produced stronger discriminative performance in the development population.

Figure 3 Forest plot showing the performance of traditional regression versus machine learning models using the development data from each relevant study



Abbreviations: C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); CI, Confidence Interval; CPRD, Clinical Practice Research Datalink; D, derivation; MHS, Maccabi Healthcare Services; NHIRD, National Health Insurance Research Database; NHIS-NSC, National Health Insurance Service-based National Sample Cohort

2.6.4 Risk of bias assessment

Table 10 shows the results of the risk of bias and applicability assessment for each PROBAST domain for each model in the included studies. Figure 3 gives an overall

summary of PROBAST domain assessments across all included studies. Overall, 96% of model results were at high risk of bias predominantly driven by high risk of bias in the analysis domain (88%). This resulted from exclusion of participants with missing data from analysis (72%) or not mentioning missing data (16%).

Table 10 Risk of bias and applicability assessment for each Prediction model Risk of Bias ASsessment Tool domain

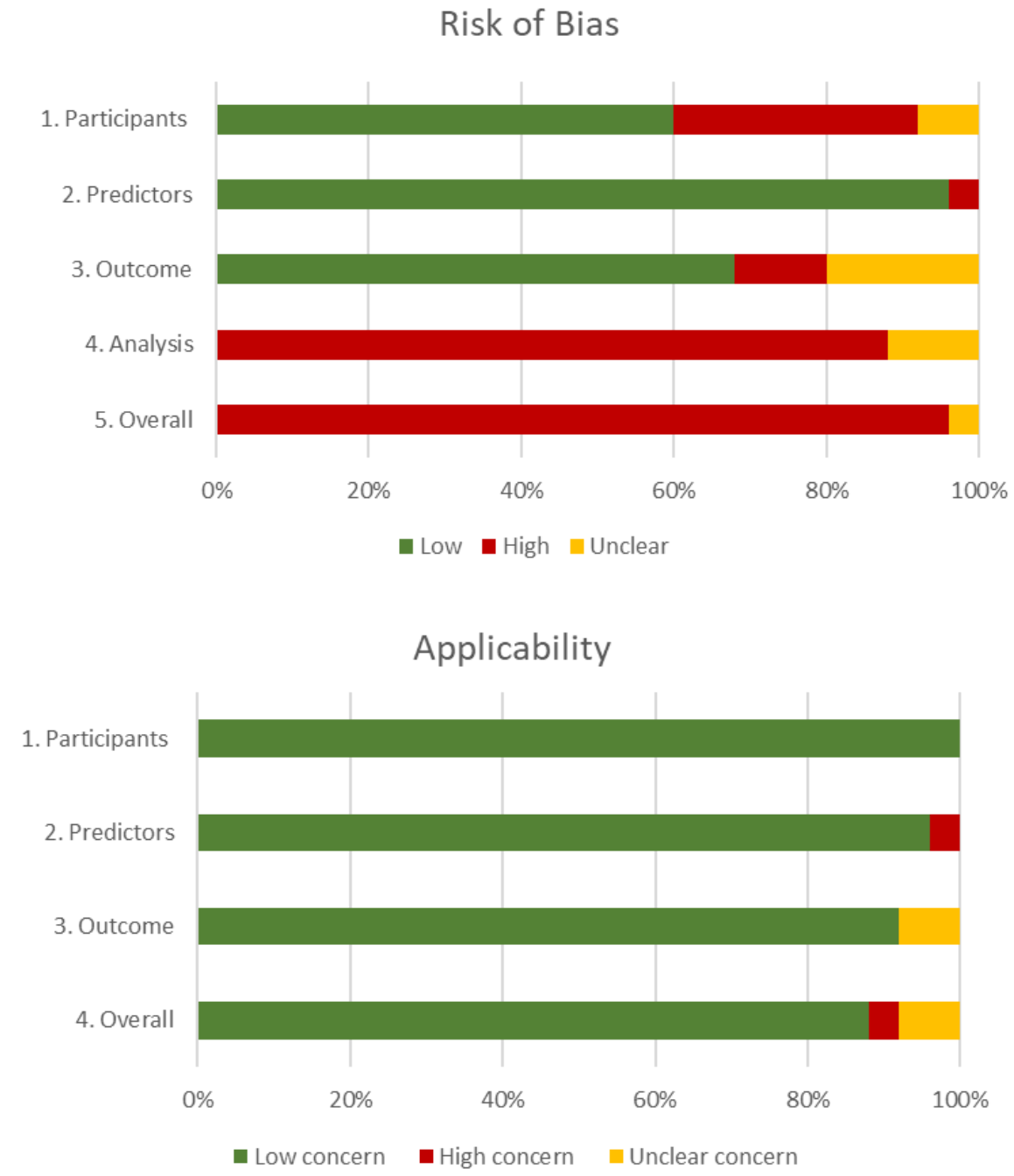
Study	Model	Aim	RoB Participants	RoB Predictors	RoB Outcome	RoB Analysis	Applicability Participants	Applicability Predictors	Applicability Outcomes	Overall RoB	Overall Applicability
Aronson 2018	MHS	D	L	L	L	H	L	L	L	H	L
Chao 2013	$CHADS_2$	EV	L	L	U	H	L	L	L	H	L
Hill 2019	CPRD	D	L	L	H	U	L	L	L	H	L
Hill 2019	CHARGE-AF	EV	L	L	H	U	L	L	L	H	L
Himmelreich 2020	CHARGE-AF	EV	L	L	L	H	L	L	L	H	L
Himmelreich 2020	CHA_2DS_2-VASc	EV	L	L	L	H	L	L	L	H	L
Hu-WS 2019	NHIRD	D	L	H	L	H	L	H	L	H	H

Hu-WS 2020	C_2HEST	EV	L	L	U	H	L	L	U	H	U
Hu-WS 2020	HATCH	EV	L	L	U	H	L	L	U	H	U
Kim 2020	NHIS-NSC	D	H	L	L	H	L	L	L	H	L
Kim 2020	$CHADS_2$	EV	H	L	L	H	L	L	L	H	L
Kim 2020	CHA_2DS_2-VASc	EV	H	L	L	H	L	L	L	H	L
Kim 2020	HATCH	EV	H	L	L	H	L	L	L	H	L
Li 2019	C_2HEST	D	L	L	L	H	L	L	L	H	L
Li 2019	C_2HEST	EV	H	L	L	H	L	L	L	H	L
Li 2019	$CHADS_2$	EV (YMID)	L	L	L	H	L	L	L	H	L
Li 2019	$CHADS_2$	EV (NHIS-HEALS)	H	L	L	H	L	L	L	H	L

Li 2019	CHA_2DS_2-VASc	EV (YMID)	L	L	L	H	L	L	L	H	L
Li 2019	CHA_2DS_2-VASc	EV (NHIS-HEALS)	H	L	L	H	L	L	L	H	L
Li 2019	HATCH	EV (YMID)	L	L	L	H	L	L	L	H	L
Li 2019	HATCH	EV (NHIS-HEALS)	H	L	L	H	L	L	L	H	L
Saliba 2016	$CHADS_2$	EV	U	L	U	H	L	L	L	H	L
Saliba 2016	CHA_2DS_2-VASc	EV	U	L	U	H	L	L	L	H	L
Sekelj 2020	CPRD	EV	L	L	H	H	L	L	L	H	L
Suenari 2017	HATCH	EV	L	L	L	U	L	L	L	U	L

Abbreviations: $CHADS_2$, Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, prior Stroke or transient ischemic attack [2 points]; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); ClalitHS, Clalit Health Service; CPRD, Clinical Practice Research Datalink; D, derivation; EV, external validation; H, high; HATCH, Hypertension, Age, stroke or Transient ischemic attack, Chronic obstructive pulmonary disease, Heart failure; L, low; MHS, Maccabi Healthcare Services; NHIRD, National Health Insurance Research Database; NHIS-HEALS, National Health Insurance Service - Health screening Cohort; NHIS-NSC, National Health Insurance Service (of Korea)-based National Sample Cohort; RoB, risk of bias; U, unclear; YMID, Yunnan Medical Insurance Database

Figure 4 Judgements on the four Prediction model Risk of Bias ASsessment Tool risk of bias domains and three applicability domains presented as percentages across all included studies.



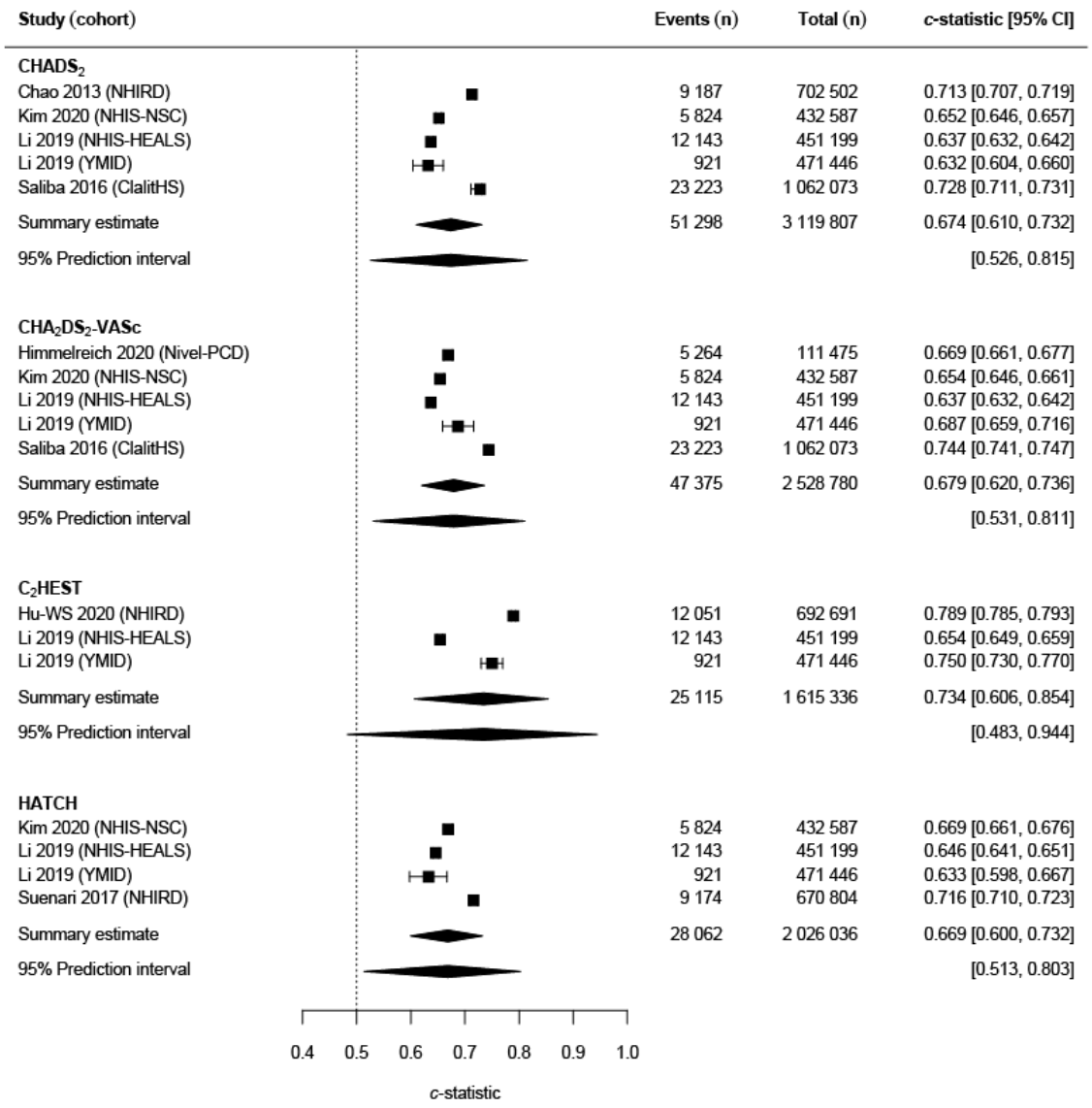
Abbreviations: ROB, risk of bias

2.6.5 Meta-analysis

Four models were eligible for the primary meta-analysis, incorporating 9,289,959 patients (Figure 5). Only C_2HEST was derived specifically for the purpose of predicting

incident AF.[199] There were three models that resulted in a summary c-statistic with statistically significant 95% PI in our primary meta-analysis: $CHADS_2$ (summary c-statistic 0.674; 95% CI 0.610 – 0.732; 95% PI 0.526 – 0.815; n = 5 studies; n = 3,119,807), CHA_2DS_2-VASc (summary c-statistic 0.679; 95% CI 0.620 – 0.736; 95% PI 0.531 – 0.811; n = 5 studies; n = 2,528,780) and HATCH (summary c-statistic 0.669; 95% CI 0.600 – 0.732; 95% PI 0.513 – 0.803; n = 4 studies; n = 2,026,036). There was high heterogeneity, as shown by the wide 95% PIs (Figure 5).

Figure 5 Forest plot of primary analysis of c-statistics.



Abbreviations: C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥ 75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); $CHADS_2$, Congestive heart failure, Hypertension,

Age > 75, Diabetes mellitus, prior Stroke or transient ischemic attack [2 points]; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age > 75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex category; CI, confidence interval; ClalitHS, Clalit Health Services; HATCH, Hypertension, Age, stroke or Transient ischaemic attack, Chronic obstructive pulmonary disease, and Heart failure; NHIRD, National Health Insurance Research Database; NHIS-HEALS, National Health Insurance Service - Health screening Cohort; NHIS-NSC, National Health Insurance Service-based National Sample Cohort; Nivel-PCD, Netherlands Institute for Health Services Research Primary Care Database; YMID, Yunnan Medical Insurance Database

Table 11 shows the results of the sensitivity analyses. Only CHA_2DS_2-VASc maintained a summary c-statistic with statistically significant 95% PI when either restricting the primary analysis to studies with ‘low’ or ‘unclear’ risk of bias for the participants domain of PROBAST, or using later data when a cohort had been analysed multiple times, or excluding data from either of the NHIS-HEALS or NHIS-NSC cohorts. However, when restricting primary analysis to models with ‘low’ or ‘unclear’ risk of bias for overall PROBAST assessment, no models met eligibility for inclusion.

Table 11 Sensitivity analyses

Comparison	**Summary c-statistic**	**95%CI**	**95%PI**	**Studies (n)**	**Patients (n)**
$CHADS_2$					
Primary meta-analysis	0.674	0.610-0.732	0.526-0.815	5	3,119,807
Excluding studies with High ROB in participants domain of PROBAST	0.694	0.581-0.798	0.478-0.887	3	2,236,021
Exclude data from NHIS-NSC	0.680	0.595-0.754	0.492-0.836	4	2,687,220
NHIRD data by Hu-WS 2020 not Suenari 2017 and data from NHIS-NSC rather than NHIS-HEALS	0.684	0.606-0.759	0.514-0.843	4	2,668,608
CHA_2DS_2-VASc					
Primary meta-analysis	0.679	0.620-0.736	0.531-0.811	5	2,528,780
Excluding studies with High ROB in participants domain of PROBAST	0.702	0.603-0.795	0.510-0.877	3	1,644,994
Exclude data in NHIS-NSC	0.690	0.602-0.758	0.520-0.850	4	2,096,193
NHIRD data by Hu-WS 2020 not Suenari 2017 and data from NHIS-NSC rather than NHIS-HEALS	0.690	0.618-0.760	0.530-0.835	4	2,077,581

HATCH					
Primary meta-analysis	0.669	0.600-0.732	0.513-0.803	4	2,026,036
NHIRD data by Hu-WS 2020 not Suenari 2017	0.684	0.586-0.782	0.467-0.880	4	2,047,923
Exclude data from NHIS-NSC	0.668	0.561-0.769	0.460-0.861	3	2,286,140
NHIRD data by Hu-WS 2020 not Suenari 2017 and data from NHIS-NSC rather than NHIS-HEALS	0.696	0.558-0.822	0.436-0.931	3	1,596,724

Abbreviations: $CHADS_2$, Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, prior Stroke or transient ischemic attack [2 points]; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex category; CI, Confidence Interval; HATCH, Hypertension, Age, stroke or Transient ischemic attack, Chronic obstructive pulmonary disease, Heart failure; NHIRD, National Health Insurance Research Database; NHIS-HEALS, National Health Insurance Service - Health screening Cohort; NHIS-NSC, National Health Insurance Service - based National Sample Cohort; PI, Prediction Interval; PROBAST, Prediction model Risk of Bias ASsessment Tool; ROB, Risk of bias

2.6.6 Certainty of evidence

The initial certainty level of the included prediction modelling studies was set at 'high' because the association between the predictors and outcomes was considered irrespective of any causal connection.[230] The overall certainty level was, however, downgraded to 'moderate' and then 'low' because of inconsistent results given high heterogeneity and the overall risk of bias was considered high in 96% of studies . The final overall certainty of 'low' implies that our confidence in the effect estimates is limited and further research is very likely to change the effect estimate.

2.7 Discussion

This systematic review and meta-analysis identified nine models that have been derived and/or validated in community-based EHR for incident AF. Five had been derived in EHR for this purpose; three by machine learning methods. Three models ($CHADS_2$, CHA_2DS_2-VASC and HATCH) produced a summary c-statistic with statistically significant 95% PI for prediction of incident AF despite high heterogeneity. However the summary c-statistics were only 0.669 – 0.679. For an outcome such as AF that is considered difficult to predict a c-statistic of 0.75 may be adequate for the models to be useful.[207] This threshold has been achieved by prediction models for incident AF in the community in non-EHR-based external validation studies,[191, 242, 243] as well as in EHR by the machine learning CPRD model.[239] Furthermore, in sensitivity analyses no model met eligibility for inclusion in meta-analysis if studies at overall high risk of bias were excluded.

A previous meta-analysis investigated prediction models for incident AF that had been derived or validated in community cohorts.[183] Nevertheless, this review included predominantly carefully-curated prospective cohort designs, the results from which will have limited generalisability. In addition, a number of the included models require variables, such as ECG parameters, that are not routinely available in community-based EHR.[244] The authors found CHA_2DS_2-VASc and CHARGE-AF (Cohorts for Heart and Aging Research in Genomic Epidemiology) resulted in a summary c-statistic with statistically significant 95% PI on meta-analysis. There is conflicting evidence as to how well CHARGE-AF performs in EHR, especially given the incompleteness of structured EHR fields for height, weight and ethnicity,[196, 204] and for our study it did not meet eligibility for inclusion into meta-analysis. Another systematic review summarised a similar selection of prediction models for the detection of AF in the community and externally validated these models head-to-head in a commercial screening cohort.[221] However, the outcome was prevalence, rather than future incident AF. Both of these

reviews predated the emergence of machine learning models in this field, which are summarised for the first time regarding the prediction of incident AF here.

The use of age alone to target screening strategies for incident AF has yet to show a benefit for systematic versus opportunistic screening, which is reflected in international guidelines.[1, 219] Prediction models could target screening and if implemented through primary care EHR would minimise extra resources. The use of CHA_2DS_2-VASc for prediction of incident AF has advantages given it uses variables available with high completeness in primary care EHR and would simultaneously provide an assessment of stroke risk as an indicator of eligibility for anticoagulation. Even so, there are a number of limitations. First, the discriminative performance was only moderate, overall certainty in the estimate effects was 'low' and the vast majority of studies were at high risk of bias. Second, it has predominantly been validated in Asia Pacific countries, where cohorts had different baseline characteristics compared with European counterparts. Third, it was outperformed by CHARGE-AF and C_2HEST when compared head-to-head in individual external validation studies.[196, 199]

Efforts may be best served to develop and externally validate novel prediction models for incident AF in community-based EHR. These data sources offer large samples sizes, providing the opportunity to investigate a larger number of predictors and utilise novel techniques. Machine learning models in this review showed strong discriminative performance in development datasets but were not included in meta-analysis due to a sparsity of external validation.

This study has a number of strengths. We had a comprehensive search strategy and thorough analysis approach. We included any model that had been used to predict the risk of incident AF, which allowed us to include models that were not originally intended for predicting AF but may have merits. We only included models that had been tested in databases relevant to the general population, which ensures the applicability of our results for screening in a primary care setting. We also did not present meta-regression or subgroup meta-analysis to investigate heterogeneity between studies based on study-level characteristics or sub-groups in the absence of available individual patient data given that such analyses would be prone to ecological bias.[245]

There are limitations to our study. Meta-analysis of model calibration performance was prohibited by poor reporting. We did not assess for 'reporting biases' visually through a funnel plot for several reasons. First, some studies reported multiple models in the

same cohort so incorporating all these data points would skew the plot; second, producing funnel plots for individual models would not be informative as there would be too few data points; third the sample sizes for all included studies was very large making small-study effects less likely. The vast majority of studies were at high risk of bias, which is consistent with previous literature on clinical prediction models due to limitations in conduct and reporting.[246] We restricted our search to studies written in English, though this has not been found to lead to significant bias.[247] Finally, routinely-collected databases are associated with a number of potential biases relating to their retrospective, observational nature.

2.8 Conclusions

In this systematic review with meta-analysis, we identified nine multivariable prediction models relevant to screening for incident AF using community-based EHR. On meta-analysis three models produced a summary c-statistic with statistically significant 95% PI, but discriminative performance was only moderate. At present, due to a combination of high risk of bias and inconsistency, there is no high performing prediction model for incident AF using primary care EHR. Future research could aim to develop models in primary care EHR using machine learning, but must better handle missing data, report calibration and provide external validation.

Chapter 3 Risk of atrial fibrillation and association with other diseases: protocol of the derivation and international external validation of a prediction model using nationwide population-based electronic health records

Ramesh Nadarajah, Jianhua Wu, Ronen Arbel, Moti Haim, Doron Zahger, Talish Razi Benita, Lior Rokach, Campbell Cowan, Chris P Gale

3.1 Summary of the publication

- This paper presents the approach that has been used to develop prediction models for incident AF, as well as understand the association between predicted AF risk and non-AF outcomes, during my PhD studies.
- The approach comprises 3 methodological steps within the UK dataset:
    - Develop a model for predicting short-term AF risk from data routinely available in community-based EHRs, comparing the performance of a random forest classifier with a multivariable logistic regression model and currently available models.
    - Quantify the association of predicted AF risk with a range of non-AF diseases and death.
    - Develop a parsimonious prediction model using logistic regression and clinically-recognised risk factors for AF.

3.2 Publication status

- Published 9 December 2023
- BMJ Open. 2023 Oct 4.

3.3 Abstract

3.3.1 Introduction

Atrial fibrillation (AF) is a major public health issue and there is rationale for the early diagnosis of AF, before the first complication occurs. Previous AF screening research is limited by low yields of new cases and strokes prevented in the screened populations. For AF screening to be clinically and cost-effective, the efficiency of

identification of newly diagnosed AF needs to be improved and the intervention offered may have to extend beyond oral anticoagulation for stroke prophylaxis. Previous prediction models for incident AF have been limited by their data sources and methodologies.

3.3.2 Methods and analysis

We will investigate the application of Random Forest and multivariable logistic regression to predict incident AF within a 6 months prediction horizon, that is a time-window consistent with conducting investigation for AF. The Clinical Practice Research Datalink (CPRD)-GOLD dataset will be used for derivation, and the Clalit Health Services dataset will be used for international external geographical validation. Analyses will include metrics of prediction performance and clinical utility. We will create Kaplan-Meier plots for individuals identified as higher and lower predicted risk of AF and derive the cumulative incidence rate for non-AF cardio-renal-metabolic diseases and death over the longer term to establish how predicted AF risk is associated with a range of new non-AF disease states.

3.3.3 Ethics and dissemination

Permission for CPRD-GOLD was obtained from CPRD (ref no: 19_076). The CPRD ethical approval committee approved the study. CHS Helsinki committee approval 21-0169 and data utilization committee approval 901. The results will be submitted as a research paper for publication to a peer-reviewed journal and presented at peer-reviewed conferences.

3.3.4 Trial registration details

A systematic review to guide the overall project was registered on PROSPERO (registration number CRD42021245093). The study was registered on Clinical Trials.gov (NCT05837364).

3.4 Introduction

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia. Over the last 20 years the number of new cases of AF diagnosed each year has risen by 72%, and now surpasses the four most common causes of cancer combined.[11] Moreover, it is

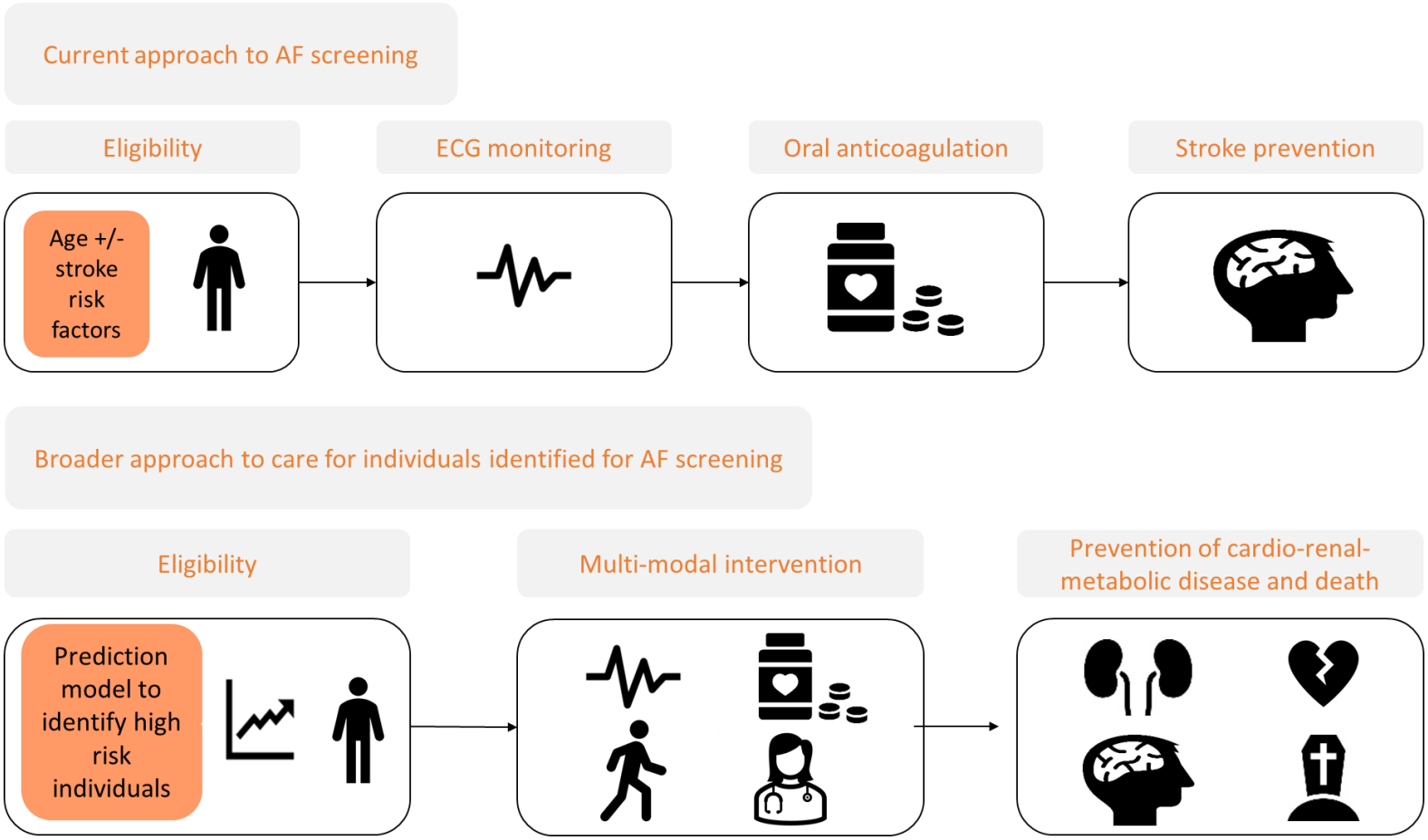
estimated that up to 35% of disease burden remains undiagnosed,[67] and 15% of strokes occur in the context of undiagnosed AF.[248]

Oral anticoagulants can reduce the risk of stroke by up to two thirds in those with AF at higher risk of stroke,[35] and international guidelines recommend their use in patients with AF at elevated thromboembolic risk.[1] Early detection of AF may permit the initiation of oral anticoagulation to reduce embolic stroke risk,[35] and early antiarrhythmic therapy to reduce the risk of death and stroke.[249] Accordingly early AF detection is a key cardiovascular priority in the UK NHS Long Term Plan,[98] and the European Society of Cardiology recommends opportunistic screening by pulse palpation or electrocardiogram (ECG) rhythm strip in persons aged ≥65 years and systematic ECG screening in those aged ≥75 years.[1]

Furthermore, AF frequently develops due to, and in parallel with, other cardiovascular, renal and metabolic conditions,[250] and individuals with AF are at an increased risk of major cardiovascular events in excess of stroke including ischemic heart disease, heart failure, chronic kidney disease, peripheral vascular disease and death.[32] Thus, AF screening, with or without AF diagnosis, may be a key opportunity for holistic management of cardiometabolic risk factors and unhealthy lifestyle behaviours to reduce an individual's risk of later adverse events beyond that of stroke prophylaxis alone.

Several studies have shown that serial or continuous non-invasive electrocardiogram (ECG) monitoring in older people with stroke risk factors / elevated N-terminal pro B-type natriuretic peptide (NT-proBNP), leads to a higher detection rate of previously undiagnosed AF compared with routine standard of care, though yields remain relatively low (3.0%-4.4%).[68, 87, 88, 251] The STROKESTOP randomised controlled trial, where AF screening was offered to individuals aged 75 and 76 years without exclusions, achieved only a 3% yield of new AF cases with a modest benefit in a composite outcome of ischaemic or haemorrhagic stroke, systemic embolism, bleeding leading to hospitalisation and all-cause death; and not for each of ischaemic stroke, haemorrhagic stroke, or hospitalisation for major bleeding.[70] Accordingly, for AF screening to be effective the yield of newly diagnosed AF amongst participants needs to be improved and the intervention offered may have to extend beyond only oral anticoagulation for stroke prophylaxis (Figure 1).

Figure 1 A schematic representation comparing current atrial fibrillation screening approaches, which focus on stroke prevention, with a broader approach to atrial fibrillation screening that considers that individuals eligible for screening will be at risk of multiple outcomes beyond stroke



Abbreviations: AF, atrial fibrillation; ECG, electrocardiogram

A large proportion of the population are registered in primary care with a routinely-collected electronic health record (EHR).[202, 252] A prediction model that utilises data available in the community to calculate AF risk could discriminate patients into risk categories, with screening offered only to higher risk individuals,[253] enabling scalable and efficient targeted AF screening. To date, several multivariable prediction models have been created or tested for prediction of incident AF in community-based electronic health records, but are of limited clinical utility for AF screening on account of moderate discriminative performance, long prediction horizons and limited scalability due to missing data.[197] None have yet reached widespread clinical practice. Moreover, reports of prediction models have yet to quantify the association between AF risk and new disease states outside that of AF and stroke.

3.5 Research aim

The aims of this study are to:

- Develop a model for predicting short-term AF risk from data routinely available in community-based EHRs.
- Quantify the association of predicted AF risk with a range of non-AF diseases.
- Externally validate the prediction model in an international context to assess transportability.
- Produce a calculator derived from a parsimonious prediction model.

3.6 Methods and analysis

3.6.1 Data sources and permissions

The derivation dataset will be the Clinical Practice Research Datalink-GOLD (CPRD-GOLD) dataset. This is an ongoing primary care database, established in 1987, that comprises anonymised medical records and prescribing data contributed by general practices using Vision® software. It contains data for approximately 17.5 million patients, with 30% of contributing practices in England, and represents the United Kingdom (UK) population in terms of age, sex and ethnicity.[202] In order to contribute to the database, general practices and other health centres must meet prespecified standards for research-quality data ('up-to-standard').[202, 254]

Recorded information includes patients' demography, clinical symptoms, signs, investigations, diagnoses, prescriptions, referrals, behavioural factors and test results entered by clinicians and other practice staff. All clinical information is coded using

Read Codes.[255] Extracted patients will have patient-level data linked to Hospital Episode Statistics (HES) Admitted Patient Care (APC) and Office for National Statistics (ONS) Death Registration. The CPRD dataset has been used to develop or validate a range of risk prediction models, including in cardiovascular disease.[256]

The extracted dataset, including linked data, comprises all patients for the period between 2nd January 1998 and 30th November 2018 from the snapshot of CPRD-GOLD in October 2019. Over this study period, the CPRD-GOLD dataset comprises approximately 2 million patients eligible for data linkage at an up-to-standard practice, with over 200,000 patients having a record of AF during follow-up.

To ascertain whether the prediction model is transportable to geographies outside of the UK, we will externally validate its performance in the Clalit Health Services database in Israel. As a result of the National Health Insurance Law, Israeli citizens are required to enrol in 1 of 4 payer-provider health funds and receive free basic health care. Clalit Health Services (CHS) provides health insurance coverage to 4.8 million insured members, and about two thirds of the population aged >65 years. CHS is recognised globally as the primary source of evaluation of Covid-19 vaccinations and therapies.[257-260] All clinical information is coded in International Classifications of Diseases, Ninth Revision (ICD-9). Receipt of vital status from the Ministry of the Interior ensures 100% follow-up of mortality. We will include participants insured by Clalit with continuous membership for at least 1 year before 01/01/2019: 2,159,663 patients with 4,330 of them having a new incident of AF (Atrial fibrillation and/or atrial flutter) in the first half of 2019.

3.6.2 Patient and public involvement

The Arrhythmia Alliance and AF association provided input on the FIND-AF scientific advisory board. The FIND-AF patient and public involvement group have given input to reporting and dissemination plans of the research.

3.6.3 Inclusion and exclusion criteria

The study population for derivation and internal validation will comprise all available patients in CPRD-GOLD eligible for data linkage and with at least 1-year follow-up in the period between 2nd January 1998 and 30th November 2018. For the external validation the study population will comprise participants insured by CHS, including those with

continuous membership for at least 1 year, before 01/01/2019 . Patients will be excluded if they were ≤30 years of age, or diagnosed with AF or atrial flutter (AFl) at the point of study entry, registered for less than 1 year or, in CPRD, ineligible for data linkage. Patients younger than 30 years of age are not included in the cohort for AF prediction because the incidence of AF over even a 10-year horizon is very low in this group.[11]

3.6.4 Prediction model outcome ascertainment

The outcome of interest is first diagnosed AF or AFl after baseline. We have included AFl as an outcome since it has similar clinical relevance, including thromboembolic risk and anticoagulation guidelines, as AF.[261] These will be identified using Read codes in CPRD dataset. For HES APC events and underlying cause of death variable in the ONS Death Registration data file, ICD-10 codes will be used. For CHS events will be identified using ICD-9 codes.

3.6.5 Sample size

To develop a prognostic prediction model, the required sample size may be determined by three criteria suggested by *Riley et al*.[262] For example, suppose a maximum of 200 parameters will be included in the prediction model and the Cox-Snell generalised R^2 is assumed to be 0.01. A total of 377,996 patients will be required to meet Riley's criterion (i) with global shrinkage factor of 0.95; this sample size also ensures a small absolute difference ($\Delta < 0.05$) in the apparent and adjusted Nagelkerke R^2 (Riley's criterion (ii)) and ensures precise estimate of overall risk with a margin of error < 0.001 (Riley's criterion (iii)). According to the Quality and Outcomes Framework (QOF), the prevalence of AF in England is 1.7%.[263, 264] Given an AF prevalence of 1.7%, only 6,425 patients will be expected to develop AF from 377,996 patients. Within the Clalit Health Services database there are 2,159,663 patients. Therefore, the number of patients in the CPRD and Clalit health services datasets with AF will provide sufficient statistical power to develop and validate a prediction model with the predefined precision and accuracy.

3.6.6 Predictor Variables

A systematic review has been conducted to establish predictor variables included in varying combinations by preceding prediction models developed to detect incident AF

in community-based EHRs (Table 1),[183] and supplemented with a literature search for variables associated with incident AF.

Table 1 Baseline demographic and comorbidity variables used in prediction models derived and/or validated for predicting incident atrial fibrillation in community-based electronic health records

Algorithm	**Demographics**	**Comorbidities**
$CHADS_2$[187]	Age	Hypertension, CHF, diabetes mellitus, CVA
CHA_2DS_2-VASc[188]	Age, sex	Hypertension, CHF, stroke/TIA/thromboembolism, vascular disease
CHARGE-AF[205]	Age, race, smoking status	Anti-hypertensive medication, MI, CHF, DM
C_2HEST[199]	Age	Hypertension, ischaemic heart disease, CHF, COPD, thyroid disease
HATCH[189]	Age	Hypertension, CHF, stroke/TIA, COPD
InGef[265]	Age, sex	Anti-hypertension medication, heart failure medication, chronic kidney disease, disorderd of lipoprotein metabolism and other lipidaemias, pulmonary heart diseases cardiac arrhythmias, other cerebrovascular disease, diverticular disease of intestine, dorsalgia, breathing abnormalities
MHS[234]	Age, sex	Anti-hypertensive medication, MI, CHF, peripheral vascular disease, inflammatory disease in a female, COPD
NHIRD[236]	Age (years), age group, sex	Hypertension, CHF, COPD, rheumatological disease, dyslipidaemia, DM, CVA or TIA, sleep disorder, cancer, hyperthyroidism, vascular disease, gout, CKD or ESRD, anaemia

NHIS-NSC[241]*	Age, sex, smoking (pack-year), alcohol	Hypertension, CHF, MI, vascular disease, stroke/TIA, COPD
PuLSE-AI[194]^	Age, sex, race, smoking status	Hypertension, anti-hypertensive medication, CHF, congenital heart disease, MI, LVH, type 1 DM, type 2 DM
Taiwan AF[266]	Age, sex, alcohol excess	Hypertension, CHF, IHD, ESRD

Abbreviations: AF, Atrial Fibrillation; $CHADS_2$, Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, prior Stroke or transient ischemic attack [2 points]; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points]; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); CHF, chronic heart failure; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; CPRD, Clinical Practice Research Datalink; CVA, cerebrovascular accident; DM, diabetes mellitus; ESRD, end-stage renal disease; HATCH, Hypertension, Age, stroke or Transient ischemic attack, Chronic obstructive pulmonary disease, Heart failure; IHD, ischaemic heart disease; LVH, left ventricular hypertrophy; MHS, Maccabi Healthcare Services; MI, myocardial infarction; NHIRD, National Health Insurance Research Database; NHIS-HEALS, National Health Insurance Service - Health screening Cohort; NHIS-NSC, National Health Insurance Service-based National Sample Cohort; TIA, transient ischaemic attack

N.B. * In Kim 2020 prediction model development using machine learning was completed both with and without the predictor $PM_{2.5}$ - which is fine particular matter air pollution. In this analysis we have only included the model without $PM_{2.5}$ as it is judged not to be a predictor that would be routinely available in primary care or population EHR. ^PuLSE-AI was also referred to as CPRD in Chapter 2.

Candidate variables include

- Sociodemographic variables including age, sex and ethnicity (SocioEconomic Score and population sector will serve as surrogate for ethnicity in CHS).
- All disease conditions during follow-up, including hospitalised diseases and procedures, such as other cardiovascular diseases, diabetes mellitus, chronic lung disease, renal disease, inflammatory disease, cancer, hypothyroidism and surgical procedures.
- Lifestyle factors including smoking status and alcohol consumption that are coded in structured Read codes.

Predictive factors will be identified using the appropriate codes, with Read codes for diagnoses and lifestyle factors. Code lists for predictors will be used from publications if available, otherwise the CPRD code browser will be used and codes checked by at least two clinicians. The code lists for predictors in CPRD-GOLD will be adapted from CALIBER and Health Data Research UK repositories or publications. If none are available from these sources then new code lists developed using the OpenCodelists and checked by at least two clinicians. Diagnostic code lists will comprise the primary care coding system (Read codes), to ensure that only information readily available within a primary care EHR could be incorporated within the prediction model. Within CHS, the code lists for predictors will be developed using similar methods based on the medical records and coding of CHS, which also includes a validated chronic diseases registry.

Candidate variable data types are deliberately limited to ensure widespread applicability of the model given the reality of 'missing' data in routinely-collected electronic health records.[252] Observations and laboratory results are not included. Ethnicity information is routinely collected in the UK NHS and so has increasingly high completeness,[267] and we will include an 'ethnicity unrecorded' category where it is unavailable because missingness is considered informative.[268] Ethicity in a UK context does not directly translate to an Israeli context so sociodemographic surrogates will be used: i) .population sectors- General Jewish, ultra-orthodox Jewish and Arab ii). Socioeconomic score on a scale of 1-10. For diagnoses, if medical codes are absent in a patient record we will assume that the patient does not have that diagnosis, or that the diagnosis was not considered sufficiently important to have been recorded by the general practitioner in case of symptoms.[35] Concordantly, the analytical cohorts are not expected to have missing data for any of the predictor variables.

3.6.7 Data analysis plan

3.6.7.1 Data pre-processing

The CPRD-GOLD and Clalit Health Services data will be cleaned and preprocessed for model development, internal validation and external validation. Specifically, for patient features with binary values, 0 and 1 will be mapped to the binary values. Variables with multiple categories (ethnicity) will be split into their component categories, and each given a binary value to indicate the presence or not of the variable for each patient. Continuous variables (age) will be kept as continuous.

3.6.7.2 Descriptive analysis

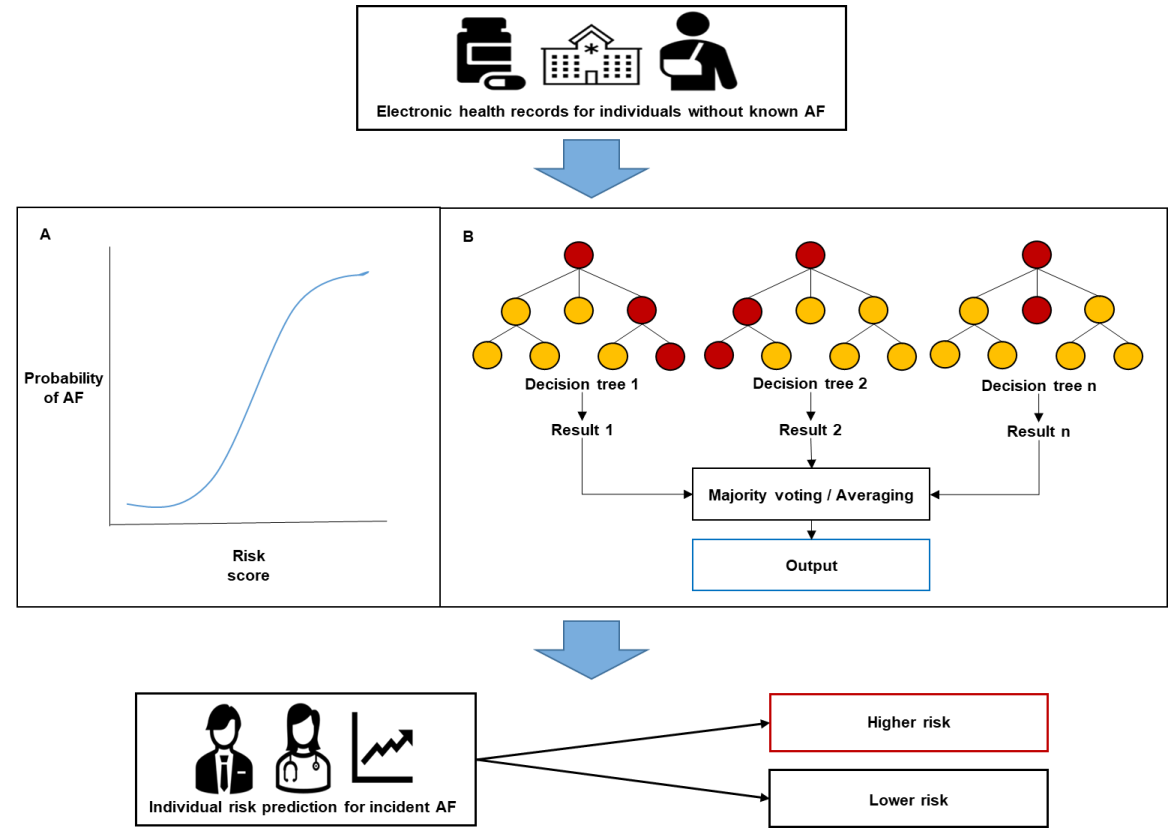
Continuous variables will be reported as mean ± standard deviation (SD) and categorical variables as frequencies with corresponding percentages.

3.6.7.3 Prediction model development

We will compare a machine learning and logistic regression approach to prediction model development for incident AF in CPRD-GOLD. Logistic regression model offers a more manageable approach for implementation, interpretation and training compared to machine learning algorithms, but machine learning methods can better handle non-linearities and interactions among variables and may lead to better discriminative performance.[197]

We will investigate the use of a Random Forest classifier for AF prediction in the CPRD-GOLD dataset. In our systematic review of AF prediction in EHRs it had the most evidence for use and showed robust performance in different datasets and geographies.[197] Random Forest (RF) is an ensemble technique that combines a large number of decision trees using a bagging approach to improve the overall performance (Figure 2).[269] In brief, the bagging approach grows multiple classification trees in parallel where each tree gives a classification which are called votes. These votes are then aggregated to provide a more accurate and stable prediction. Furthermore the degree of variation of each feature in a RF classifier for the prediction task can be calculated using the mean decrease in the Gini coefficient, a measure of how each variable contributes to the homogeneity of nodes and leaves in the resulting RF. Showing the importance of variables used in prediction (explainability) is considered important for clinical uptake of prediction models,[270] and a limitation of using deep learning techniques.

Figure 2 A schematic representation of a multivariable logistic regression model or random forest model using data from electronic health records to provide risk prediction for incident atrial fibrillation



Abbreviations: AF, atrial fibrillation

Preprocessed patient-level data in CPRD-GOLD will be randomly split into an 80:20 ratio to create derivation and internal validation (or training and testing) samples. The split ratio is not a significant factor, given the volume of the sample size. The model parameters and dropout rate, will be chosen through a grid search and 10-fold cross-validation will be used (i.e. 10% of the training data will be randomly selected as the cross-validation set). The multivariable logistic regression model will be developed with backward model selection with Akaike information criterion.[271] The prediction window will be set at 6 months, as this is considered in keeping with the logistical time frames for organising AF investigation at scale.[37]

3.6.7.4 Internal validation

We will evaluate the model performance using a validation cohort with internal bootstrap validation with 200 samples. The AUROC will be used to evaluate predictive ability (concordance index) with 95% confidence intervals calculated using the DeLong method.[38] Youden's index will be established for the outcome measure as a method of

empirically identifying the optimal dichotomous cut-off to assess sensitivity, specificity, positive predictive value and negative predictive value. We will calculate the Brier score, a measure of both discrimination and calibration, by taking the mean squared difference between predicted probabilities and the observed outcome. To assess the clinical impact of utilising FIND-AF as opposed to other risk prediction scores, we will calculate the net reclassification index at the risk threshold that equates to the average 6 months incidence rate in the cohort and conduct a decision curve analysis, which assesses across threshold probabilities whether the predictive model would do more benefit than harm. Calibration will be assessed graphically by plotting predicted AF risk against observed AF incidence and quantified using a calibration slope.

The same methods will be employed in subgroups by age (<65 years, ≥65 years, <75 years, ≥75 years), sex (women, men) and ethnicity (White, Black, Asian, others and unspecified) to assess the model's predictive performance across clinically relevant groups.

Performance of the prediction model will be compared with the CHA_2DS_2-VASc and C_2HEST scores. The CHA_2DS_2-VASc score was originally developed to predict stroke risk in individuals with AF, and the C_2HEST score for Asian people without structural heart disease.[197] These algorithms are robust to missing data in routinely-collected primary care EHRs and have been tested for AF risk prediction in European cohorts.[197] Other algorithms that can only be applied to a minority of European primary care EHRs (Pfizer-AI, CHARGE-AF) will not be considered as they cannot be implemented at scale to inform AF screening.[52, 252]

3.6.7.5 Quantification of the association between short-term predicted atrial fibrillation risk and long-term atrial fibrillation and other diseases

We will include all patients randomly assigned to the testing dataset in CPRD-GOLD by the Mersenne twister pseudorandom number generator, categorized as lower or higher predicted AF risk by the developed prediction model. For long-term AF risk we will plot Kaplan-Meier plots for individuals identified as higher and lower predicted risk of AF to assess the event rate for AF censored at 10 years, and calculate the hazard ratio for AF between higher and lower predicted risk of AF using the Cox proportional hazard model with adjustment for the competing risk of death. This will inform us of whether short-term AF risk is also associated with long-term AF risk, and whether an individual

who undergoes risk-guided AF screening should be considered for repeated AF screening at a later time point (e.g. 1 or 5-years).

For non-AF disease states we will consider the initial presentation of a cardiovascular, renal, or metabolic disease or death. This is because AF is not a disease in isolation and is known to be associated with high risk of adverse clinical outcomes. To best characterise highly prevalent and morbid diseases, associated with the development or consequence of AF and that may be appropriate for prevention or targeted diagnostic pathways subsequent to AF screening,[250] we will individually examine the following nine conditions: heart failure, valvular heart disease (and specifically aortic stenosis), myocardial infarction, stroke (ischaemic and haemorrhagic) or transient ischaemic attack, peripheral vascular disease, chronic kidney disease, diabetes mellitus, as well as chronic obstructive pulmonary disease (COPD). These disease states have been further selected for investigation because interventions could be implemented and / or tested to reduce their clinical progression. We will also quantify the occurrence of death by any cause recorded in primary care or by death certification from the UK Death Register of the Office for National Statistics, which will be mapped on to 9 disease categories (Table 2). For each condition, a list of diagnostic codes from the CALIBER code repository, including from International Classification of Diseases 10th revision (used in secondary care) and Read coding schemes (used in primary care) will be defined to comprehensively identify diagnoses from EHRs. Incident diagnoses will be defined as the first record of that condition in primary or secondary care records from any diagnostic position. For definition of new cases, we will exclude individuals for the analysis of each condition who had a diagnosis of that condition before the patient's entry to the study. If no indication of a specific disease is recorded, then the patient will be assumed to be free from the disease. CPRD is a positive recording dataset, which reduces the likelihood of the non-recording of a clinically identified disease state.

Table 2 Definition of disease categories for causes of deaths

Causes of death	**Code**
Cardiovascular disorders	ICD chapter 'Diseases of the circulatory system' (code range: I00–I99), excluding codes relating to infections or cerebrovascular disease.
Cerebrovascular disorders	ICD chapter 'Diseases of the circulatory system' (I60-I69)
Neoplasms	ICD chapter 'Neoplasms' (C00–D48).
Infections	Infectious and parasitic diseases, respiratory infections, urinary tract infections, and cellulitis, as defined by individual codes as Conrad et al.
Chronic respiratory diseases	Individual codes from Conrad et al.*
Digestive diseases	ICD chapter 'Diseases of the digestive system' (K00–K93), excepting selected codes categorized as infections.
Mental and neurological disorders	ICD chapter 'Mental and behavioral disorders' (F00–F99) and ICD chapter 'Diseases of the nervous system' (G00–G99)
Injuries	ICD chapters 'Injury, poisoning and certain other consequences of external causes' (S00–T98) and 'External causes of morbidity and mortality' (V01–Y98)
Kidney diseases	ICD sub-chapters 'Renal failure' (N17-N19), 'Glomerular diseases' (N00-N08), 'Renal tubulo-interstitial diseases' (N10-N16), 'Other disorders of kidney and ureter' (N25-N29)

Abbreviations: ICD, international classification of diseases

*N.B. To categorise cause of death as infections or chronic respiratory diseases we used the same codelists as previously published by *Conrad et al.*[272]

We will create Kaplan-Meier plots for individuals identified as higher and lower predicted risk of AF and derive the cumulative incidence rate for each outcome at 1, 5 and 10 years considering the competing risk of death, as well as death at 5 and 10 years. For each specified outcome, we will calculate the hazard ratio (HR) between higher and lower predicted risk of AF using the Fine and Gray's model with adjustment for the competing risk of death. We will also report adjusted HR where the model is adjusted for age, sex, ethnicity and the presence of any of the other outcomes at baseline. As some of the outcomes have incidence rates that are strongly associated with age (e.g. aortic stenosis) or differ by sex (e.g. heart failure),[273, 274] we will conduct sub-group analyses of incidence rates for higher and lower risk individuals for each outcome by age group (30 to 64 years and ≥65 years) and sex. As some of the non-AF outcomes are more likely to occur in the setting of prevalent AF (e.g. stroke or heart failure),[250] we will also conduct a sensitivity analysis whereby people with incident AF during follow-up are excluded.

3.6.7.6 External validation

The CHS dataset will then be used to externally validate the model performance to assess transportability. A lack of external validation hampers the implementation of prediction models in routine clinical practice.[275] The prediction model will be applied to each individual in the external validation cohort to give the predicted probabilities of experiencing AF at 6 months. Prediction performance will be quantified by calculating the AUROC, Brier score, and by using calibration plots, and the same aforementioned clinical utility and subgroup analysis will be conducted. Performance of the prediction model will be compared with the CHA_2DS_2-VASc, C_2HEST scores.

3.6.7.7 Prediction model calculator

The full models are developed to take advantage of rich longitudinal community-based EHRs present in many high income countries. However there are other geographies (low-lower middle income countries) and care setting (emergency care, secondary care clinics) where searching for AF may be desired and an easy-to-use, simple model is preferable. From the derived prediction model, we will generate a parsimonious model based on factors with clinical rationale to predict new-onset AF over a 6 months time horizon.[250] This will be based upon the same core principles as detail above, but use logistic regression to ensure transparency in how prediction results are calculated. We will aim to develop a user-friendly version of a model that may be applied as a calculator in a clinical and public setting, yet have good model performance indices.

3.6.7.8 Software

All analysis will be conducted through R.

3.7 Ethics and dissemination

The study has been approved by CPRD (ref no: 19_076). Those handling data have completed University of Leeds information security training. All analyses will be conducted in concordance with the CPRD study dataset agreement between the Secretary of State for Health and Social Care and the University of Leeds.

The Clalit Health Services (CHS) Community Helsinki Committee and the CHS Data Utilization Committee approved the study. The study was exempt from the requirement to obtain informed consent.

The study has been registered at clinical trials.gov (NCT05837364). The study is informed by the Prognosis Research Strategy (PROGRESS) framework and recommendations.[275] The subsequent research papers will be submitted for publication in a peer-reviewed journal and will be written following TRIPOD: *transparent reporting of a multivariable prediction model for individual prognosis or diagnosis* and RECORD: *reporting of studies conducted using observational routinely-collected health data* guidelines,[276, 277] as well as the CODE-EHR best-practice framework for using structured electronic healthcare records in clinical research.[278]

If the model shows better prediction performance than previous models and evidence for clinical utility in analysis, it could be made readily available through EHR platforms. The model will be designed to be amenable to in-situ updating with new information so that prediction of an individual's AF risk is updated contemporaneously. If the parsimonious model shows good prediction performance, the user friendly version could be accessible through the internet. Future research would be needed to assess the clinical impact of this risk model. At the point when utilisation in clinical practice is possible the applicable regulation on medicine devices will be adhered to.[279] When in clinical use, the model itself could also be reviewed and updated by a pre-specified expert consensus group on an annual basis after incorporating evidence from post-service utilization and the curation of more data.

3.8 Conclusions

Atrial fibrillation is a common clinical problem with important clinical sequelae that extend beyond stroke. A prediction model that may identify in a community-based EHR which individuals will develop AF could enable targeted screening. This British Heart Foundation funded study is designed to fill a knowledge gap and enable the leveraging of EHRs to provide risk prediction and targeted AF screening. By understanding if individuals identified as higher risk of new onset AF are also at elevated risk of other cardio-renal-metabolic diseases, this study may demonstrate the opportunity to deliver a more comprehensive clinical approach to improve patient outcomes from AF screening.

Chapter 4 Prediction of short-term atrial fibrillation risk using primary care electronic health records

Ramesh Nadarajah, Jianhua Wu, David C Hogg, Keerthenan Raveendra, Yoko M Nakao, Kazuhiro Nakao, Ronen Arbel, Moti Haim, Doron Zahger, John Parry, Chris Bates, Campbell Cowan, Chris P Gale

4.1 Summary of the publication

- This analysis was performed using the methodology presented in Chapter 3 to develop a prediction model for incident AF within the next 6 months in UK primary care EHRs using a random forests (RF) classifier.
- The study found that the RF classifier (FIND-AF) could be applied to all EHRs in the dataset, without hindrance by missing data.
- Prediction performance of FIND-AF was superior to a multivariable logistic regression model, and the C_2HEST and CHA_2DS_2-VASc scores.
- Prediction performance for FIND-AF was robust in both sexes and across ethnic groups, whereas the performance of the C_2HEST and CHA_2DS_2-VASc scores varied.

4.2 Publication status

- Published 9 February 2023
- Heart. 2023 Jul 1;109(14):1072-9

4.3 Abstract

4.3.1 Objective

Atrial fibrillation (AF) screening by age achieves a low yield and misses younger individuals. We aimed to develop an algorithm in nationwide routinely-collected primary care data to predict the risk of incident AF within 6 months (FIND-AF).

4.3.2 Methods

We used primary care electronic health record data from individuals aged ≥30 years without known AF in the UK Clinical Practice Research Datalink-GOLD dataset

between Jan 2, 1998 and Nov 30, 2018; randomly divided into training (80%) and testing (20%) datasets. We trained a random forest classifier using age, sex, ethnicity and comorbidities. Prediction performance was evaluated in the testing dataset with internal bootstrap validation with 200 samples, and compared against the CHA_2DS_2-VASc and C_2HEST scores, as these algorithms are robust to missing data in routinely-collected primary care EHRs and have been tested for AF risk prediction in European cohorts. Cox proportional hazard models with competing risk of death were fit for incident longer-term AF between higher and lower FIND-AF predicted risk.

4.3.3 **Results**

Of 2 081 139 individuals in the cohort, 7 386 developed AF within 6 months. FIND-AF could be applied to all records. In the testing dataset (n = 416 228), discrimination performance was strongest for FIND-AF (AUROC 0·824, 95% CI 0·813-0·829) compared with CHA_2DS_2-VASc (0·784, 0·773-0·794) and C_2HEST (0·757, 0·744-0·770), and robust by sex and ethnic group. The higher predicted risk cohort, compared to lower predicted risk, had a 20-fold higher 6-month incidence rate for AF and higher long-term hazard for AF (HR 8·75, 95% CI 8·44-9·06).

4.3.4 **Conclusions**

FIND-AF, a machine learning algorithm applicable at scale in routinely-collected primary care data, identifies people at higher risk of short-term AF.

4.4 Introduction

Atrial fibrillation (AF) is a major public health issue. There are now more new cases of AF diagnosed each year in the English National Health Service (NHS) than the four most common causes of cancer combined.[11] Moreover, it is estimated that up to 35% of disease burden remains undiagnosed,[67] and 15% of strokes occur in the context of undiagnosed AF.[248]

Early detection of AF may permit the initiation of oral anticoagulation to reduce embolic stroke risk,[35] and early antiarrhythmic therapy to reduce the risk of death and stroke.[249] Accordingly early AF detection is a key cardiovascular priority in the UK NHS Long Term Plan,[98] and the European Society of Cardiology recommends opportunistic screening by pulse palpation or electrocardiogram (ECG) rhythm strip in persons aged

≥65 years and systematic ECG screening in those aged ≥75 years.[1] However, there is an increasing cohort of individuals aged younger than 65 years who are being diagnosed with AF and are eligible for anticoagulation.[11]

A large proportion of the population are registered in primary care with a routinely-collected electronic health record (EHR).[202, 252] An algorithm that utilises routinely-collected EHR data to calculate AF risk could give a scalable, efficient and fair approach to targeting AF detection. However, previous algorithms tested in community-based EHRs have a number of shortcomings (Table 1-2). First, many algorithms developed using traditional regression techniques show only moderate discriminative performance.[197] Second, algorithm prediction horizons are often 5 or 10 years, making it difficult to judge the merits of investigating individuals in the short-term.[194, 252] Third, reports have infrequently investigated for variation in algorithm prediction performance by sex and ethnicity.[194] Fourth, algorithms often require variables frequently missing (absent in more than half of the records) from routinely-collected data such as height, weight and blood pressure thereby restricting the population to which they can be applied.[194, 196]

Table 1 Prediction models that have been derived and/or validated in community-based electronic health records for predicting atrial fibrillation

Algorithm	**Study Aim**	**Study**	**EHR cohort (country)**	**Age eligibility (years)**	**Discrimination (c-statistic)**	**Follow-up**	**Variable frequently missing in routinely-collected primary care EHR**
Models originally derived for another purpose but tested for prediction of incident atrial fibrillation							
$CHADS_2$	EV	Chao 2013[235]	NHIRD (TW)	≥18	0.713	10	N/A
	EV	Saliba 2016[238]	ClalitHS (IL)	≥50	0.728	3	
	EV	Li 2019[199]	YMID (CN)	≥18	0.632	11	
	EV	Li 2019[199]	NHIS-HEALS (KR)	≥18	0.637	11	
	EV	Kim 2020[241]	NHIS-NSC (KR)	≥18	0.652	5	
CHA_2DS_2-VASc	EV	Saliba 2016[238]	ClalitHS (IL)	≥50	0.744	3	N/A
	EV	Li 2019[199]	YMID (CN)	≥18	0.687	11	
	EV	Li 2019[199]	NHIS-HEALS (KR)	≥18	0.637	11	

	EV	Himmelreich 2020[196]	Nivel-PCD (NL)	≥40	0.669	5	
	EV	Kim 2020[241]	NHIS-NSC (KR)	≥18	0.654	5	
HATCH	EV	Suenari 2017[240]	NHIRD (TW)	≥20	0.716	9	N/A
	EV	Li 2019[199]	YMID (CN)	≥18	0.633	11	
	EV	Li 2019[199]	NHIS-HEALS (KR)	≥18	0.646	11	
	EV	Kim 2020[241]	NHIS-NSC (KR)	≥18	0.669	5	
	EV	Hu-WS 2020[237]	NHIRD (TW)	≥18	0.771	14	
Machine Learning models							
PuLSE-AI*	D	Hill 2019[194]	CPRD (UK)	≥30	0.827	11	Height, weight, BMI, SBP, DBP
	EV	Sekelj 2020[239]	Discover (UK)	≥30	0.870	8	
NHIRD	D	Hu-WS 2019[236]	NHIRD (TW)	≥18	0.948	14	Follow-up duration (years)

NHIS-NSC	D	Kim 2020[241]	NHIS-NSC (KR)	≥18	0.845	5	BMI, SBP, Triglycerides, total cholesterol, HDL cholesterol, LDL cholesterol, eGFR, GGT, fasting blood glucose, Haemoglobin, AST, Socioeconomic status
Regression Models derived in electronic health records							
C_2HEST	D	Li 2019[199]	YMID (CN)	≥18	0.750	11	N/A
	EV	Li 2019[199]	NHIS-HEALS (KR)	≥18	0.654	11	
	EV	Hu-WS 2020[237]	NHIRD (TW)	≥18	0.790	14	
	EV	Lip 2020[200]	DCRS, DNPR, DPR (DK)	65	0.588	5	
				70	0.594		
				75	0.593		
MHS	D	Aronson 2018[234]	MHS (IL)	≥50	0.743	10	BMI, SBP
Taiwan AF	D	Chao 2021[266]	NHIRD (TW)	≥40	0.857	1	N/A
					0.825	5	
					0.797	10	

					0.756	16	
InGef	D	Schnabel 2022[265]	InGef (G)	≥45	0.829	1	N/A
Regression model derived in a prospective cohort design							
CHARGE-AF	EV	Hill 2019[194]	CPRD (UK)	≥30	0.725	11	Height, weight, SBP, DBP

Abbreviations: AF, Atrial Fibrillation; $CHADS_2$, Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, prior Stroke or transient ischemic attack [2 points]; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); ClalitHS, Clalit Health Services; CPRD, Clinical Practice Research Datalink; D, derivation; DCRS, Danish Civil Registration system; DK, Denmark; DNPR, Danish National Patient Register; DPR, Danish Prescription Regster; EHR, electronic health record; EV, external validation; G, Germany; HATCH, Hypertension, Age, stroke or Transient ischemic attack, Chronic obstructive pulmonary disease, Heart failure; IL, Israel; KR, Republic of Korea; MHS, Maccabi Healthcare Services; NHIRD, National Health Insurance Research Database; NHIS-HEALS, National Health Insurance Service - Health screening Cohort; NHIS-NSC, National Health Insurance Service-based National Sample Cohort; Nivel-PCD, Netherlands Institute for Health Services Research Primary Care Database; NL, Netherlands; TW, Taiwan; UK, United Kingdom; YMID, Yunnan Medical Insurance Database

*N.B. *PuLSE-AI model was previously referred to as CPRD.

Table 2 Algorithms that have been derived and/or validated in European community-based electronic health records for predicting incident atrial fibrillation

Algorithm	**Study Aim**	**Study**	**EHR cohort (country)**	**Age eligibility (years)**	**Discrimination (c-statistic)**	**Follow-up**	**Variable frequently missing in routinely-collected primary care EHR**
Models originally derived for another purpose but tested for prediction of incident atrial fibrillation							
CHA_2DS_2-VASc	EV	Himmelreich 2020	Nivel-PCD (NL)	≥40	0.669	5	N/A
Machine Learning models							
PuLSE-AI*	D	Hill 2019	CPRD (UK)	≥30	0.827	11	Height, weight, BMI, SBP, DBP
	EV	Sekelj 2020	Discover (UK)	≥30	0.870	8	
Regression Models derived in electronic health records							
C_2HEST	EV	Lip 2020	DCRS, DNPR, DPR (DK)	65	0.588	5	N/A
				70	0.594		
				75	0.593		
InGef	D	Schnabel 2022	InGef (G)	≥45	0.829	1	N/A
Regression model derived in a prospective cohort design							

CHARGE-AF	EV	Hill 2019	CPRD (UK)	≥30	0.725	11	Height, weight, SBP, DBP

Abbreviations: AF, Atrial Fibrillation; $CHADS_2$, Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, prior Stroke or transient ischemic attack [2 points]; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); CPRD, Clinical Practice Research Datalink; D, derivation; DCRS, Danish Civil Registration system; DK, Denmark; DNPR, Danish National Patient Register; DPR, Danish Prescription Regster; EHR, electronic health record; EV, external validation; G, Germany; Nivel-PCD, Netherlands Institute for Health Services Research Primary Care Database; NL, Netherlands; UK, United Kingdom

N.B. *PuLSE-AI model was previously referred to as CPRD

Therefore, our objective was to train and test an algorithm (Future Innovations in Novel Detection of Atrial Fibrillation, FIND-AF) that predicts an individual's risk of AF in the next 6 months using routinely-recorded data in primary care EHRs. We compared performance against other AF prediction algorithms and investigated for variation in performance by sex and ethnicity.

4.5 Methods

4.5.1 Study design and population

In this population-based study we used primary care EHRs from the UK Clinical Practice Research Datalink (CPRD)-GOLD dataset. CPRD is one of the largest databases of longitudinal medical records from primary care worldwide and contains anonymised patient data from approximately 7% of the UK population.[202] CPRD-GOLD represents the UK population in terms of age, sex and ethnicity,[202] and has been used to develop algorithms for predicting AF.[194] Data collection happens as part of routine clinical care in participating practices and patients are included in the primary care dataset from their first until their last contact with a participating practice.[202] Diagnostic coding for AF in CPRD has been shown to be consistent and valid, with a positive predictive value of 98%.[280]

All individuals in the CPRD dataset were linked to Hospital Episode Statistics (HES) Admitted Patient Care (APC) records to obtain comprehensive coverage of AF cases diagnosed in secondary care. We included all adults registered at practices within CPRD who were ≥30 years of age at entry with no prior history of AF from either data source and at least one-year follow-up between January 2, 1998 and November 30, 2018. This study period enabled the inclusion of a sufficient sample size to have enough cases of AF within a 6 month prediction horizon to derive robust statistical results. Individuals were censored to a diagnosis of AF (or atrial flutter (AFl), since it has similar thromboembolic risk and anticoagulation guidelines),[1] withdrawal from CPRD, or six months, whichever came first. Diagnoses of AF or AFl in primary care were identified using Read codes in CPRD and in secondary care with the tenth revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes in HES-APC (Table 3). Individuals were randomly split 4:1 to establish a training dataset (80%) and a testing dataset (20%) using the Mersenne twister pseudorandom number generator.

We followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline and the CODE-EHR best-practice framework for using structured electronic healthcare records in clinical research.[276, 278]

Table 3 Read codes and ICD-10 codes used to define the outcomes of atrial fibrillation or atrial flutter

Code	Description
Read codes	
G573200	Paroxysmal atrial fibrillation
G573400	Permanent atrial fibrillation
G573500	Persistent atrial fibrillation
3272	ECG: atrial fibrillation
G573000	Atrial fibrillation
G573300	Non-rheumatic atrial fibrillation
G573.00	Atrial fibrillation and flutter
G573z00	Atrial fibrillation and flutter NOS
3273	ECG: atrial flutter
G573100	Atrial flutter
ICD-10 codes	
I48	Atrial fibrillation and flutter

Abbreviations: ECG, electrocardiogram; ICD-10, the tenth revision of the International Statistical Classification of Diseases and Related Health Problems

4.5.2 FIND-AF algorithm development

A random forest (RF) classifier was trained to predict AF at 6 months. Our systematic review evidenced strong discriminative performance for AF prediction using RF across different EHR datasets.[197] RF is a machine learning method consisting of many individual decision trees that operate as an ensemble.[269] FIND-AF was trained using 10-fold cross-validation on the full training set. Each decision tree used Gini impurity, commonly used in classification and regression tree (CART) algorithms, to measure the split quality.[281] The minimum impurity split threshold for each node, above which a

node will split into two or more branches, was set to 10^{-7}. The minimum number of samples required to split a node was set to two. The minimum samples per leaf was set to one. All the algorithm's hyperparameters were tuned using the grid search method, in which all possible combinations were evaluated, resulting in 1000 trees, mtry = 8 (the number of random features to consider in each tree) and nodesize = 12 (number of patients classified at that node).

To create an algorithm that could be implemented at scale in national primary care EHRs we restricted candidate variables to age, sex, comorbidities (72 binary variables, indicating presence or absence of recorded diagnosis) and ethnicity (6 categories). Observations and laboratory results were not included. Ethnicity information is routinely collected in the UK NHS and so has increasingly high completeness,[267] and we included an 'ethnicity unrecorded' category where it was unavailable because missingness was considered to be informative.[268] Predictor variables were selected a priori from systematic review of variables included in previous AF risk prediction algorithms,[197] plus an updated literature review. Predictor variables included in previous AF risk prediction algorithms derived and/or validated in community-based EHRs are summarised in Chapter 3 Table 1. Additional variables identified from a literature review are summarised in Table 4. Candidate variables were categorised (for example chronic kidney disease [CKD] into CKD stage 1-2, stage 3, stage 4, stage 5) based on how this affected the association of the comorbidity to the incidence of AF in the literature but ensuring the prevalence of a categorized variable was greater than 0.1% in the CPRD-GOLD dataset. The final list of predictor variables is summarised in Table 5.

Table 4 Candidate variables added after literature search with accompanying reference demonstrating association

Comorbidity associated with / predictive of atrial fibrillation	**Categorisation**	**Reference demonstrating association with AF and rationale for categorisation**
Cardiac surgery	Valvular,	Greenberg JW, Lancaster TS, Schuessler RB, et al. Postoperative atrial fibrillation following cardiac surgery: a persistent complication. Eur J Cardiothorac Surg 2017;52(4):665-72.
	Non-valvular (including coronary artery bypass grafting)	

		Within overall cardiac surgical procedures incidence of post-operative AF is 35%, isolated CABG has an incidence of 20—30% and isolated valve surgeries have an incidence of 35-40
Deep venous thrombosis	-	Lutsey P, Norby F, Alonso A, et al. Atrial fibrillation and venous thromboembolism: evidence of bidirectionality in the Atherosclerosis Risk in Communities Study. J Thromb Haemost 2018;16(4):670-79.
Infective Endocarditis	-	Ferrera C, Vilacosta I, Fernández C, et al. Usefulness of new-onset atrial fibrillation, as a strong predictor of heart failure and death in patients with native left-sided infective endocarditis. The American journal of cardiology 2016;117(3):427-33.
Electrophysiology procedure affecting the atria	-	Strickberger SA, Man KC, Daoud EG, et al. Adenosine-induced atrial arrhythmia: a prospective analysis. Ann Intern Med 1997;127(6):417-22. Khachab, H., and B. Brembilla-Perrot. "Prevalence of atrial fibrillation in patients with history of paroxysmal supraventricular tachycardia." International journal of cardiology 166.1 (2013): 221-224.
Hypertrophic cardiomyopathy	-	Siontis KC, Geske JB, Ong K, et al. Atrial fibrillation in hypertrophic cardiomyopathy: prevalence, clinical correlations, and mortality in a large high-risk population. Journal of the

		American Heart Association 2014;3(3):e001002.
Inflammatory bowel disease	-	Boos CJ. Infection and atrial fibrillation: inflammation begets AF. Eur Heart J 2020
Intensive care unit admission	-	Klein Klouwenberg PM, Frencken JF, Kuipers S, et al. Incidence, predictors, and outcomes of new-onset atrial fibrillation in critically ill patients with sepsis. A cohort study. Am J Respir Crit Care Med 2017;195(2):205-11.
Infection	Gastrointestinal	Gundlund A, Olesen JB, Butt JH, et al. One-year outcomes in atrial fibrillation presenting during infections: a nationwide registry-based study. Eur Heart J 2020;41(10):1112-19. Chang T-Y, Chao T-F, Liu C-J, et al. The association between influenza infection, vaccination, and atrial fibrillation: A nationwide case-control study. Heart Rhythm 2016;13(6):1189-94. Klein Klouwenberg PM, Frencken JF, Kuipers S, et al. Incidence, predictors, and outcomes of new-onset atrial fibrillation in critically ill patients with sepsis. A cohort study. Am J Respir Crit Care Med 2017;195(2):205-11. In a cohort study among infections precipitating AF the order of risk is as follows: Pneumonia > sepsis > urinary tract infection > gastrointestinal infection
	Influenza	
	Respiratory	
	Sepsis	

	Urinary	
Myocarditis	-	Wang Z, Wang Y, Lin H, et al. Early characteristics of fulminant myocarditis vs non-fulminant myocarditis: a meta-analysis. Medicine 2019;98(8)
Pulmonary embolus	-	Ptaszynska-Kopczynska K, Kiluk I, Sobkowicz B. Atrial fibrillation in patients with acute pulmonary embolism: clinical significance and impact on prognosis. BioMed research international 2019;2019
Pericarditis	-	Imazio M, Lazaros G, Picardi E, et al. Incidence and prognostic significance of new onset atrial fibrillation/flutter in acute pericarditis. Heart 2015;101(18):1463-67.
Pulmonary hypertension	-	Olsson KM, Nickel NP, Tongers J, et al. Atrial flutter and fibrillation in patients with pulmonary hypertension. Int J Cardiol 2013;167(5):2300-05.
Surgery (non-cardiac)	Colorectal	Siu CW, Tung HM, Chu KW, et al. Prevalence and predictors of new-onset atrial fibrillation after elective surgery for colorectal cancer. Pacing Clin Electrophysiol 2005;28:S120-S23. Onaitis M, D'Amico T, Zhao Y, et al. Risk factors for atrial fibrillation after lung cancer surgery: analysis of the Society of Thoracic Surgeons general thoracic surgery database. The Annals of thoracic surgery 2010;90(2):368-74. Philip I, Berroëta C, Leblanc I. Perioperative challenges of atrial
	Thoracic	
	Vascular	

		fibrillation. Current Opinion in Anesthesiology 2014;27(3):344-52. Thoracic surgery is associated with the greatest risk of post-operative AF amongst non-cardiac surgeries followed by colorectal then vascular surgery
Valvular heart disease	Mitral stenosis / rheumatic valvular disease	Iung B, Leenhardt A, Extramiana F. Management of atrial fibrillation in patients with rheumatic mitral stenosis. Heart 2018;104(13):1062-68. Levy S. Factors predisposing to the development of atrial fibrillation. Pacing Clin Electrophysiol 1997;20(10):2670-74. Grigioni F, Avierinos J-F, Ling LH, et al. Atrial fibrillation complicating the course of degenerative mitral regurgitation: determinants and long-term outcome. J Am Coll Cardiol 2002;40(1):84-92. The association of mitral stenosis and rheumatic valve disease with AF is greater than mitral regurgitation followed by diseases of other valves
	Non-mitral valve / other valves	
	Mitral regurgitation	
Vascular dementia	-	Ott A, Breteler MM, De Bruyne MC, et al. Atrial fibrillation and dementia in a population-based study: the Rotterdam Study. Stroke 1997;28(2):316-21.
Weight	Obese	Lavie CJ, Pandey A, Lau DH, et al. Obesity and atrial fibrillation prevalence, pathogenesis, and

		prognosis: effects of weight loss and exercise. J Am Coll Cardiol 2017;70(16):2022-35. Frost L, Hune LJ, Vestergaard P. Overweight and obesity as risk factors for atrial fibrillation or flutter: the Danish Diet, Cancer, and Health Study. The American journal of medicine 2005;118(5):489-95. Lee S-R, Choi E-K, Park CS, et al. Direct oral anticoagulants in patients with nonvalvular atrial fibrillation and low body weight. J Am Coll Cardiol 2019;73(8):919-31. Obesity is associated with a greater risk of AF than being overweight. Low body weight is associated with a higher risk of AF than normal weight.

Table 5 Variable categorisations with rationale

Comorbidity associated with / predictive of atrial fibrillation	**Categorisation**	**References and Rationale for categorisation**
Demographics		
Age	-	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 of Cardio-

		Thoracic Surgery (EACTS). Eur Heart J 2020 Incidence of AF increases with age (therefore included as a continuous variable)
Sex	Men	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 of Cardio-Thoracic Surgery (EACTS). Eur Heart J 2020 AF is more common in men
	Women	
Ethnicity	Asian	Shen AY-J, Contreras R, Sobnosky S, et al. Racial/ethnic differences in the prevalence of atrial fibrillation among older adults—a cross-sectional study. J Natl Med Assoc 2010;102(10):906-14. Chiang C-E, Zhang S, Tse HF, et al. Atrial fibrillation management in Asia: from the Asian expert forum on atrial fibrillation. Int J Cardiol 2013;164(1):21-32. White, Asian, pacific Asian, and black ethnicities have different odds ratios of development of AF
	Black	
	Mixed	
	Other	
	Pacific Asian	
	White	
Alcohol use	Ex-	Samokhvalov AV, Irving HM, Rehm J. Alcohol consumption as a risk factor for atrial fibrillation: a systematic review and meta-analysis. European Journal of
	Light,	
	Moderate	
	Excess	

	Unspecified	Preventive Cardiology 2010;17(6):706-12. There is a monotonic dose-response relationship between alcohol consumption and AF incidence
Smoking	Current	Heeringa J, Kors JA, Hofman A, et al. Cigarette smoking and risk of atrial fibrillation: the Rotterdam Study. Am Heart J 2008;156(6):1163-69. Watanabe I. Smoking and risk of atrial fibrillation: Elsevier, 2018. Current and ex-smokers are at increased risk of AF, with a higher risk in current smokers.
	Ex	
Weight	Obese	See Table 4
	Overweight	
	Under-weight	
Comorbidities		
Adult congenital heart disease	-	See Chapter 3 Table 1
Anaemia	-	See Chapter 3 Table 1
Cancer	Leukaemia	Thompson PA, Lévy V, Tam CS, et al. Atrial fibrillation in CLL patients treated with ibrutinib. An international retrospective study. Br J Haematol 2016;175(3):462-66. Sorigue M, Gual-Capllonch F, Garcia O, et al. Incidence, predictive factors, management, and survival impact of atrial fibrillation in non-Hodgkin
	Lymphoma	
	Metastasis	
	Skin cancers other than melanoma	
	Solid organ	

		lymphoma. Ann Hematol 2018;97(9):1633-40. Han H, Chen L, Lin Z, et al. Prevalence, trends, and outcomes of atrial fibrillation in hospitalized patients with metastatic cancer: findings from a national sample. Cancer medicine 2021;10(16):5661-70. AF risk is higher in patients with leukaemia and lymphoma, especially treated with iritunib. Solid organ cancers (such as lung and colorectal cancer) are more likely to undergo surgery. Metastatic disease is associated with higher risk of AF compared to non-metastatic disease. Skin cancers other than melanoma have a lower risk of metastasis and hence AF.
Cardiac surgery	Valvular,	See Table 4
	Non-valvular (including coronary artery bypass grafting)	
Chronic kidney disease	Stage 1-2	Alonso A, Lopez FL, Matsushita K, et al. Chronic kidney disease is associated with the incidence of atrial fibrillation: the Atherosclerosis Risk in Communities (ARIC) study. Circulation 2011;123(25):2946-53. Risk of AF increases as CKD stage worsens and if there is proteinuria
	Stage 3	
	Stage 4	
	Stage 5	
	Unspecified	
	Other	
COPD	-	See Chapter 3 Table 1
Cerebro-vascular accident	Intracerebral haemorrhage	Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the

	Subarachnoid haemorrhage	diagnosis and management of atrial fibrillation developed in collaboration with the European Association of Cardio-Thoracic Surgery (EACTS). Eur Heart J 2020 Association with AF is higher for ischaemic strokes than haemorrhagic strokes
	Unspecified	
Diabetes Mellitus	Good control	Dublin S, Glazer NL, Smith NL, et al. Diabetes mellitus, glycemic control, and risk of atrial fibrillation. J Gen Intern Med 2010;25(8):853-58. Poorer glycaemic control is associated with a higher risk of AF compared to better glycaemic control or no diabetes
	Poor control	
	Unspecified / secondary	
Deep venous thrombosis	-	See Table 4
Dyslipidaemia	-	See Chapter 3 Table 1
Infective Endocarditis	-	See Table 4
Electrophysiology procedure affecting the atria	-	See Table 4
Gout	-	See Chapter 3 Table 1
Hypertrophic cardiomyopathy	-	See Table 4
Heart failure	-	See Chapter 3 Table 1
Hypertension	Poor control	Dzeshka MS, Shantsila A, Shantsila E, et al. Atrial fibrillation and hypertension. Hypertension 2017;70(5):854-61.
	Unspecified / secondary	

		Poorer control of hypertension and end organ damage is associated with a higher risk of developing AF
Hyperthyroidism	-	See Chapter 3 Table 1
Inflammatory bowel disease	-	See Table 4
Intensive care unit admission	-	See Table 4
Ischaemic heart disease	Chronic	Huxley RR, Lopez FL, Folsom AR, et al. Absolute and attributable risks of atrial fibrillation in relation to optimal and borderline risk factors: the Atherosclerosis Risk in Communities (ARIC) study. Circulation 2011;123(14):1501-08. Pizzetti F, Turazza F, Franzosi M, et al. Incidence and prognostic significance of atrial fibrillation in acute myocardial infarction: the GISSI-3 data. Heart 2001;86(5):527-32. There is a high risk of AF in the acute setting of myocardial infarction as well as evidence in the context of underlying chronic coronary syndromes.
	Myocardial infarction	
	Percutaneous coronary intervention	
Infection	Gastrointestinal	See Table 4
	Influenza	
	Respiratory	
	Sepsis	
	Urinary	
Left ventricular hypertrophy	-	See Chapter 3 Table 1
Myocarditis	-	See Chapter 3 Table 1

Obstructive sleep apnoea	-	See Chapter 3 Table 1
Pulmonary embolus	-	See Table 4
Pericarditis	-	See Table 4
Pulmonary hypertension	-	See Table 4
Peripheral vascular disease	-	See Chapter 3 Table 1
Rheumatological condition	Autoimmune connective tissue diseases	Lee E, Choi E-K, Jung J-H, et al. Increased risk of atrial fibrillation in patients with Behçet's disease: a nationwide population-based study. Int J Cardiol 2019;292:106-11. Moon I, Choi E-K, Jung J-H, et al. Ankylosing spondylitis: a novel risk factor for atrial fibrillation—a nationwide population-based study. Int J Cardiol 2019;275:77-82. Melduni RM, Cooper LT, Gersh BJ, et al. Association of Autoimmune Vasculitis and Incident Atrial Fibrillation: A Population-Based Case-Control Study. Journal of the American Heart Association 2020;9(18):e015977. Naaraayan A, Meredith A, Nimkar A, et al. Arrhythmia prevalence among patients with Polymyositis-Dermatomyositis in the United States: an observational study. Heart Rhythm 2021
	Rheumatoid arthritis	
	Spondyloarthropathies	
	Vasculitides	

		Songnan W, Shengma C. GW24-e2483 Catheter ablation of atrial fibrillation in patients with autoimmune rheumatic diseases. Heart 2013;99(Suppl 3):A197-A97. Giallafos I, Triposkiadis F, Oikonomou E, et al. Incident atrial fibrillation in systemic sclerosis: the predictive role of B-type natriuretic peptide. Hellenic J Cardiol 2014;55:313-21. Pugnet G, Gouya H, Puéchal X, et al. Cardiac involvement in granulomatosis with polyangiitis: a magnetic resonance imaging study of 31 consecutive patients. Rheumatology 2017;56(6):947-56. Lindhardsen J, Ahlehoff O, Gislason GH, et al. Risk of atrial fibrillation and stroke in rheumatoid arthritis: Danish nationwide cohort study. BMJ 2012;344 Each of the subtypes of rheumatological disease are associated with differing risks of development of AF. Here they have been categorised in clinical sub-type.
Smoking	Current	See Table 4
	Ex	
Surgery (non-cardiac)	Colorectal	See Table 4
	Thoracic	
	Vascular	

Systemic Embolism	-	See Chapter 3 Table 1
Valvular heart disease	Mitral stenosis / rheumatic valvular disease	See Table 6
	Non-mitral valve / other valves	
	Mitral regurgitation	
Vascular dementia	-	See Table 4

Diagnostic code lists only included the primary care coding system (Read codes), ensuring that only information readily available within a primary care EHR could be incorporated within the algorithm. Concordantly, our entire analytical cohort had no missing data for any of the predictor variables and the algorithm could be applied to all records.

4.5.3 Statistical analyses

The baseline characteristics are summarised by incident AF status. Continuous variables were reported as mean ± standard deviation (SD). Categorical variables were reported as frequencies with corresponding percentages.

The degree of variation of each feature in FIND-AF to classification was calculated using the mean decrease in the Gini coefficient, a measure of how each variable contributes to the homogeneity of nodes and leaves in the resulting random forest.

Model performance of FIND-AF was determined using the full holdout test set with internal bootstrap validation with 200 samples and compared to a multivariable logistic regression (MLR) model developed with backward model selection with Akaike information criterion.[271] Performance was compared with the CHA_2DS_2-VASc and C_2HEST scores. The CHA_2DS_2-VASc score was originally developed to predict stroke risk in individuals with AF, and the C_2HEST score for Asian people without structural heart disease.[197] These algorithms are robust to missing data in routinely-collected

primary care EHRs and have been tested for AF risk prediction in European cohorts (Table 2).[197] Other algorithms that can only be applied to a minority of European primary care EHRs (PuLSE-AI, CHARGE-AF) were not considered.[52, 252] The area under the receiver operating characteristic curve (AUROC) was used to evaluate predictive ability (concordance index) with 95% confidence intervals (CIs) calculated using the DeLong method. Youden's index was established for the outcome measure as a method of empirically identifying the optimal dichotomous cut-off to assess sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Youden's index was calculated and optimised for each test set for each score to derive the optimal cut-off threshold. Calibration was assessed by plotting predicted AF risk against observed AF incidence and by the calibration slope. We calculated the Brier score, a measure of both discrimination and calibration, by taking the mean squared difference between predicted probabilities and the observed outcome. To assess the clinical impact of utilising FIND-AF as opposed to other risk prediction scores, we calculated the net reclassification index at 0.4% AF risk threshold (the average 6-month incidence rate in the cohort) and conducted a decision curve analysis.

We investigated the performance of FIND-AF, CHA_2DS_2-VASc and C_2HEST within relevant subgroups defined by sex, ethnicity (White vs. Black vs. Asian vs. other Non-White ethnic minorities) and age (≥65 years and ≥75 years). We plotted Kaplan-Meier plots for individuals identified as higher and lower FIND-AF predicted risk of AF to assess the event rate for AF censored at 10 years, and calculated the hazard ratio for AF between higher and lower FIND-AF predicted risk of AF using the Cox proportional hazard model with adjustment for the competing risk of death. We used R version 4·1·0 for all analyses.

4.5.4 Patient and public involvement

The Arrhythmia Alliance an AF association provided input on the FIND-AF scientific advisory board. The FIND-AF patient and public involvement group have given input to reporting and dissemination plans of the research.

4.6 Results

4.6.1 Patient population

There were 2 081 139 individuals registered in our UK primary care cohort (1 664 911 in the training dataset, 416 228 in testing dataset), with average age 49.9 (SD 15.4), 50.7% women, and 86.7% white. Baseline characteristics and clinical outcomes were similar in the training and testing datasets (Table 6).

Table 6 Baseline characteristics of training and testing datasets

	Training set **n (%)**	**Testing set** **n (%)**
	1 664 911	416 228
Demographics		
Age, years	49.90 (15.43)	49.90 (15.42)
Sex (women)	844 083 (50.7)	211 478 (50.8)
Comorbidities		
Diabetes mellitus	58 513 (3.5)	14 268 (3.4)
Stroke or TIA	30 871 (1.9)	7 794 (1.9)
Ischaemic heart disease	62 980 (3.8)	15 622 (3.8)
Hypertension	200 217 (12.0)	50 106 (12.0)
Heart failure	11 577 (0.7)	2 790 (0.7)
Dyslipidaemia	48 719 (2.9)	12 170 (2.9)
Hyperthyroidism	13 069 (0.8)	3 233 (0.8)
COPD	20 294 (1.2)	5 129 (1.2)
Chronic kidney disease	23 794 (1.4)	6 014 (1.4)
Anaemia	53 962 (3.2)	13 383 (3.2)
Cancer	58 725 (3.5)	14 783 (3.6)
Valvular heart disease	7 946 (0.5)	1 927 (0.5)
Mean CHA_2DS_2-VASc score	0.98 (1.04)	0.98 (1.04)

Abbreviations: AF, atrial fibrillation; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 years [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74 years, Sex Category; COPD, chronic obstructive pulmonary disease; SD, standard deviation; TIA, transient ischaemic attack

Within 6 months, 7 386 individuals (0.4%) were recorded as having AF. Those who developed AF were older and had a higher prevalence of baseline comorbidities than

individuals who did not develop AF (Table 7). Of new cases, 1 546 (20.9%) were younger than 65 years old.

Table 7 Baseline characteristics of analytical cohort with and without atrial fibrillation

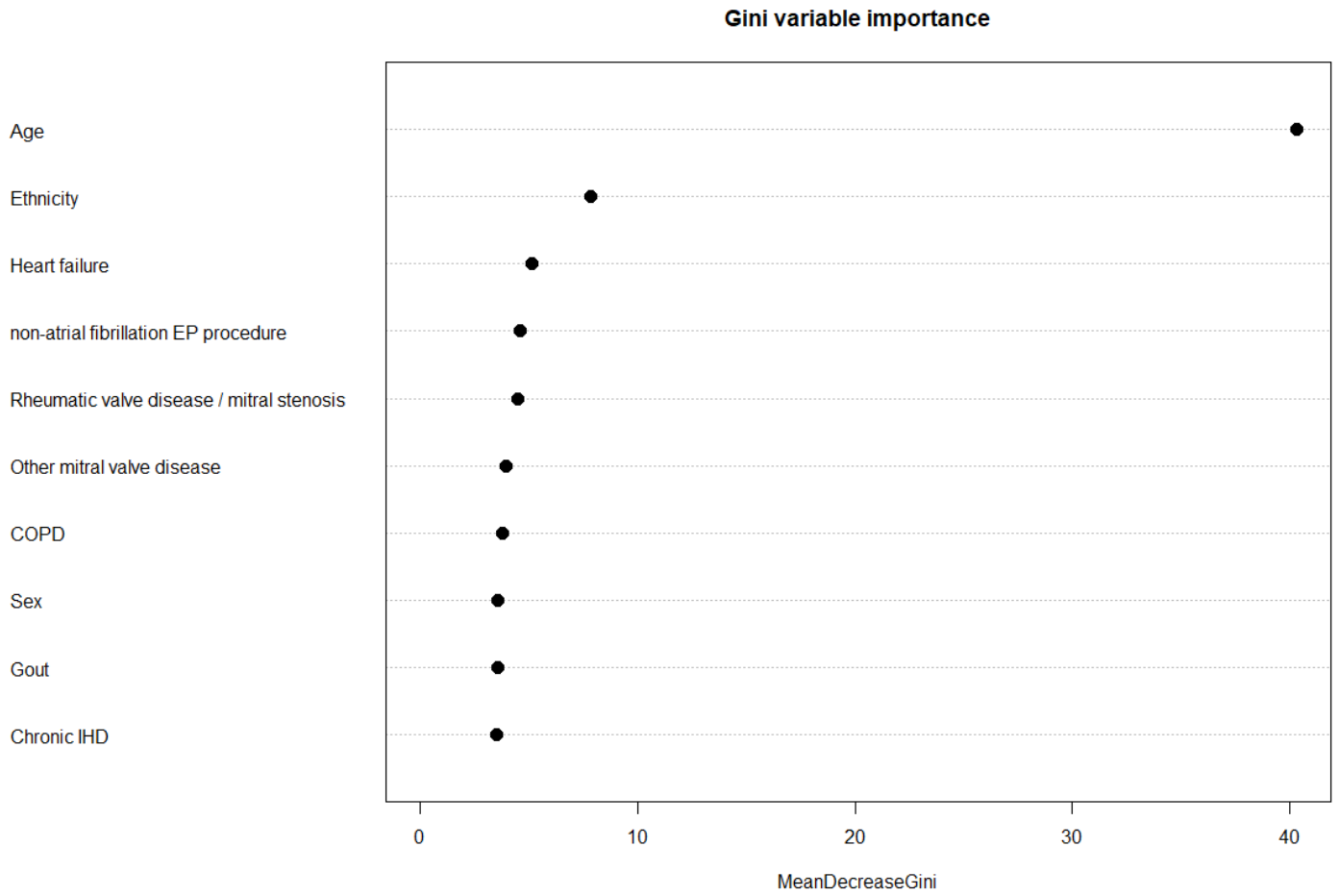
	Incident AF	
	No AF **n (%)**	**AF** **n (%)**
	2 073 753	7 386
Demographics		
Age, years	49·82 (15.37)	73·72 (12.62)
Sex (women)	1 051 942 (50.7)	3 619 (49.0)
Comorbidities		
Diabetes mellitus	71 966 (3.5)	815 (11.0)
Stroke or TIA	37 773 (1.8)	892 (12.1)
Ischaemic heart disease	77 060 (3.7)	1 542 (20.9)
Hypertension	247 436 (11.9)	2 887 (39.1)
Heart failure	13 717 (0.7)	650 (8.8)
Dyslipidaemia	60 357 (2.9)	532 (7.2)
Hyperthyroidism	16 147 (0.8)	155 (2.1)
COPD	24 962 (1.2)	461 (6.2)
Chronic kidney disease	29 359 (1.4)	449 (6.1)
Anaemia	66 844 (3.2)	501 (6.8)
Cancer	72 621 (3.5)	887 (12.0)
Valvular heart disease	9 497 (0.5)	376 (5.1)
Mean CHA_2DS_2-VASc score (SD)	0·97 (1.03)	2·72 (1.42)

Abbreviations: AF, atrial fibrillation; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 years [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74 years, Sex Category; COPD, chronic obstructive pulmonary disease; SD, standard deviation; TIA, transient ischaemic attack

4.6.2 Prediction factors and model accuracy

According to mean decrease in the Gini coefficient, age contributed the most to the prediction, followed by ethnicity and history of heart failure (Figure 1).

Figure 1 The top 10 most important variables for FIND-AF prediction in individuals aged ≥30 years quantified by mean decrease in Gini coefficient



COPD, chronic obstructive pulmonary disease; EP, electrophysiology; IHD, ischaemic heart disease

AF discrimination and accuracy of predictions, by AUROC and Brier scores, were better using FIND-AF than the MLR, CHA_2DS_2-VASc and C_2HEST algorithms (Table 8, Figure 2). Sensitivity was highest for the CHA_2DS_2-VASc algorithm, but specificity lowest.

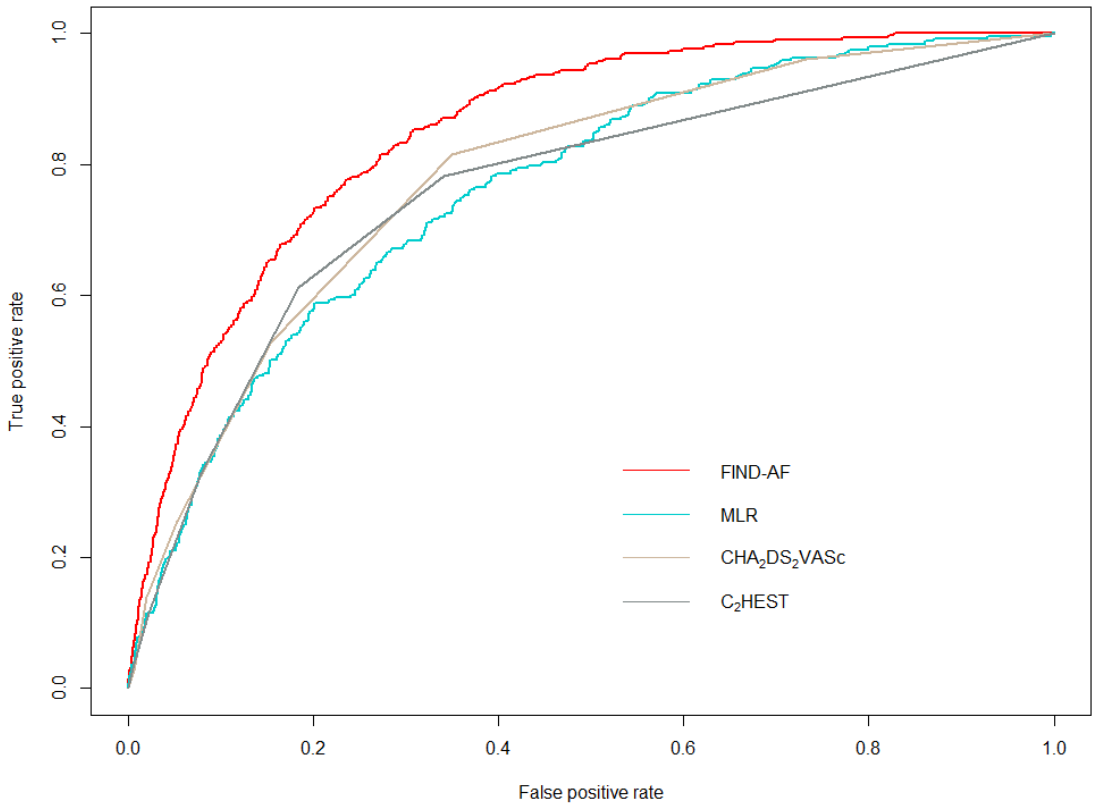
Table 8 Performance for 6-month incident atrial fibrillation with optimal threshold determined by Youden Index

	FIND-AF	**MLR**	**CHA_2DS_2-VASc**	**C_2HEST**
AUROC (95% CI)	0.824 (0.814-0.834)	0.765 (0.755-0.769)	0.784 (0.773-0.794)	0.757 (0.744-0.770)
Sensitivity (95% CI)	0.781 (0.731-0.829)	0.760 (0.653-0.814)	0.847 (0.829-0.866)	0.642 (0.619-0.791)
Specificity (95% CI)	0.731 (0.693-0.771)	0.679 (0.635-0.776)	0.611 (0.608-0.612)	0.790 (0.622-0.792)
PPV (% [95% CI])	2.5% (2.3-2.7)	2.0% (1.8-2.6)	2.2% (2.1-2.3)	2.0% (1.5-2.2)
NPV (% [95% CI])	99.8% (99.8-99.8)	99.7% (99.6-99.7)	99.8% (99.8-99.8)	99.7% (99.7-99.8)
Calibration slope* (95% CI)	0.782 (0.743-0.824)	0.698 (0.654-0.735)	0.621 (0.589-0.652)	0.608 (0.576-0.648)
Brier score	0.069	0.097	0.093	0.102

Abbreviations: AF, atrial fibrillation; AUROC, area under received operating characteristic; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); CI, confidence interval; FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation; MLR, Multivariable logistic regression; NPV, negative predictive value; PPV, positive predictive value

N.B. *calibration slope was derived from linear regression models by forcing the intercept through origin (0,0).

Figure 2 Receiver operating characteristic curves for FIND-AF, multivariable logistic regression, CHA_2DS_2-VASc, and C_2HEST algorithms

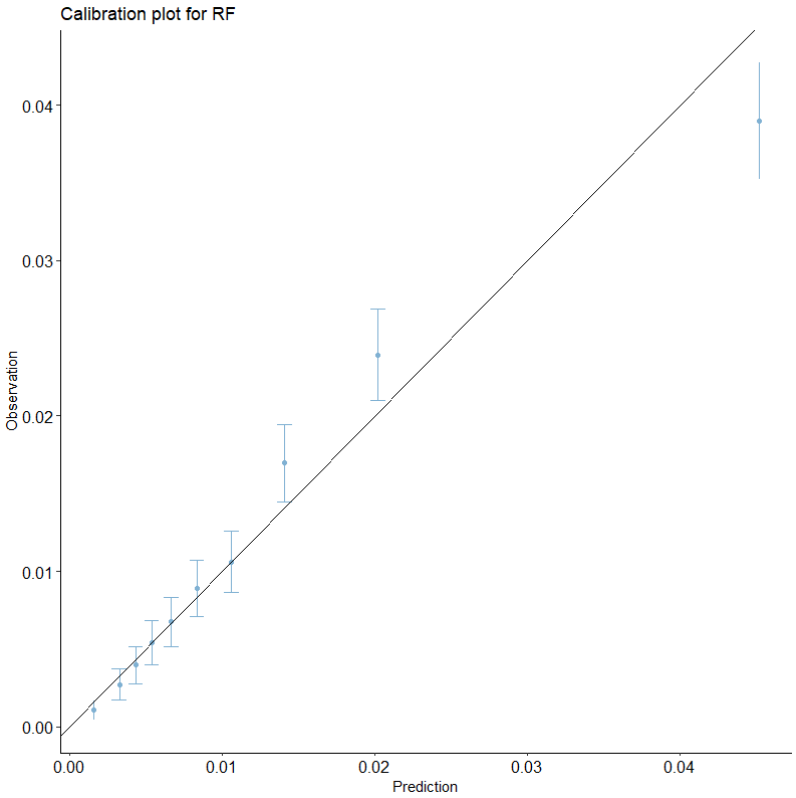


Abbreviations: CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age > 75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex category; CI, confidence interval; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥ 75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation; MLR, multivariable logistic regression; Multivariable logistic regression.

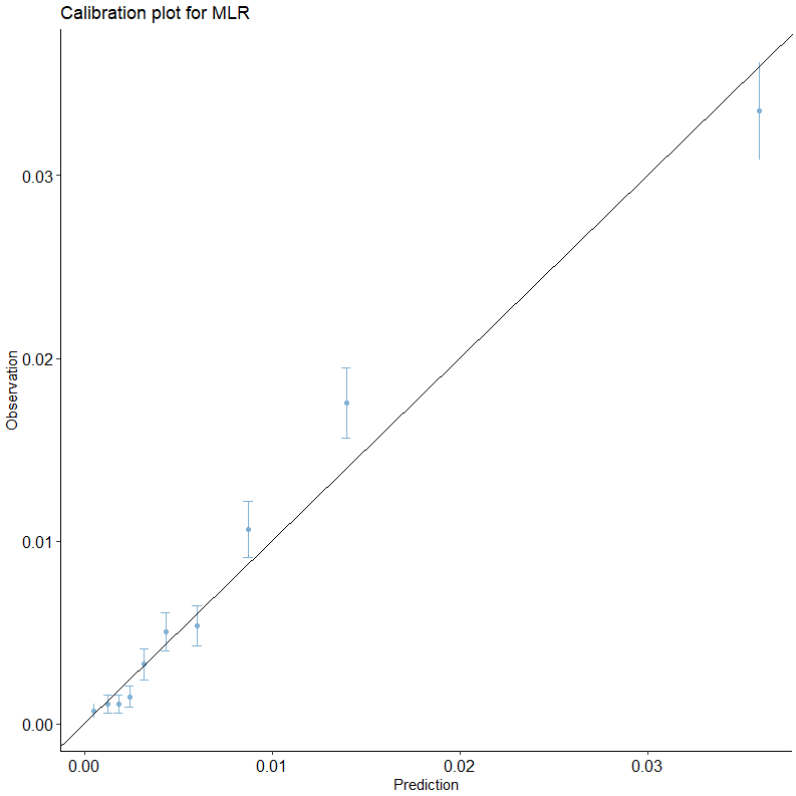
According to the Youden index, the optimal cut-off was 0.0032, leading to a sensitivity of 78% and a specificity of 73%, with a PPV of 2.5% and NPV of 99.8%. The low incidence of AF over 6 months led to similar values for PPV and NPV across the algorithms. Of the algorithms, FIND-AF was the best calibrated (calibration slope 0.782 [95% CI 0.743 – 0.824], Table 2, Figure 3), yet showed underestimation of risk in the mid-risk strata and over-estimation in the highest risk strata.

Figure 3 Calibration plots for FIND-AF, multivariable logistic regression, CHA_2DS_2-VASc, and C_2HEST algorithms

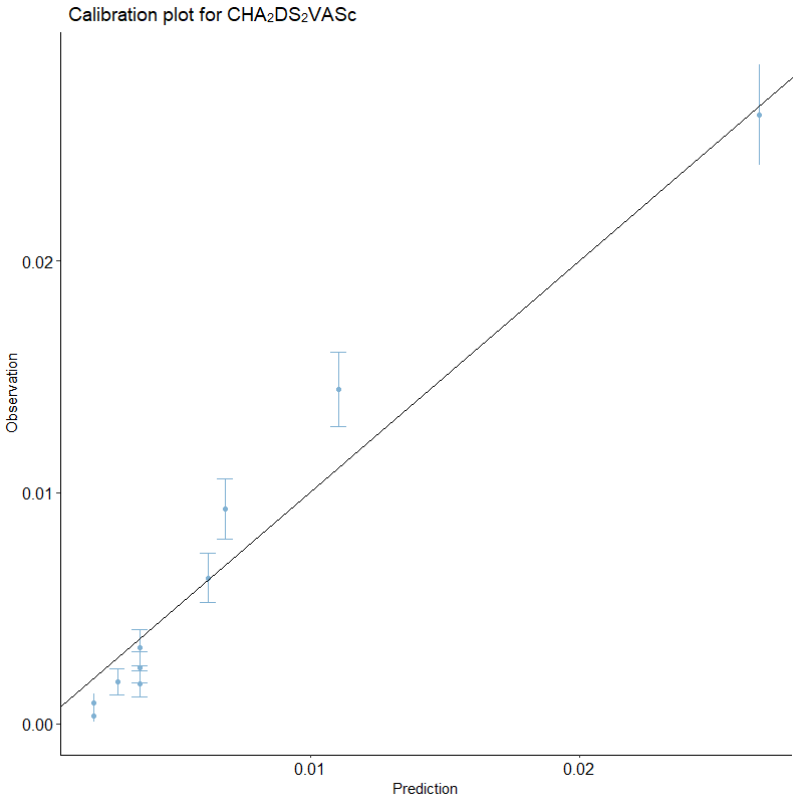
FIND-AF



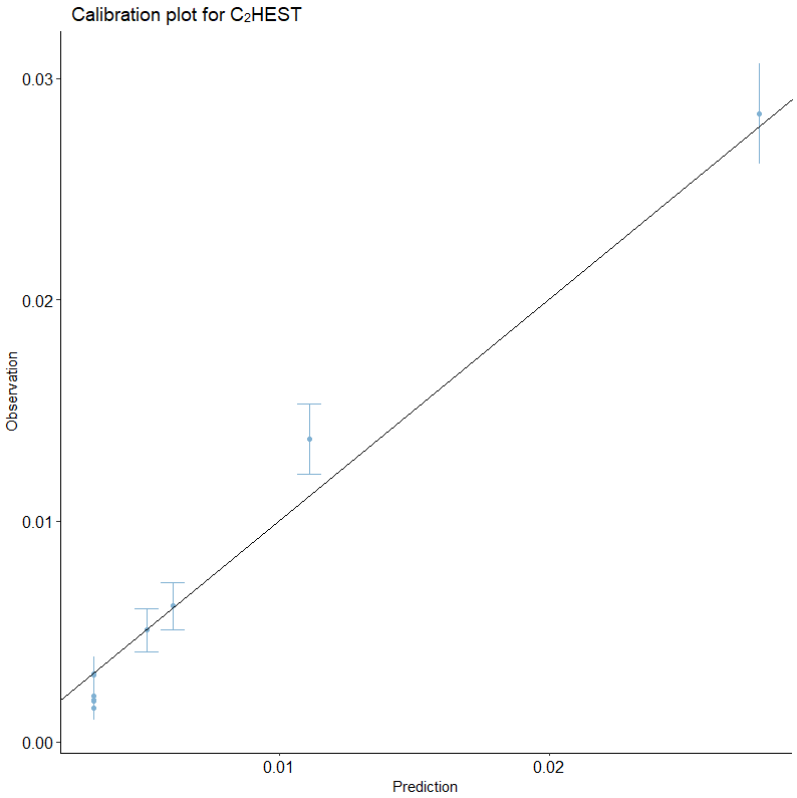
Multivariable logistic regression



CHA_2DS_2VASc



C_2HEST



Abbreviations: CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation

4.6.3 Risk classification

Of the 416 228 individuals in the testing set, 82 942 (19.9%) were classified as higher risk using FIND-AF, 84 282 (20.2%) using the CHA_2DS_2-VASc score and 84 542 (20.3%) using the C_2HEST score, respectively. Net reclassification analyses at the 0.4% risk threshold demonstrated modestly favourable reclassification using FIND-AF as opposed to using CHA_2DS_2-VASc (net reclassification 0.032, 95% CI 0.029-0.051) and strong favourable reclassification using FIND-AF as opposed to using C_2HEST (net reclassification 0.113, 95% CI 0.098-0.135; Table 9).

Table 9 Net reclassification using FIND-AF

AF cases

	FIND-AF			FIND-AF	
CHA_2DS_2-VASc	≥0.4%	<0.4%	**C_2HEST**	≥0.4%	<0.4%
≥0.4%	1 121	37	≥0.4%	893	10
<0.4%	82	191	<0.4%	310	218

Appropriate upclassification

Inappropriate downclassification

Non-AF cases

	FIND-AF			FIND-AF	
CHA_2DS_2-VASc	≥0.4%	<0.4%	**C_2HEST**	≥0.4%	<0.4%
≥0.4%	65 322	17 511	≥0.4%	38 640	3 053
<0.4%	16 417	315 547	<0.4%	43 099	330 005

Appropriate downclassification

Inappropriate upclassification

Net reclassification indices

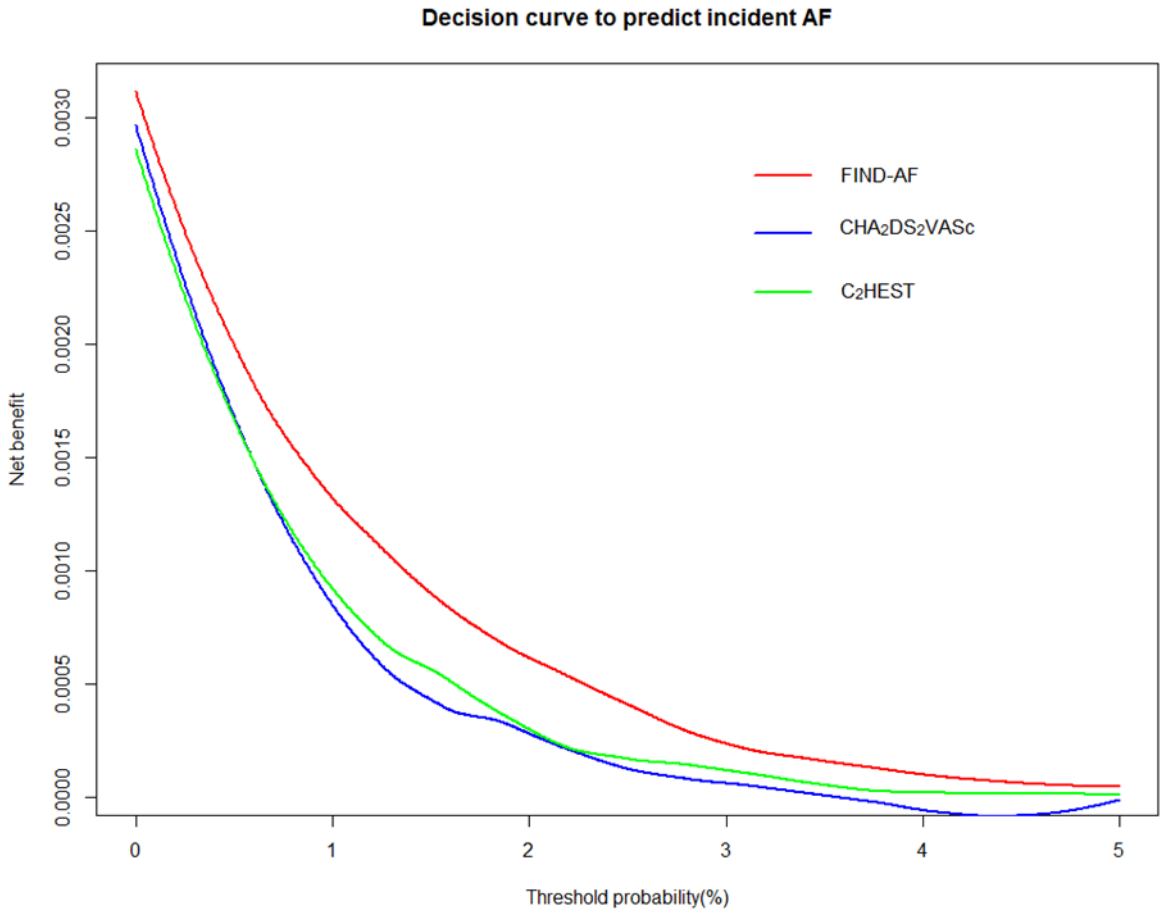
Index	**CHA_2DS_2-VASc**	**C_2HEST**
Case reclassification (NRI+ [95% CI])	0.031 (0.026-0.048)	0.021 (0.19-0.23)
Non-case reclassification (NRI- [95% CI])	0.0026 (0.0015-0.0032)	-0.096 (-0.098 - -0.095)
Net reclassification (NRI [95% CI])	0.032 (0.029-0.051)	0.113 (0.098-0.135)

Abbreviations: CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic

heart failure, Thyroid disease (hyperthyroidism); CI, confidence interval; FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation; NRI, net reclassification index

In a decision curve analysis, FIND-AF had a superior net benefit compared to the CHA_2DS_2-VASc and C_2HEST risk scores across all threshold probabilities (Figure 4).

Figure 4 Decision curve analysis for FIND-AF versus CHA_2DS_2-VASc and C_2HEST

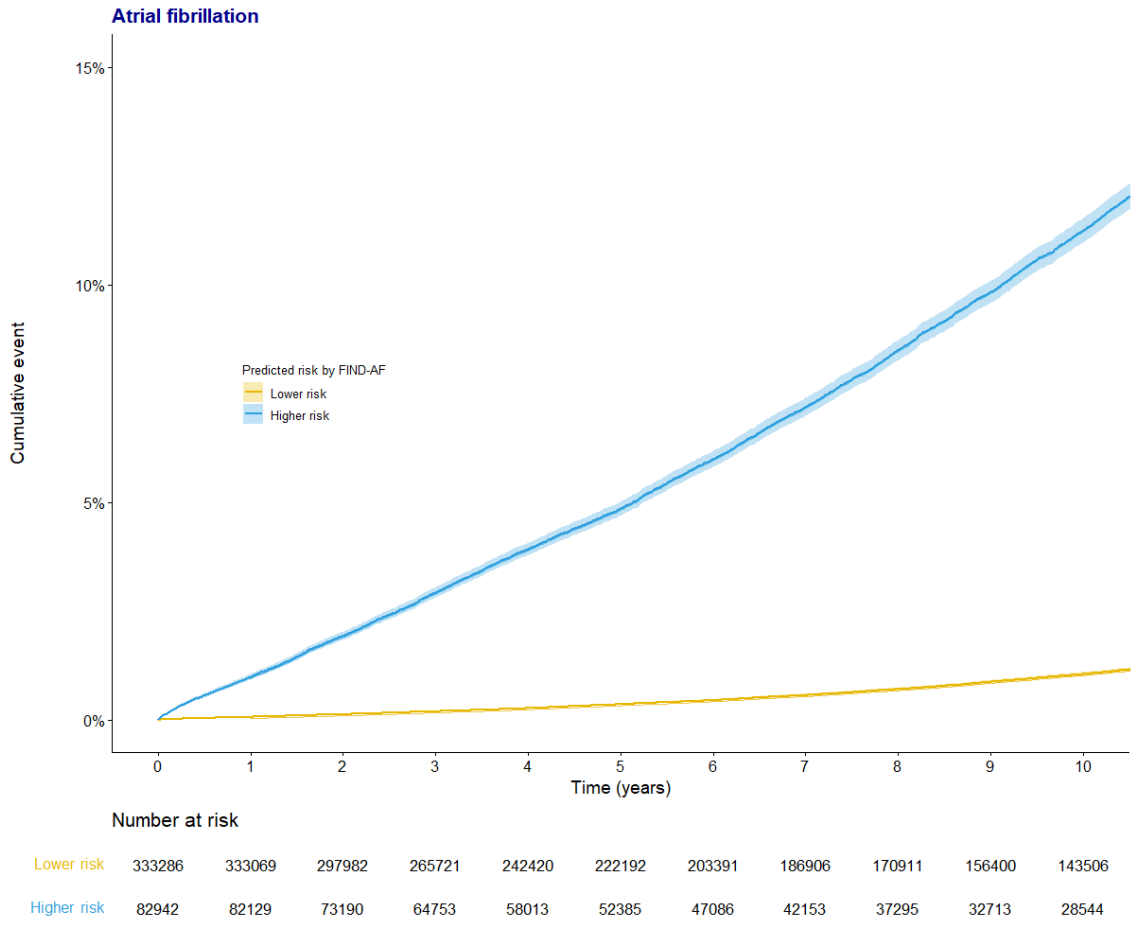


Abbreviations: CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation

Of the 82 942 individuals identified as higher risk by FIND-AF, 3 483 were <65 years of age, of whom 3 448 had a CHA_2DS_2-VASc score of at least 1. The incidence rate of AF in routine clinical practice at 6 months was 20-fold higher amongst individuals identified as a higher predicted risk of AF by FIND-AF compared with individuals identified as lower risk (2.0% vs 0.1%). In routine clinical practice, 1 in every 71 individuals aged ≥65 years were diagnosed with AF within 6 months, 1 in every 58 individuals aged ≥75 years and 1 in every 40 individuals identified at higher predicted AF risk.

Higher predicted AF risk was also associated with increased long-term AF occurrence. Within 5 and 10 years, respectively, 5.1% and 11.9% of the higher predicted risk cohort had been diagnosed with AF; with an 8.75-fold increased hazard (95% CI 8.44-9.06) relative to individuals at lower predicted risk (Figure 5).

Figure 5 Kaplan-Meier plots for atrial fibrillation occurrence, by predicted risk from FIND-AF



Abbreviations: FIND-AF, Future Innovations in Novel Detection of Atrial Fibrillation.

4.6.4 Model performance in clinically relevant subgroups

FIND-AF discrimination performance remained strong in both sexes, whereas for the CHA_2DS_2-VASc and C_2HEST scores performance was better in men than women (Table 10). The scores performed differently across ethnic groups. In Black individuals AF discrimination was highest for CHA_2DS_2-VASc, and in White and Asian individuals FIND-AF had the strongest discrimination performance.

Table 10 Discrimination performance of FIND-AF, CHA_2DS_2-VASc, and C_2HEST by sex, age and ethnicity

	FIND-AF	**CHA_2DS_2-VASc**	**C_2HEST**
	AUROC (95% CI)	AUROC (95% CI)	AUROC (95% CI)
Overall	0.824 (0.814-0.834)	0.784 (0.773-0.794)	0.757 (0.744-0.770)
Sex			
Men	0.819 (0.809-0.829)	0.807 (0.793-0.821)	0.793 (0.777-0.810)
Women	0.821 (0.810-0.831)	0.776 (0.760-0.793)	0.746 (0.727-0.765)
Age			
≥65 years	0.712 (0.698-0.727)	0.669 (0.654-0.684)	0.675 (0.661-0.690)
≥75 years	0.657 (0.638-0.675)	0.612 (0.593-0.632)	0.589 (0.570-0.608)
Ethnicity			
White	0.810 (0.799-0.821)	0.781 (0.769-0.792)	0.756 (0.743-0.770)
Asian	0.796 (0.693-0.899)	0.758 (0.639-0.876)	0.731 (0.611-0.850)
Black	0.801 (0.680-0.923)	0.843 (0.764-0.923)	0.707 (0.511-0.902)

Oher non-white ethnic minority	0.805 (0.765-0.845)	0.768 (0.729-0.807)	0.805 (0.765-0.846)
Ethnicity unrecorded	0.823 (0.770-0.875)	0.838 (0.777-0.900)	0.788 (0.705-0.870)

Abbreviations: AUROC, area under receiver operating characteristic; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); CI, confidence interval; FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation

N.B. The total number of individuals in each subgroup and number of incident AF cases is as follows: Men (N = 211 378, AF = 720), Women (N = 204 850, AF = 753), Age ≥65 years (N = 81 258, AF = 1 168), Age ≥75 years (N = 36 358, AF = 796), White (N = 279 027, AF = 1 301), Asian (N = 8 422, AF = 16), Black (N = 6478, AF = 11), Other non-white ethnic minority (N = 28 303, AF = 96), Ethnicity unrecorded (N = 93 998, AF = 49).

4.7 Discussion

In this population-based study, we trained a machine learning algorithm (FIND-AF) on more than 1.5 million individuals registered in UK primary care to predict the risk of incident AF within the next 6 months. When tested in over 400 000 individuals, FIND-AF demonstrated good predictive accuracy, which was superior to other risk scores and robust in both sexes and across ethnic groups. FIND-AF identified a cohort of younger people at higher risk of AF and more efficiently identified individuals diagnosed with AF within 6 months compared with age-based risk stratification. Finally, short-term predicted AF risk also translated to long-term AF occurrence.

Current approaches to targeting investigation for undiagnosed AF are based on age.[1] Our analysis demonstrated that a fifth of newly detected AF cases within 6 months occur in people aged ≤65 years, emphasising the opportunity lost when enhanced AF investigation is restricted to older populations. Electrocardiograms can be used to accurately predict AF risk,[270] but they are not widely available in the community

whereas 98% of the UK population are registered in primary care with an accompanying EHR.[202] Our meta-analysis of AF prediction algorithms using EHRs demonstrated that algorithms developed using traditional regression techniques provided only moderate discrimination performance.[197] In our study a machine learning prediction algorithm (FIND-AF) outperformed the C_2HEST and CHA_2DS_2-VASc scores.

For a machine learning prediction algorithm to be useful in clinical practice it must be implementable within the clinical workflow, provide prediction that meaningfully informs decision making, and engender confidence in how outputs were arrived at.[282] FIND-AF has been designed to be implemented and displayed through EHR systems, so will be available in a platform that healthcare professionals are interacting with as part of routine care. By design, FIND-AF provides AF risk prediction over a short time-frame and so could assist clinicians at point of care in identifying patients for targeted diagnostics such as ECG monitoring. Finally, the most important predictors in FIND-AF are already well-recognised risk factors for AF (for example age, heart failure, valvular heart disease), which provides reassurance in the associations being made by the algorithm.[1]

Fairness is a critical characteristic when considering the impact of prediction algorithms in healthcare. The CHARGE-AF and PuLSE-AI algorithms have strong AF prediction performance,[194, 196] yet incorporate variables that are frequently missing (height, weight and systolic and diastolic blood pressure).[197] Consequently, their applicability is limited to 17% and 35% of primary care EHRs, respectively.[194, 196] Often health data poverty disproportionately affects individuals from minority ethnicities and deprived backgrounds, so the application of these algorithms could reinforce health inequities.[283] Furthermore, whether their performance varies by sex and in minority ethnic groups in European populations is unknown. In our study the C_2HEST and CHA_2DS_2-VASc scores were less accurate in women compared with men, and their performance varied substantially across different ethnic groups. FIND-AF's design enabled its application to every single patient record in a nationally representative dataset of routinely-collected primary care EHRs; and performance was robust in both sexes and across minority ethnic groups.

Three barriers need to be overcome for FIND-AF to be accepted into clinical practice. First, it requires external validation, which is planned to be conducted in the TPP UK primary care EHR system (ResearchOne) and the Israeli Clalit Health Services. Second, prospective validation of FIND-AF is critical before implementation into clinical practice. We are launching a pilot implementation study across primary care sites

where individuals identified at higher risk will be offered rhythm monitoring (The BHF Bristol Myers Squibb Cardiovascular Catalyst Award – CC/22/250026). Third, a cost utility analysis and budget impact analysis of the use of FIND-AF will need to be conducted.

Primary care EHRs in the UK are nationwide and held centrally, so FIND-AF could be activated at scale across geographically disparate sites to identify a subpopulation at elevated AF risk. The cohort identified as higher risk in this study included younger people who would currently be excluded from screening pathways, and higher predicted AF risk was associated with elevated AF occurrence both in the short- and long-term. Therefore, FIND-AF could facilitate efficient population-based AF screening or comprehensive programs designed to improve risk factor profiles (including targeted weight loss and optimisation of blood pressure control).[284]

Screening for AF would adhere to many of the Wilson & Junger principles for a screening programme.[45] Opportunistic screening guided by age has not been demonstrated to increase AF detection rates,[285] but this may change in a more precisely defined higher risk cohort. Systematic screening of older patients with intermittent or continuous (invasive or non-invasive) rhythm monitors is associated with increased AF detection rates, compared to routine care.[45] However, the yield of new cases is low (3% in the STROKESTOP trial)[70] and in our study FIND-AF more efficiently identified a cohort with a higher rate of clinically detected AF than age-based approaches. Accurate risk assessment would be an integral component of a systematic screening process but ongoing research is needed to address the issues of the effectiveness and safety of treatment of screen-detected AF, and the costs of widespread use of ECG monitoring and prescription of oral anticoagulation, after the mixed results of the recently published LOOP and STROKESTOP trials.[70, 72]

There are some limitations to our study. First, the CPRD database is routinely-collected, retrospective primary care data. Underestimation of AF incidence is possible since there will have been individuals with unrecorded asymptomatic AF. Second, important predictor variables may have been ‘missing by design’; nonetheless, we aimed to develop an algorithm that used routinely recorded data. Third, our choice of a random forest classifier was based on a systematic review of AF prediction in EHRs,[197] and it is possible other machine learning methods may have performed differently in our study. Fourth, the algorithm will need to be updated as population characteristics change, data quality of EHRs improves and new or additional risk factors emerge. Fifth, electrophysiology procedures not specified as treating atrial fibrillation (including

pacemaker implantations and percutaneous ablations) were a strong predictor of AF risk, and this may be a result of detection bias.

4.8 Conclusions

We trained and tested a novel machine learning algorithm (FIND-AF) that was applicable at scale within a nationwide routinely-collected primary care EHR dataset. FIND-AF was able to accurately predict AF risk within 6 months and identify a cohort at elevated risk of AF in the longer-term.

Chapter 5 Incident cardiovascular, renal, metabolic diseases and death in individuals identified for risk-guided atrial fibrillation screening: a nationwide cohort study

Jianhua Wu*, Ramesh Nadarajah*, Yoko M Nakao, Kazuhiro Nakao, David C Hogg, Keerthenan Raveendra, Ronen Arbel, Moti Haim, Doron Zahger, Campbell Cowan, Chris P Gale

*Jianhua Wu and Ramesh Nadarajah are joint first authors

5.1 Summary of the publication

- This analysis was performed using the methodology presented in Chapter 3 to quantify the association of predicted AF risk with non-AF outcomes.
- The study found that higher predicted AF risk, compared with lower predicted AF risk, was associated with higher risk and shorter median time to event for a range of incident cardio-renal-metabolic diseases and death.

5.2 Publication status

- Published July 10 2023
- Open Heart 2023;10:e002357

5.3 Abstract

5.3.1 Objective

Risk-guided AF screening may be an opportunity to prevent adverse events in addition to stroke. We compared events rates for new diagnoses of cardio-renal-metabolic diseases and death in individuals identified at higher versus lower predicted AF risk.

5.3.2 Methods

From the UK Clinical Practice Research Datalink-GOLD dataset, Jan 2, 1998 to Nov 30, 2018, we identified individuals aged ≥30 years without known AF. The risk of AF

was estimated using the FIND-AF risk score. We calculated cumulative incidence rates and fitted Fine and Gray's models at 1, 5, and 10 years for nine diseases and death adjusting for competing risks.

5.3.3 Results

Of 416 228 individuals in the cohort, 82 942 were identified as higher risk for AF. Higher predicted risk, compared with lower predicted risk, was associated with incident chronic kidney disease (cumulative incidence per 1000 persons at 10 years 245.2; HR 6.85, 95% CI 6.70-7.00; median time to event 5.44 years), heart failure (cumulative incidence per 1000 persons at 10 years 124.7; HR 12.54, 95% CI 12.08-13.01; median time to event 4.06), diabetes mellitus (cumulative incidence per 1000 persons at 10 years 123.3; HR 2.05, 95% CI 2.00-2.10; median time to event 3.45), stroke/transient ischaemic attack (cumulative incidence per 1000 persons at 10 years 118.9; HR 8.07, 95% CI 7.80-8.34; median time to event 4.27), myocardial infarction (cumulative incidence per 1000 persons at 10 years 69.6; HR 5.02, 95% CI 4.82-5.22; median time to event 4.32), peripheral vascular disease (cumulative incidence per 1000 persons at 10 years 44.6; HR 6.62, 95% CI 6.28-6.98; median time to event 4.28), valvular heart disease (cumulative incidence per 1000 persons at 10 years 37.8; HR 6.49, 95% CI 6.14-6.85; median time to event 4.54), aortic stenosis (cumulative incidence per 1000 persons at 10 years 18.7; HR 9.98, 95% CI 9.16-10.87; median time to event 4.41) and death from any cause (cumulative incidence per 1000 persons at 10 years 273.9; HR 10.45, 95% CI 10.23-10.68; median time to event 4.75). The higher risk group constituted 74% of deaths from cardiovascular or cerebrovascular causes (8 582/11 676).

5.3.4 Conclusions

Individuals identified for risk-guided AF screening are at risk of new diseases across the cardio-renal-metabolic spectrum and death, and may benefit from interventions beyond ECG monitoring.

5.4 Introduction

Atrial fibrillation (AF) screening research has hitherto primarily focused on stroke prophylaxis through early detection of AF and initiation of oral anticoagulation. Randomised controlled trials have demonstrated that non-invasive electrocardiogram (ECG) monitoring in older people with or without stroke risk factors increases detection

rates of previously undiagnosed AF compared with routine standard of care,[68, 87, 88] but yields are relatively low (<5%) and the net benefit small.[70]

Atrial fibrillation (AF) frequently develops due to, and in parallel with, other cardiovascular, renal and metabolic conditions.[250] Over 70% of new diagnoses have at least two concomitant, chronic comorbidities,[211] and thereafter are at an increased risk of major cardiovascular events beyond stroke, including ischemic heart disease, heart failure, chronic kidney disease, peripheral vascular disease and death.[32]

Risk-guided AF screening has the potential to achieve a higher yield of AF detection than age-guided screening.[286] Furthermore, individuals identified at elevated risk of AF may have an age and comorbidity profile similar to individuals with diagnosed AF, and thus also be at risk of subsequent adverse events. If so, a risk-guided AF screening strategy may provide an opportunity for the identification and management of concomitant diseases and cardiometabolic risk factors to prevent a range of adverse events beyond stroke.[250]

To determine whether individuals identified for risk-guided AF screening are at increased risk of adverse events we used a large nationwide longitudinal database of linked primary and secondary care records to study event rates in the subpopulation at higher predicted AF risk for a range of new-onset cardio-renal-metabolic diseases and death.

5.5 Methods

5.5.1 Data source

We used EHRs from the Clinical Practice Research Datalink (CPRD). The CPRD database contains anonymised patient data from approximately 7% of the UK population and is broadly representative in terms of age, sex, and ethnicity.[202] CPRD is one of the world's largest databases of longitudinal medical records from primary care. The dataset used for this analysis was primary care records from CPRD that had been linked to secondary care admission records from Hospital Episodes Statistics Admitted Patient Care (HES-APC) data and death certificates from the Office for National Statistics (ONS). Linkage is available for a subset of English practices from Jan 3, 1998 to Nov 30, 2018, covering approximately 50% of all CPRD records. Previous research has demonstrated the representativeness of patients eligible for linkage in terms of

age, sex and geography.[287] More than 200 independent studies have investigated the validity of diagnoses recorded in CPRD, which reported an average positive predictive value of about 90% for a broad range of conditions.[254]

5.5.2 Study population

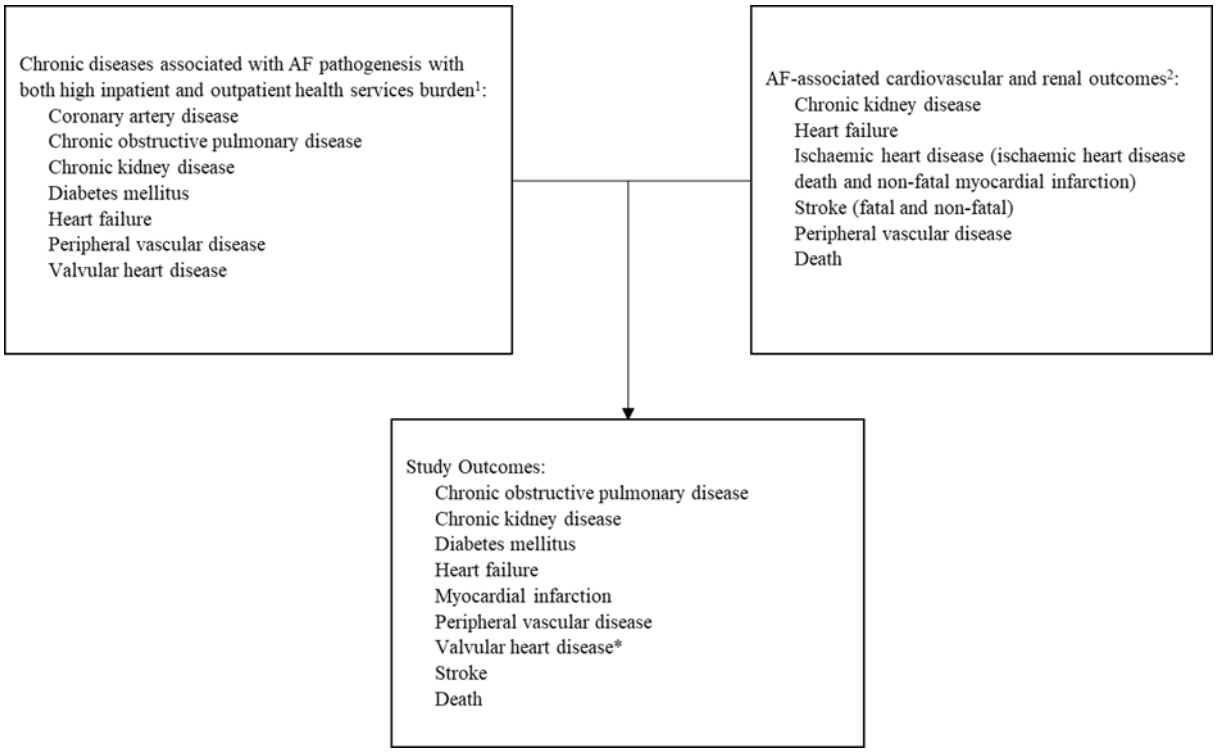
We included adults registered at practices within CPRD who were ≥30 years of age at entry with no prior history of AF and at least one-year follow-up, between January 2, 1998 and November 30, 2018. All individuals were categorized as lower or higher predicted AF risk by the FIND-AF risk score,[286] with the higher risk cohort reflecting individuals who would be identified for risk-guided AF screening.

The FIND-AF risk score predicts incident AF at 6 months for individuals ≥30 years of age without a preceding diagnosis of AF.[286] The risk score is scalable through community-based EHRs because it only requires data for age, sex, comorbidities and ethnicity (included an 'ethnicity unrecorded' category where it was unavailable because missingness was considered to be informative)[268]. The risk score was found to have stronger discriminative performance, reclassification and net benefit for short-term incident AF than the CHA_2DS_2-VASc and C_2HEST scores, and more efficiently identify individuals who develop AF than an age-guided approach.[286]

5.5.3 Outcomes

The primary endpoint for the analysis was the initial presentation of a cardiovascular, renal, or metabolic disease or death. To best characterise highly prevalent and morbid diseases, associated with the development or consequence of AF (Figure 1),[32, 250] we individually examined the following nine conditions: heart failure, valvular heart disease (and specifically aortic stenosis), myocardial infarction, stroke (ischaemic and haemorrhagic) or transient ischaemic attack, peripheral vascular disease, chronic kidney disease, diabetes mellitus, as well as chronic obstructive pulmonary disease (COPD). Aortic stenosis was further specified in addition to valvular heart disease given the increasing availability and randomised controlled trial evidence for earlier treatment, and increasing therapeutic options across operative risk profiles.[274]

Figure 1. Study design process leading to selection of study outcomes



Abbreviations: AF, atrial fibrillation

N.B. * Aortic stenosis was further specified in addition to valvular heart disease given the increasing availability and randomised controlled trial evidence for earlier treatment, and increasing therapeutic options across operative risk profiles.[274]

We also investigated for occurrence of death by any cause recorded in primary care or by death certification from the UK Death Register of the ONS, which was mapped on to 9 disease categories following previously established methods, as summarised in Chapter 3 Table 2.[272]

For each condition, a list of diagnostic codes from the CALIBER code repository, including from International Classification of Diseases 10th revision (used in secondary care) and Read coding schemes (used in primary care) was defined to comprehensively identify diagnoses from EHRs. Incident diagnoses were defined as the first record of that condition in primary or secondary care records from any diagnostic position. For definition of new cases, we excluded individuals for the analysis of each condition who had a diagnosis of that condition before the patient's entry to the study. If no indication of a specific disease was recorded, then the patient was assumed to be free from the disease.

5.5.4 Statistical analysis

The baseline characteristics are summarised by predicted AF status. Continuous variables were reported as mean ± standard deviation (SD). Categorical variables were reported as frequencies with corresponding percentages.

We created Kaplan-Meier plots for individuals identified as higher and lower predicted risk of AF and derived the cumulative incidence rate for each outcome at 1, 5 and 10 years considering the competing risk of death, as well as death at 5 and 10 years. For each specified outcome, we calculated the hazard ratio (HR) between higher and lower predicted risk of AF using the Fine and Gray's model with adjustment for the competing risk of death. We reported unadjusted HR and adjusted HR where the model was adjusted for age, sex, ethnicity and the presence of any of the other outcomes at baseline.

Given that age and sex were two key variables in the FIND-AF algorithm,[286] and some of the outcomes have incidence rates that are strongly associated with age (e.g. aortic stenosis) or differ by sex (e.g. heart failure),[273, 274] we conducted sub-group analyses of incidence rates for higher and lower risk individuals for each outcome by age group (30-64 years and ≥65 years) and sex. As some of the outcomes are more likely to occur in the setting of prevalent AF (e.g. stroke or heart failure),[250] we also conducted a sensitivity analysis where people with incident AF during follow-up were excluded.

Study findings are reported in accordance with the Reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) recommendations,[288] and the CODE-EHR best-practice framework for using structured electronic healthcare records in clinical research.[278] We used R version 4·1·0 for all analyses.

5.5.5 Patient and public involvement

The Arrhythmia Alliance an AF association provided input on the FIND-AF scientific advisory board. The FIND-AF patient and public involvement group have given input to reporting and dissemination plans of the research.

5.6 Results

5.6.1 Patient population

In the cohort of 416 228 individuals (average age 49.9 [SD 15.4] years, 50.8% women, 86.8% white), 82 942 (19.9%) were identified as higher predicted risk of AF, 3 483 of whom were <65 years of age, with 1 203 and 8 876 diagnosed with AF over 6 months and 10 years of follow up, respectively. At point of risk prediction, those at higher compared with lower predicted AF risk had a higher average age and prevalence of baseline comorbidities (Table 1).

Table 1 Baseline characteristics of analytical cohort stratified by predicted atrial fibrillation risk

	FIND-AF predicted risk	
	Lower risk n (%)	**Higher risk n (%)**
	333 286	82 942
Demographics		
Age, years	44.1 (10.40)	73.2 (8.75)
Sex (women)	170 568 (51.2)	41 210 (49.7)
Ethnicity		
Asian	7 385 (2.2)	894 (1.1)
Black	5 786 (1.7)	613 (0.7)
Other	22 033 (6.6)	5 878 (7.1)
Unknown	91 505 (27.5)	2 161 (2.6)
White	206 577 (62.0)	73 396 (88.5)
Comorbidities		
Anaemia	9 118 (2.7)	4 251 (5.1)
Aortic stenosis	63 (<0.1)	316 (0.4)
Cancer	6 120 (1.8)	8 303 (10.0)
COPD	1 111 (0.3)	4 019 (4.8)
Chronic kidney disease	2 938 (0.9)	2 990 (3.6)

Diabetes mellitus	6 328 (1.9)	8 072 (9.7)
Dyslipidaemia	6 095 (1.8)	5 984 (7.2)
Ischaemic heart disease	3 299 (1.0)	12 486 (15.1)
Heart failure	163 (<0.1)	2 748 (3.3)
Hypertension	20 139 (6.0)	29 594 (35.7)
Hyperthyroidism	1 883 (0.6)	1 370 (1.7)
Stroke/TIA	1 376 (0.4)	6 375 (7.7)
Valvular heart disease	562 (0·2)	1 414 (1·7)

Abbreviations: AF, atrial fibrillation; COPD, chronic obstructive pulmonary disease; TIA, transient ischaemic attack

The cohort with higher predicted AF risk had similar baseline characteristics and mean $CHA_2D_2$2-VASc score to the cohort who developed AF during follow up, but a lower prevalence of ischaemic heart disease (15.1% vs 20.2%), prior stroke or transient ischaemic attack (7.7% vs 12.2%), hypertension (35.7 % vs 40.0%), valvular heart disease (1.7% vs 5.4%) and chronic kidney disease (3.6% vs 6.4%) (Table 2).

Table 2 Baseline characteristics of testing set, stratified by incident atrial fibrillation and predicted atrial fibrillation risk

	Incident atrial fibrillation		**FIND-AF predicted risk**	
	no AF **n (%)**	**AF** **n (%)**	**Lower risk** **n (%)**	**Higher risk** **n (%)**
	414 676	1 552	333 286	82 942
Demographics				
Age, years	49·82 (15·38)	73·87 (12·47)	44·11 (10·40)	73·24 (8·75)
Sex (women)	210 646 (50·8)	755 (48·6)	170 568 (51·2)	41 210 (49·7)
Ethnicity				
Asian	8 258 (2·0)	21 (1·5)	7 385 (2·2)	894 (1·1)
Black	6 390 (1·5)	9 (0·6)	5 786 (1·7)	613 (0·7)
Other	27 805 (6·7)	106 (7·4)	22 033 (6·6)	5 878 (7·1)
Unknown	93 630 (22·6)	36 (2·5)	91 505 (27·5)	2 161 (2·6)
White	278 714 (67·2)	1 259 (88·0)	206 577 (62·0)	73 396 (88·5)
Comorbidities				
Diabetes mellitus	14 649 (3·5)	171 (11·0)	6328 (1·9)	8072 (9·7)
Stroke or TIA	7 467 (1·8)	189 (12·2)	1376 (0·4)	6375 (7·7)

Ischaemic heart disease	15 483 (3·7)	314 (20·2)	3299 (1·0)	12486 (15·1)
Hypertension	49 494 (11·9)	621 (40·0)	20139 (6·0)	29594 (35·7)
Heart failure	2 745 (0·7)	132 (8·5)	163 (0·0)	2748 (3·3)
Dyslipidaemia	12 122 (2·9)	121 (7·8)	6095 (1·8)	5984 (7·2)
Hyperthyroidism	3 203 (0·8)	44 (2·8)	1883 (0·6)	1370 (1·7)
COPD	4 987 (1·2)	106 (6·8)	1111 (0·3)	4019 (4·8)
Chronic kidney disease	5 839 (1·4)	99 (6·4)	2938 (0·9)	2990 (3·6)
Anaemia	13 165 (3·2)	106 (6·8)	9118 (2·7)	4251 (5·1)
Cancer	14 710 (3·5)	186 (12·0)	6120 (1·8)	8303 (10·0)
Valvular heart disease	1 881 (0·5)	84 (5·4)	562 (0·2)	1414 (1·7)
Mean CHA_2DS_2-VASc score (SD)	0·97 (1·03)	2·74 (1·40)	0·62 (0·62)	2·42 (1·14)

Abbreviations: AF, atrial fibrillation; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 years [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74 years, Sex Category; COPD, chronic obstructive pulmonary disease; TIA, transient ischaemic attack

5.6.2 Outcomes

Higher predicted AF risk, compared with lower predicted AF risk was associated with increased occurrence for each pre-specified condition at 1, 5 and 10 years of follow-up (Figure 2; Table 3).

Figure 2 Kaplan-Meier plots for the ten outcomes

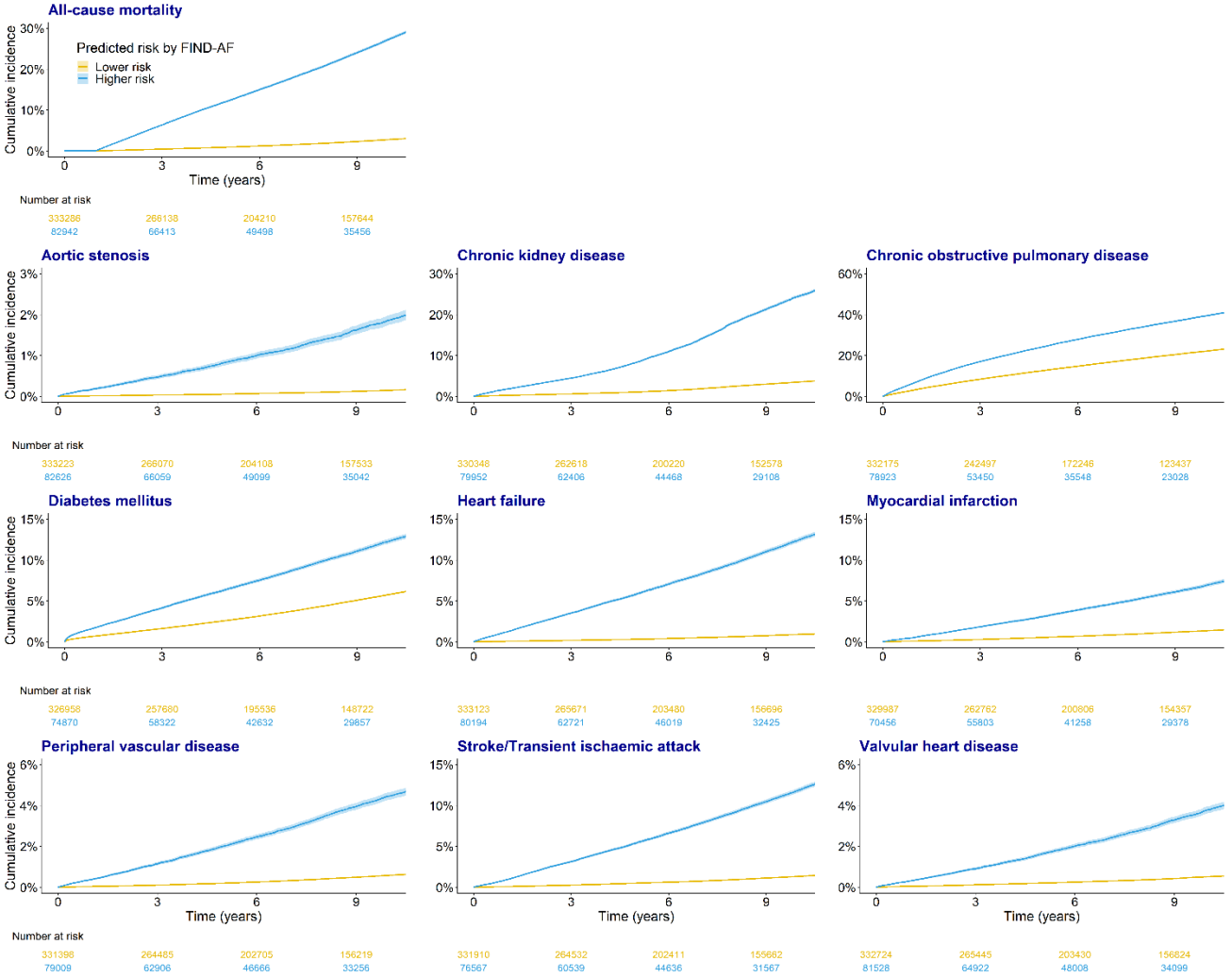


Table 3 Cumulative incidence rate for the 10 outcomes stratified by predicted atrial fibrillation risk

	Median time to event (years, IQR)		**Cumulative incidence (per 1000 persons)**					
Outcome	**Predicted lower risk**	**Predicted higher risk**	**Predicted lower risk**			**Predicted higher risk**		
			1-year	**5-year**	**10-year**	**1-year**	**5-year**	**10-year**
Aortic stenosis	5.23 (2.45-7.81)	4.41 (1.98-7.18)	0.1 (0.1-0.2)	0.5 (0.4-0.6)	1.5 (1.3-1.7)	1.6 (1.4-1.9)	8.2 (7.5-8.9)	18.7 (17.5-19.9)
COPD	3.12 (1.28-5.84)	2.68 (1.11-5.31)	32.2 (31.6-32.8)	127.4 (126.2-128.6)	222.2 (220.4-223.9)	68.4 (66.6-70.2)	244.6 (241.4-247.8)	395.8 (391.5-400.0)
Chronic kidney disease	5.95 (3.03-7.83)	5.44 (2.76-7.60)	2.3 (2.1-2.4)	10.6 (10.2-11.0)	35.3 (34.5-36.1)	17.4 (16.5-18.4)	82.9 (80.8-85.0)	245.2 (241.3-249.1)

Diabetes mellitus	4.24 (1.62-7.10)	3.45 (1.30-6.33)	7.2 (6.9-7.4)	26.1 (25.5-26.7)	57.9 (56.9-58.9)	17.9 (16.9-18.8)	64.4 (62.5-66.3)	123.3 (120.4-126.3)
Heart failure	5.49 (2.71-7.89)	4.06 (1.82-6.84)	0.6 (0.5-0.6)	3.1 (2.9-3.3)	9.0 (8.6-9.4)	11.9 (11.2-12.7)	58.3 (56.5-60.1)	124.7 (121.7-127.6)
Myocardial infarction	4.95 (2.54-7.50)	4.32 (2.03-6.88)	0.9 (0.8-1.0)	5.4 (5.1-5.7)	13.6 (13.1-14.1)	5.5 (5.0-6.1)	31.4 (30.0-32.8)	69.6 (67.2-72.0)
Peripheral vascular disease	5.59 (2.83-7.83)	4.28 (2.05-6.96)	0.4 (0.3-0.4)	2.0 (1.8-2.1)	5.8 (5.5-6.2)	3.7 (3.3-4.2)	20.1 (19.1-21.2)	44.6 (42.8-46.4)
Stroke/TIA	5.17 (2.63-7.79)	4.27 (2.01-6.92)	0.8 (0.7-0.9)	5.0 (4.7-5.2)	13.3 (12.8-13.8)	9.2 (8.6-9.9)	54.1 (52.4-55.9)	118.9 (116.0-121.8)
Valvular heart disease	4.89 (2.25-7.72)	4.54 (2.12-7.11)	0.5 (0.4-0.5)	2.0 (1.9-2.2)	5.2 (4.8-5.5)	3.0 (2.6-3.4)	16.3 (15.4-17.3)	37.8 (36.1-39.5)

All-cause mortality	5.72 (3.24-8.06)	4.75 (2.66-7.27)		9.2 (8.8-9.6)	27.9 (27.2-28.6)		121.6 (119.2-124.0)	273.9 (270.2-277.5)

1 Abbreviations: COPD, chronic obstructive pulmonary disease; IQR, interquartile range; TIA, transient ischaemic attack

A quarter of individuals in the higher predicted AF risk cohort were diagnosed with COPD within 5 years and with chronic kidney disease within 10 years. Furthermore, within 10 years each of heart failure, diabetes mellitus and stroke or transient ischaemic attack were diagnosed in more than 10% of individuals at higher predicted AF risk. Relative to individuals at lower predicted AF risk, those with higher predicted AF risk were at 12.54-fold (95% CI 12.08-13.01) increased risk for heart failure, 9.98-fold increased risk for aortic stenosis (95% CI 9.16-10.87) and 8.07-fold increased risk for stroke/transient ischaemic attack (95% CI 7.80-8.34) (Table 4).

Table 4 Hazard ratios for incident outcomes comparing individuals at higher and lower predicted atrial fibrillation risk

	Events/Cohorts			
Outcome	**Lower risk**	**Higher risk**	**Unadjusted hazard ratio (95% CI)**	**Adjusted hazard ratio (95% CI)**
Aortic stenosis	851/ 333 223	1 557/ 82 626	9.98 (9.16-10.87)	1.64 (1.43-1.87)
COPD	66 941/ 332 175	27 110/ 78 923	2.02 (2.00-2.05)	1.17 (1.14-1.20)
Chronic kidney disease	15 077/ 33 0348	17 494/ 79 952	6.85 (6.70-7.00)	1.46 (1.41-1.51)
Diabetes mellitus	21 627/ 326 958	8 338/ 74 870	2.05 (2.00-2.10)	1.06 (1.02-1.10)
Heart failure	4 135/ 333 123	9 453/ 80 194	12.54 (12.08-13.01)	1.63 (1.54-1.73)
Myocardial infarction	5 111/ 329 987	4 483/ 70 456	5.02 (4.82-5.22)	1.09 (1.03-1.17)
Peripheral vascular disease	2 470/ 331 398	3 176/ 79 009	6.62 (6.28-6.98)	1.30 (1.19-1.42)
Stroke/TIA	5 884/ 331 910	8 573/ 76 567	8.07 (7.80-8.34)	1.40 (1.33-1.48)

Valvular heart disease	2 426/ 332 724	2 946/ 81 528	6.49 (6.14-6.85)	1.56 (1.43-1.71)
All-cause mortality	12 804/ 333 286	25 814/ 82 942	10.45 (10.23-10.68)	1.06 (1.02-1.09)

Abbreviations: COPD, chronic obstructive pulmonary disease; CI, confidence interval; TIA, transient ischaemic attack

N.B. Model was adjusted for age, sex, ethnicity, and the presence of any of the other outcomes at baseline.

The higher predicted AF risk cohort were also more than five times more likely to be diagnosed with chronic kidney disease, valvular heart disease, myocardial infarction and peripheral vascular disease; and twice as likely to experience COPD or diabetes mellitus. Furthermore, the median time to event was shorter for each outcome in the higher predicted risk cohort compared with the lower predicted risk cohort, with a difference of over a year for heart failure (4.06 vs 5.49) and peripheral vascular disease (4.28 vs 5.59).

Death was common amongst persons identified as higher predicted AF risk, with over a quarter of patients having died by 10 years (Table 3). On unadjusted analysis individuals at higher predicted AF risk were at 10.5-fold increased hazard for death compared with individuals at lower predicted AF risk (95% CI 10.23-10.68; Table 4). Of the 25 814 deaths during 10 years follow-up in the higher predicted AF risk cohort, 8 582 (33%) were as a result of cardiovascular disease or cerebrovascular disease, with 5 931 (23%) attributed to cancer (Table 5).

Table 5 Cause of death stratified by FIND-AF risk classification

	Predicted AF risk	
Cause of death	**Lower risk n = 333 286**	**Higher risk n = 82 942**
Cardiovascular disease	2 506 (0.8)	6 006 (7.2)
Cerebrovascular disease	588 (0.2)	2 576 (3.1)
Chronic respiratory disease	751 (0.2)	1 952 (2.4)

Digestive disease	701 (0.2)	1 125 (1.4)
Infection	573 (0.2)	2 531 (3.1)
Injuries	494 (0.1)	471 (0.6)
Kidney disease	43 (0.0)	233 (0.3)
Mental and neurological disease	546 (0.2)	2 144 (2.6)
Neoplasms	4 889 (1.5)	5 931 (7.2)

Abbreviations: AF, atrial fibrillation

During the 10-year follow up, 70% of incident heart failure cases (9 453/13 588), and 65% of incident aortic stenosis diagnoses (1 557/2 408) occurred in individuals at higher predicted AF risk, even though they only accounted for less than a fifth of the total cohort. Of the 38 618 deaths that occurred during follow-up, two-thirds occurred in the higher predicted AF risk cohort (25 814; 67%). Specifically, individuals in the higher predicted AF risk cohort constituted three quarters of the deaths related to cardiovascular or cerebrovascular disease (8 582/11676; 74%), whereas the burden of death from neoplasm was more evenly distributed between individuals at lower and higher predicted AF risk (total deaths attributed to neoplasm 10 820; deaths in lower predicted AF risk cohort 4889 [45%]; deaths in higher predicted AF risk cohort 5931 [55%]).

5.6.3 Subgroup analysis

On subgroup analysis, higher predicted AF risk, compared with lower predicted AF risk, was associated with increased incidence for each of the outcomes in both men and women and in younger (age 30-64 years) and older (age ≥65 years) individuals (Figure 3-6).

Excluding patients with incident AF during follow up did not change the direction or magnitude of events (Table 6).

Figure 3 Kaplan-Meier plots for the ten outcomes in individuals aged 30-64 years at baseline

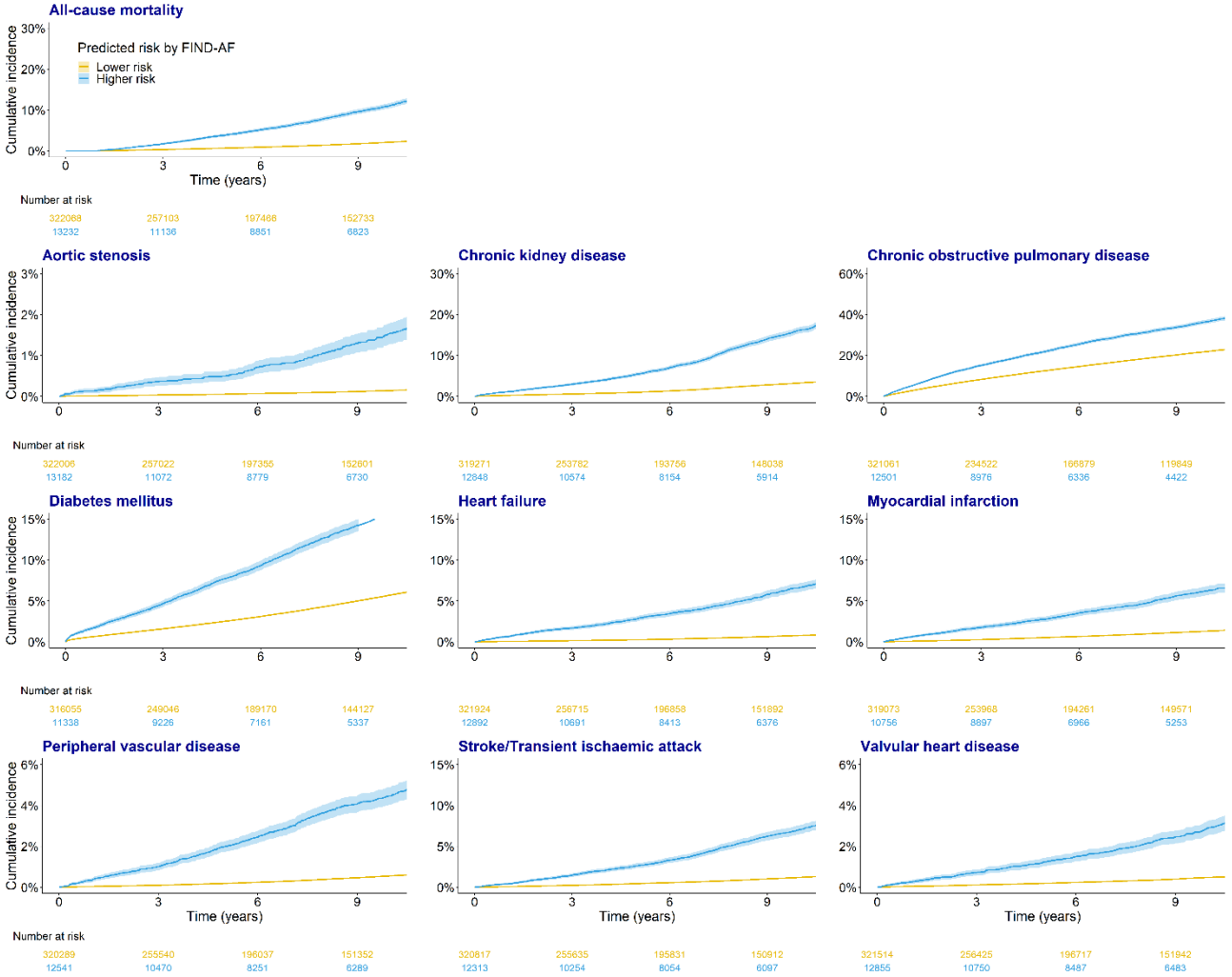


Figure 4 Kaplan-Meier plots for the ten outcomes in individuals aged ≥65 years at baseline

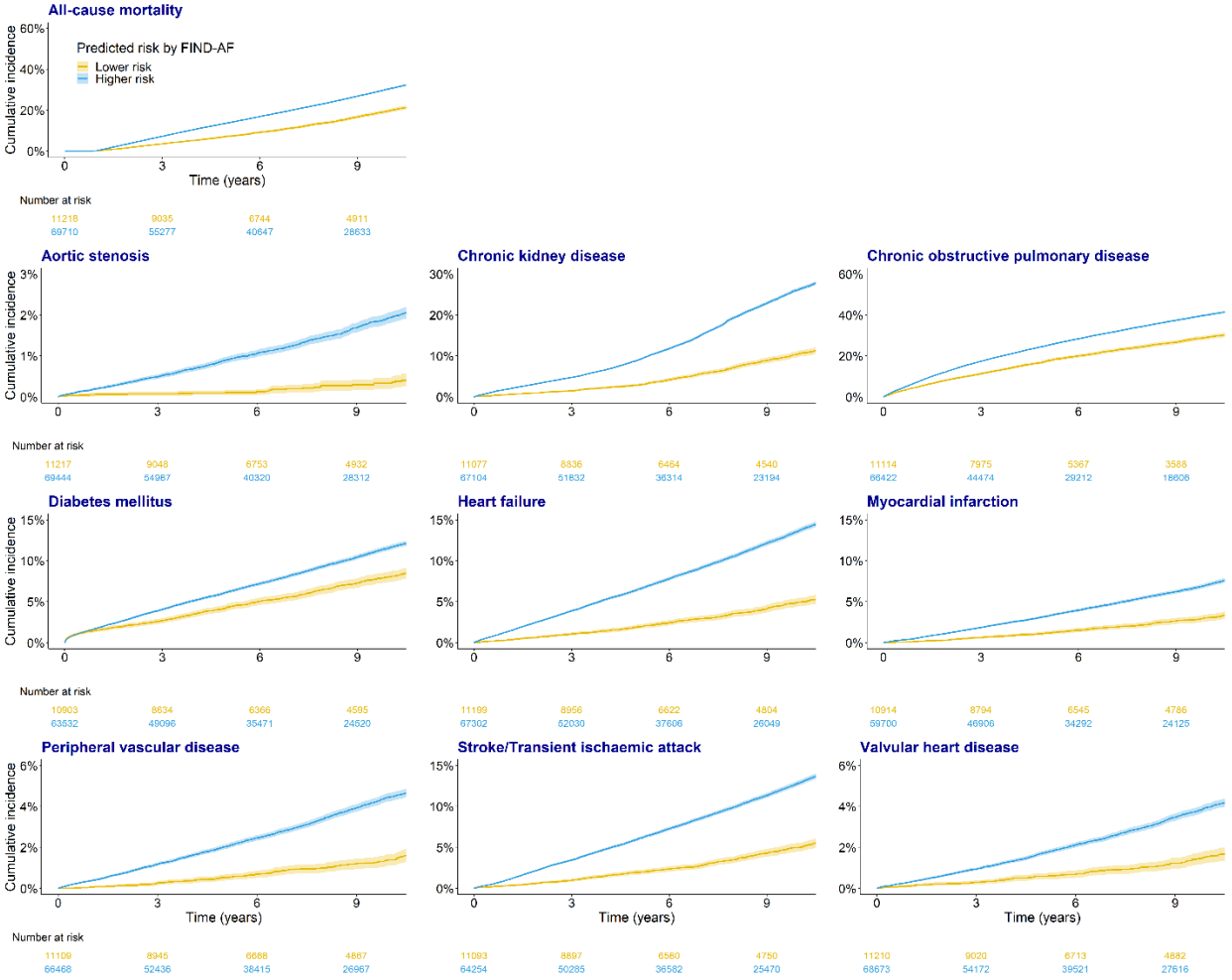


Figure 5 Kaplan-Meier plots for the ten outcomes in men

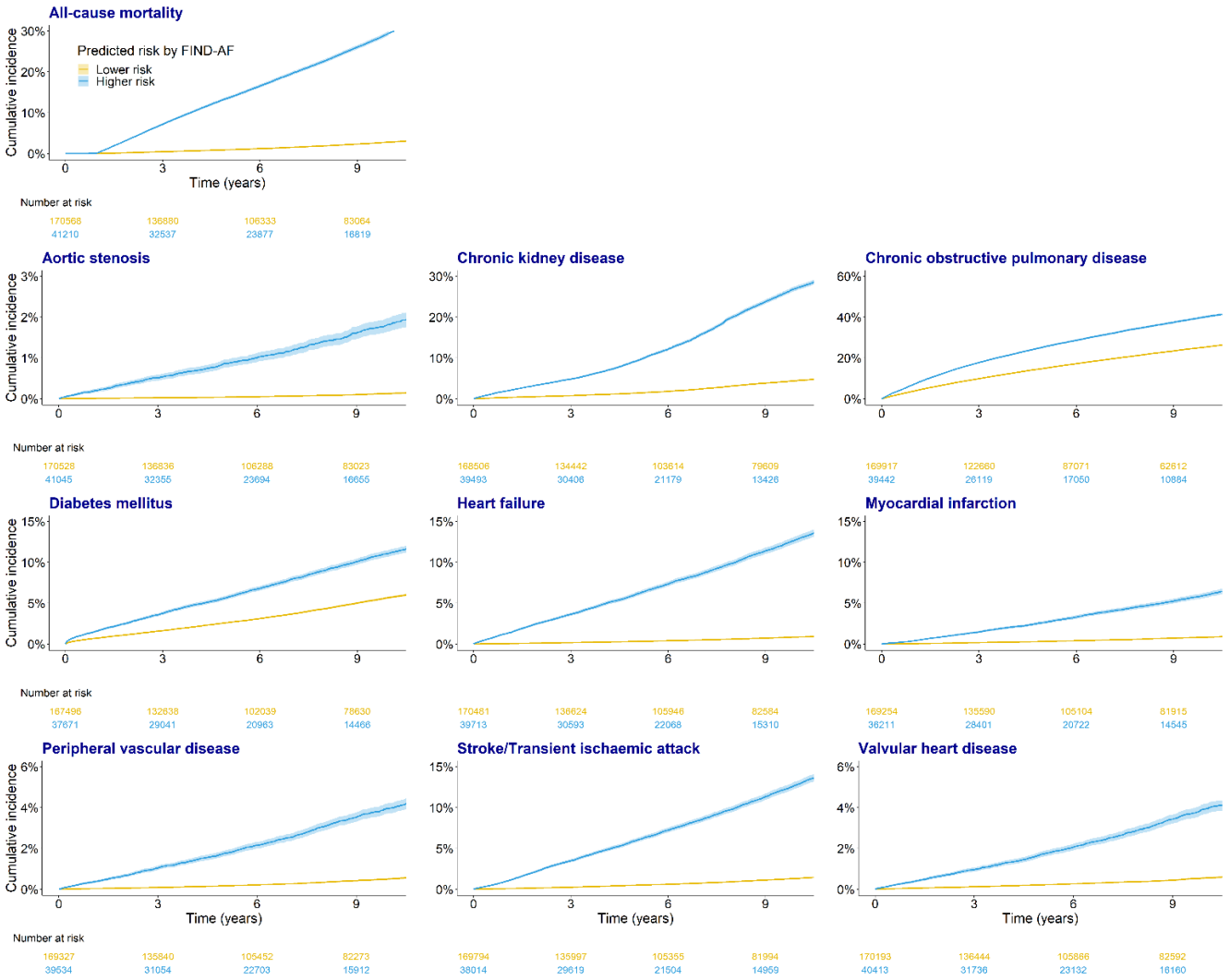


Figure 6 Kaplan-Meier plots for the ten outcomes in women

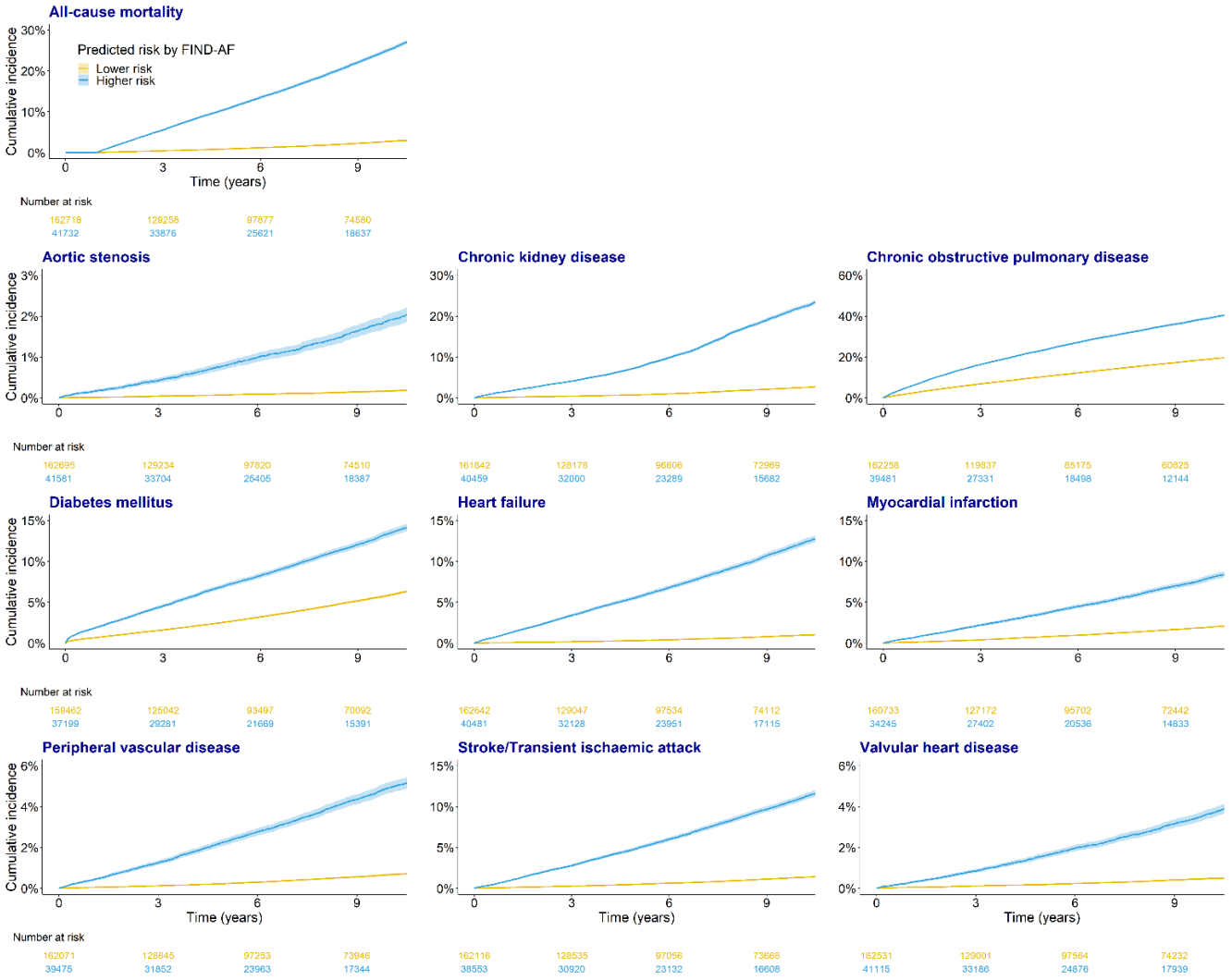


Table 6 Cumulative incidence rate for the 10 outcomes at 1, 5, and 10 years of follow up stratified by predicted atrial fibrillation risk, when incident atrial fibrillation cases are excluded

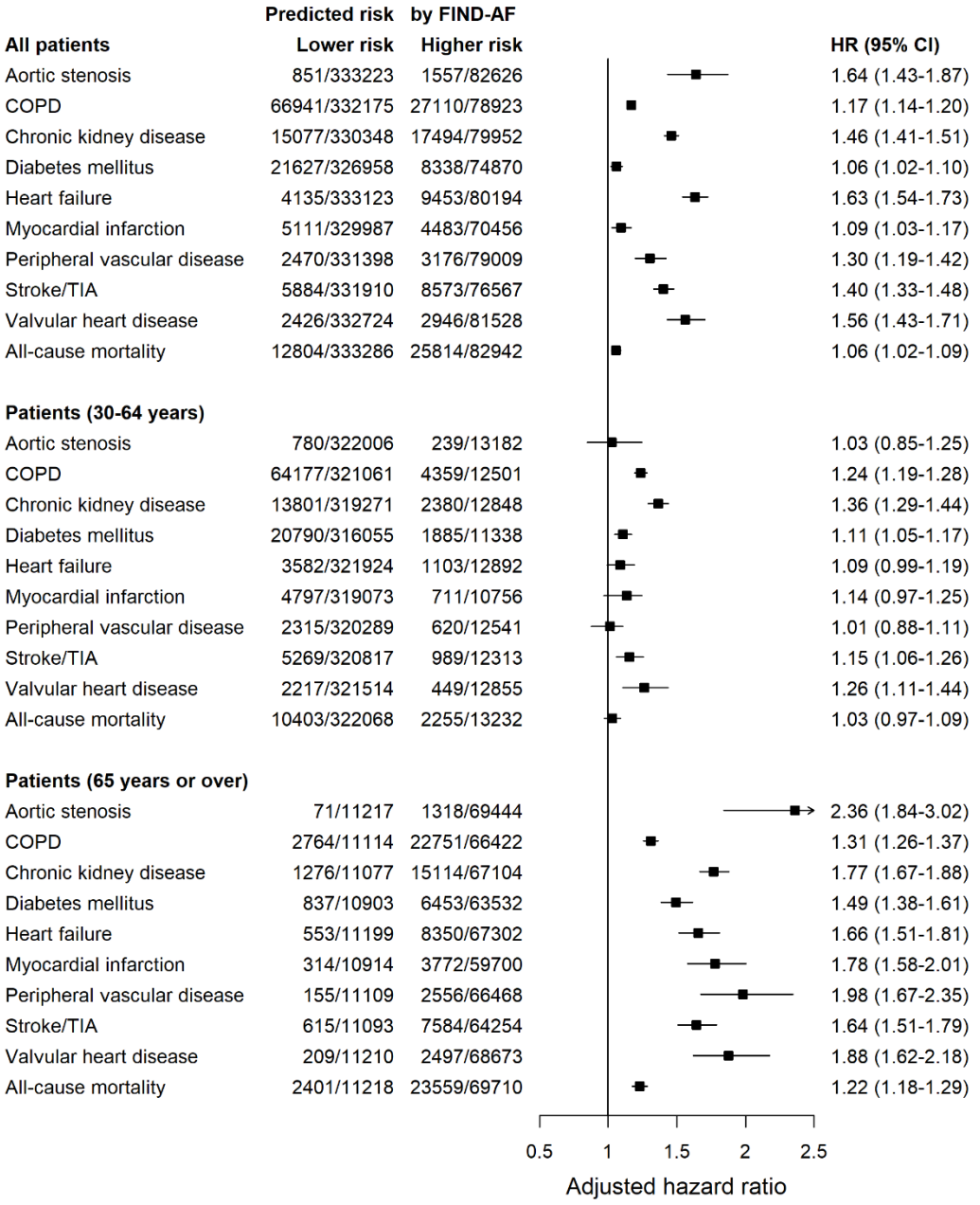
	Cumulative incidence (per 1000 persons)					
Outcome	**Predicted lower risk**			**Predicted higher risk**		
	1-year	**5-year**	**10-year**	**1-year**	**5-year**	**10-year**
Aortic stenosis	0.1 (0.1-0.1)	0.5 (0.4-0.5)	1.2 (1.1-1.4)	1.5 (1.2-1.7)	7.2 (6.5-7.8)	16.4 (15.2-17.7)
COPD	31.8 (31.2-32.4)	125.8 (124.6-127.1)	219.3 (217.5-221.0)	67.7 (65.8-69.5)	241.4 (237.9-244.8)	389.9 (385.3-394.4)
Chronic kidney disease	2.3 (2.1-2.4)	10.5 (10.1-10.9)	34.1 (33.3-34.9)	17.8 (16.8-18.8)	82.2 (80.0-84.4)	236.4 (232.2-240.5)
Diabetes mellitus	7.1 (6.8-7.4)	25.8 (25.2-26.4)	57.0 (56.0-58.1)	18.3 (17.2-19.3)	64.8 (62.7-66.8)	121.4 (118.2-124.5)
Heart failure	0.4 (0.4-0.5)	2.6 (2.4-2.7)	7.2 (6.8-7.6)	9.9 (9.2-10.6)	49.3 (47.5-51.0)	102.3 (99.4-105.1)
Myocardial infarction	0.8 (0.7-0.9)	5.1 (4.8-5.4)	12.9 (12.4-13.4)	5.3 (4.7-5.8)	30.2 (28.7-31.7)	66.6 (64.0-69.1)

Peripheral vascular disease	0.4 (0.3-0.4)	1.9 (1.7-2.1)	5.6 (5.2-5.9)	3.8 (3.3-4.2)	19.7 (18.6-20.9)	42.5 (40.6-44.5)
Stroke/TIA	0.8 (0.7-0.9)	4.7 (4.4-4.9)	12.2 (11.8-12.7)	8.8 (8.1-9.5)	51.9 (50.0-53.7)	111.4 (108.3-114.4)
Valvular heart disease	0.4 (0.3-0.5)	1.6 (1.5-1.8)	4.0 (3.7-4.3)	2.5 (2.2-2.9)	13.0 (12.1-13.9)	29.8 (28.2-31.5)
All-cause mortality		9.2 (8.8-9.6)	27.9 (27.2-28.6)		130.3 (127.6-132.9)	287.1 (283.1-291.1)

Abbreviations: TIA, transient ischaemic attack

After adjustment for age, sex, ethnicity and presence of any other outcomes at baseline, higher predicted AF risk remained associated with excess risk for all-cause death and each condition (Figure 7, Table 4). The magnitude of independent associations was greater in older compared to younger individuals. It was highest for aortic stenosis, followed in descending order by peripheral vascular disease, valvular heart disease, myocardial infarction, chronic kidney disease, heart failure, stroke or transient ischaemic attack, diabetes mellitus, COPD and death.

Figure 7 Adjusted hazard ratios for the ten outcomes, stratified by age



Abbreviations: COPD, chronic obstructive pulmonary disease; TIA, transient ischaemic attack

N.B. Hazard ratios among individuals at higher predicted AF risk compared with individuals at lower predicted AF risk for the 10 outcomes when adjusted for age, sex, ethnicity and the presence of any of the other outcomes at baseline.

5.7 Discussion

In this population-based study, we found that individuals identified for risk-guided AF screening had a similar age and comorbidity profile to individuals who develop AF, and were at increased risk for a range of cardiovascular, renal, and metabolic diseases and death. Over a decade follow-up, more than a quarter of individuals at higher predicted AF risk received a new diagnosis of chronic kidney disease, with heart failure and diabetes mellitus diagnosed in more than one in 10. Although the higher predicted AF risk cohort only made up a fifth of the total population, it constituted 70% of new heart failure diagnoses and 65% of new aortic stenosis diagnoses. The risk of death from any cause was 10-fold greater for individuals at higher predicted AF risk, who accounted for two-thirds of deaths observed during follow-up, and three-quarters of the deaths attributed to cardiovascular or cerebrovascular disease. Adjusted analysis demonstrated that AF risk was associated with incident diseases and death beyond advanced age, which has been the predominant approach hitherto used in AF screening research and advocated in guidelines.[70, 250]

Elevated AF risk portended incident diseases across the cardio-renal-metabolic axis, including when incident AF cases during follow up were excluded. Structural and electric remodelling of the atrium, which increase AF susceptibility, is contributed to by a continuum of unhealthy lifestyle, risk factors and comorbidities;[20] and systemic inflammation, myocardial ischaemia and autonomic dysfunction are implicated in AF genesis.[20] Age, smoking, obesity, inflammatory diseases and hypertension are shared risk factors between AF, vascular disease, aortic stenosis, heart failure, diabetes mellitus and chronic kidney disease.[213-215] Aortic stenosis and heart failure share neurohormonal and proinflammatory pathways with AF which induce myocardial inflammation and fibrosis.[20, 289] Thus, AF is not a disease process in isolation, but a manifestation of multi-system pathology - and AF risk may be considered a precursor stage for an AF ‘syndrome’ of clustered disease states.

Previous studies of AF risk have only investigated for occurrences of AF and stroke during follow-up, reflecting a narrower focus on stroke prevention through early AF detection and treatment.[290] Increasingly it is recognised that the majority of individuals with AF are older and/or have a higher burden of concomitant diseases, cardiometabolic risk factors and unhealthy lifestyle behaviours.[250] Accordingly, lifestyle interventions and management of specific cardiovascular risk factors/comorbidities are recommended in contemporary guidelines for patients with newly diagnosed AF.[250] People identified for risk-guided AF screening share the same characteristics as those with AF, so they may also benefit from equivalent interventions.

Our findings suggest that a risk-guided approach to AF screening may present an opportunity to intervene beyond AF detection and prescription of oral anticoagulation for stroke prophylaxis. The UK National Health Service (NHS) Health Check aims to prevent stroke and cardiovascular disease at a cost £165 million per year,[291] but includes a population comprising only 20% of all strokes and myocardial infarction.[292] By contrast, the higher predicted AF risk subpopulation experience the majority of incident heart failure and vascular events, as well as cardiovascular and cerebrovascular deaths. Based on our findings, risk-guided AF screening would be offered to a subpopulation of 339,000 people aged ≤65 years in the UK, and of this cohort 20% and 15% developed new chronic kidney disease and diabetes mellitus, respectively, over the next 10 years (Figure 3). The median time to event for these outcomes was in excess of three years, so it may be appropriate to offer this ‘targeted’ group comprehensive programmes designed to improve risk factor profiles,[202] as well as early initiation of therapeutics such as sodium-glucose cotransporter 2 inhibitors to reduce the risk of disease progression and cardiovascular morbidity.[293-295] Furthermore, older persons identified for AF screening were more than twice as likely to be diagnosed with aortic stenosis as their lower risk counterparts. Thus this cohort may benefit from targeted early diagnostics, which may not be effective and cost-effective in a purely age-guided AF screening cohort. Elevated natriuretic peptide levels may similarly uncover the presence of underlying multi-systemic or structural cardiac changes, and has been demonstrated to increase the yield of AF screening,[251] but employing wide-scale natriuretic peptide testing would be resource-intensive. Biomarker testing may be more efficiently employed as part of a step-wise approach after risk assessment.

Treatment for individuals at risk of heart failure has been demonstrated to improve outcomes,[296] and accordingly collaborative care for individuals at risk of AF may reduce the subsequent incidence of AF and other adverse events. To prospectively determine

the burden of undiagnosed or under-treated cardiovascular, renal and metabolic conditions and risk factors in individuals identified for risk-guided AF screening, participants enrolled in the FIND-AF pilot implementation study (The BHF Bristol Myers Squibb Cardiovascular Catalyst Award – CC/22/250026) could undergo biomarker and imaging characterisation and cardiologist review, with long-term digital follow-up for the outcomes investigated here.

There are some limitations to our study. First, the CPRD database is routinely-collected, retrospective primary care data and underestimation of incidence of outcomes in this study is possible since there will have been individuals with unrecorded diagnoses. Second, incomplete clinical information is contained in available structured data from EHRs. In particular, echocardiographic reports were unavailable for left ventricular ejection fraction or valve disease severity. Consequently, we could not differentiate types of heart failure, though all are associated with increased risk of death and hospitalisation.[273] We were also unable to evidence the proportion of aortic stenosis cases that were eligible for intervention. However, aortic stenosis is a progressive condition, so we considered an increased risk of clinical diagnosis as important.[274] Third, it is possible that AF risk is associated with increased risk of diseases outside of those we investigated (for example, different cancers). Here we sought to assess association with diseases where there was an underlying pathophysiological rationale and available treatment options,[250] rather than take a data-driven approach. Fourth, our cohort was risk stratified at a single time point, in keeping with how AF screening would be implemented in practice, and we did not address changes in risk profile over time. Fifth, this study included a UK-based cohort and the association between predicted AF risk and incident diseases and death in other geographies may vary. Sixth, individuals for risk-guided AF screening were identified by the FIND-AF risk score, which is scalable in European community-based EHRs and has demonstrated better prediction performance for incident AF than other scalable risk scores.[286] It seems likely that elevated AF risk calculated from other AF risk scores would be associated with incident cardio-renal-metabolic diseases and death but the magnitude of association may vary.

5.8 Conclusions

Individuals identified for risk-guided AF screening are also at higher risk of new diseases across the cardio-renal-metabolic spectrum and death. Participants in risk-guided AF screening may benefit from targeted diagnostics and prevention strategies in excess of ECG monitoring for AF detection.

Chapter 6 Future Innovations in Novel Detection for Atrial Fibrillation (FIND-AF): Pilot study of an electronic health record machine learning algorithm-guided intervention to identify undiagnosed atrial fibrillation

Ramesh Nadarajah, Ali Wahab, Catherine Reynolds, Keerthenan Raveendra, Deborah Askham, Richard Dawson, John Keene, Sagar Shanghavi, Gregory Y H Lip, David C Hogg, Campbell Cowan, Jianhua Wu, Chris P Gale

6.1 Summary of the publication

- This paper presents the protocol for a prospective clinical validation of the FIND-AF algorithm which was developed in Chapter 4.
- This is an interventional, non-randomised, single-arm, open-label study where eligible participants will have their AF risk estimated using FIND-AF and classified as higher or lower risk based on the threshold from the original development and validation paper. All participants will undergo intermittent ECG monitoring for AF detection. The hypothesis is that detection of AF during ECG monitoring will be higher in the group at higher predicted AF risk, compared with the group at lower predicted AF risk.
- The study is possible because of successful funding applications to the British Heart Foundation (grant reference CC/22/250026), Leeds Hospital Charity (grant reference A2002295), and Daiichi Sankyo (NHS Joint Working Partnership, JTW/22/0029).
- Ethical approval for the study was granted by the North West – Greater Manchester South Research Ethics Committee, and the study was approved by the Health research Authority (IRAS project ID: 318197).

6.2 Publication status

- Published September 1 2023
- Open Heart. 2023 Sep 1;10(2):e002447

6.3 Abstract

6.3.1 Introduction

Atrial fibrillation (AF) is associated with a five-fold increased risk of stroke. Oral anticoagulation reduces the risk of stroke, but AF is elusive. A machine learning algorithm (Future Innovations in Novel Detection of Atrial Fibrillation [FIND-AF]) developed to predict incident AF within six months using data in primary care electronic health records (EHRs) could be used to guide AF screening. The objectives of the FIND-AF Pilot study are to determine yields of AF during electrocardiogram (ECG) monitoring across AF risk estimates and establish rates of recruitment and protocol adherence in a remote AF screening pathway.

6.3.2 Methods and analysis

The FIND-AF Pilot is an interventional, non-randomised, single arm, open label study that will recruit 1955 participants aged 30 years or older, without a history of AF and eligible for oral anticoagulation, identified as higher risk and lower risk by the FIND-AF risk score from their primary care EHRs in a 1:1 ratio, to a period of remote ECG monitoring with a Zenicor-ECG device. The primary outcome is AF diagnosis during ECG monitoring, and secondary outcomes include recruitment rates, withdrawal rates, adherence to ECG monitoring, and prescription of oral anticoagulation to participants diagnosed with AF during ECG monitoring.

6.3.3 Ethics and dissemination

The study has ethical approval (the North West – Greater Manchester South Research Ethics Committee reference 23/NW/0180). Findings will be announced at relevant conferences and published in peer-reviewed journals in line with the Funder's open access policy.

6.3.4 Trial registration details

The study has been registered at Clinical Trials.gov (NCT05898165).

6.4 Introduction

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia worldwide, and confers a five-fold increased risk of stroke.[1] It is estimated that up to 35% of AF disease burden remains undiagnosed,[67] and 15% of strokes occur in the context of undiagnosed AF.[248] Early detection of AF may allow the initiation of oral anticoagulation to reduce the risk of AF-related stroke.[35]

Systematic population screening for AF guided by age with or without the presence of additional stroke risk factors with non-invasive electrocardiogram (ECG) devices has been shown to be feasible, increase detection rates for AF compared to routine care, and increase initiation of oral anticoagulation. However, yields of new AF diagnosed are low at between 2.6 and 5.3%.[68, 70, 87, 88] Population screening of 75 and 76 year olds with an intermittent hand-held ECG recorder demonstrated a small net benefit in a composite outcome of ischaemic or haemorrhagic stroke, systemic embolism, bleeding leading to hospitalisation and all-cause death compared to routine care, but was limited by a yield of new of AF of only 3.0%, [70] which hampers clinical and cost-effectiveness.[297]

A targeted screening approach toward a reliably identified subpopulation at higher risk of AF may be more effective and cost-effective. Guiding AF screening by predicted AF risk based on artificial intelligence analysis of ECGs in sinus rhythm has been demonstrated to improve yield of new AF,[298] but ECGs are not widely available in the community which limits the scalability of this approach in the general population. By contrast, a large proportion of the population - 98% in the United Kingdom (UK) - are registered in primary care with a routinely-collected electronic health record (EHR).[202, 252] Thus establishing AF risk from data in primary care EHRs may be a more appropriate approach to guide population AF screening.

A previous randomised clinical trial (RCT) of intermittent non-invasive ECG monitoring compared to routine care for individuals identified as higher risk by a EHR-based risk prediction algorithm (PuLSE-AI) did not find a higher yield of AF detection from ECG monitoring.[299] However that algorithm had a number of shortcomings – it could only be applied to a minority of the population (35%) due to the variables it required for prediction,[52] and it predicted 10 year AF-risk,[194] which may not be relevant to investigating for AF in the short term.

The Future Innovations in Novel Detection of Atrial Fibrillation (FIND-AF) machine learning algorithm was developed and validated in routinely-collected EHRs from over two million UK patients for prediction of incident AF within the next six months. It demonstrates an area under the receiver operating characteristic (AUROC) curve of 0.824 (95% CI 0.814-0.834), with robust prediction performance across both sexes and different ethnic groups.[286] Notably, it was designed to be applicable to 100% of UK primary care EHRs.

FIND-AF was developed and validated in retrospective cohorts of patients where AF was diagnosed during routine care. The objectives of the FIND-AF pilot study are to determine yields of AF during non-invasive ECG monitoring across AF risk estimates and to establish recruitment and protocol adherence rates for a remote AF screening intervention.

6.5 Methods and Analysis

6.5.1 Study design

This publication describes V1.0 of the FIND-AF Pilot study protocol, dated 9th May 2023. The FIND-AF Pilot study is an interventional, non-randomised, single arm, open label study in UK primary care.

6.5.2 Study population

The study will enrol 1955 participants aged ≥30 years with a primary care EHR at General Practices in the NIHR Clinical Research Network Yorkshire and Humber region, who do not have known AF or atrial flutter and are eligible for oral anticoagulation.

Individuals aged ≥30 years are included because this age group were included in the development of the FIND-AF algorithm.[286] Eligibility for oral anticoagulation is determined as men with a CHA_2DS_2-VASc score ≥2 or women with a CHA_2DS_2-VASc score ≥3.[250] We will exclude individuals receiving any form of anticoagulation and those on the palliative care register. Eligibility for oral anticoagulation is required because the aim of the intervention is to diagnose AF in those who would be considered for anticoagulation, thereby minimising the unnecessary care, cost, and anxiety for patients for whom a new diagnosis of AF would not change their management. The inclusion and exclusion criteria are summarised in Table 1.

Table 1 Inclusion and exclusion criteria

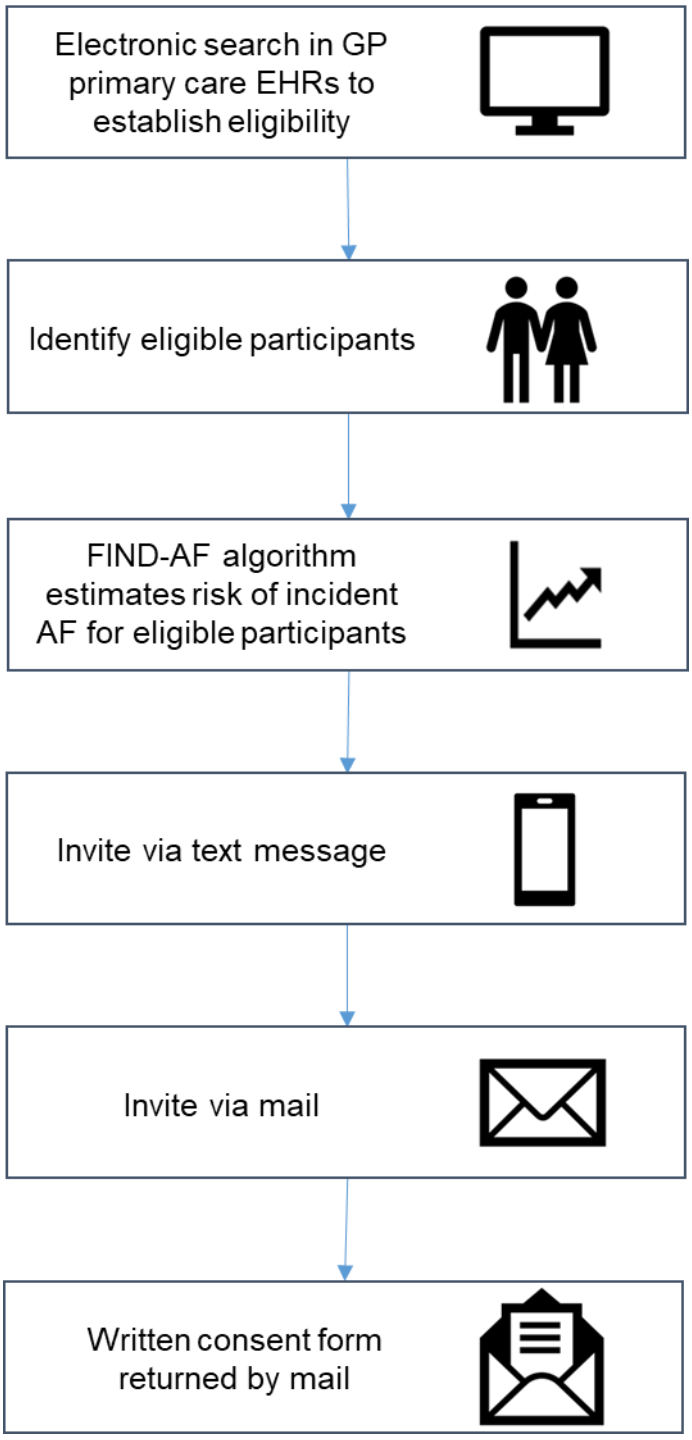
Inclusion criteria
Age at enrolment ≥30 years
Men with CHA_2DS_2-VASC ≥ 2 and women with a CHA_2DS_2-VASC ≥ 3

Exclusion criteria
Known diagnosis of AF or atrial flutter
Currently receiving anticoagulation
On the palliative care register
Unable to give written informed consent for participation in the study
Unable to adhere to the study requirements

Abbreviations: AF, atrial fibrillation; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category

The eligible population will have their AF risk estimated using FIND-AF and classified as higher or lower risk based on the threshold from the original development and validation paper (Figure 1).[286]

Figure 1 Remote consent with eligibility based on information recorded in electronic health records

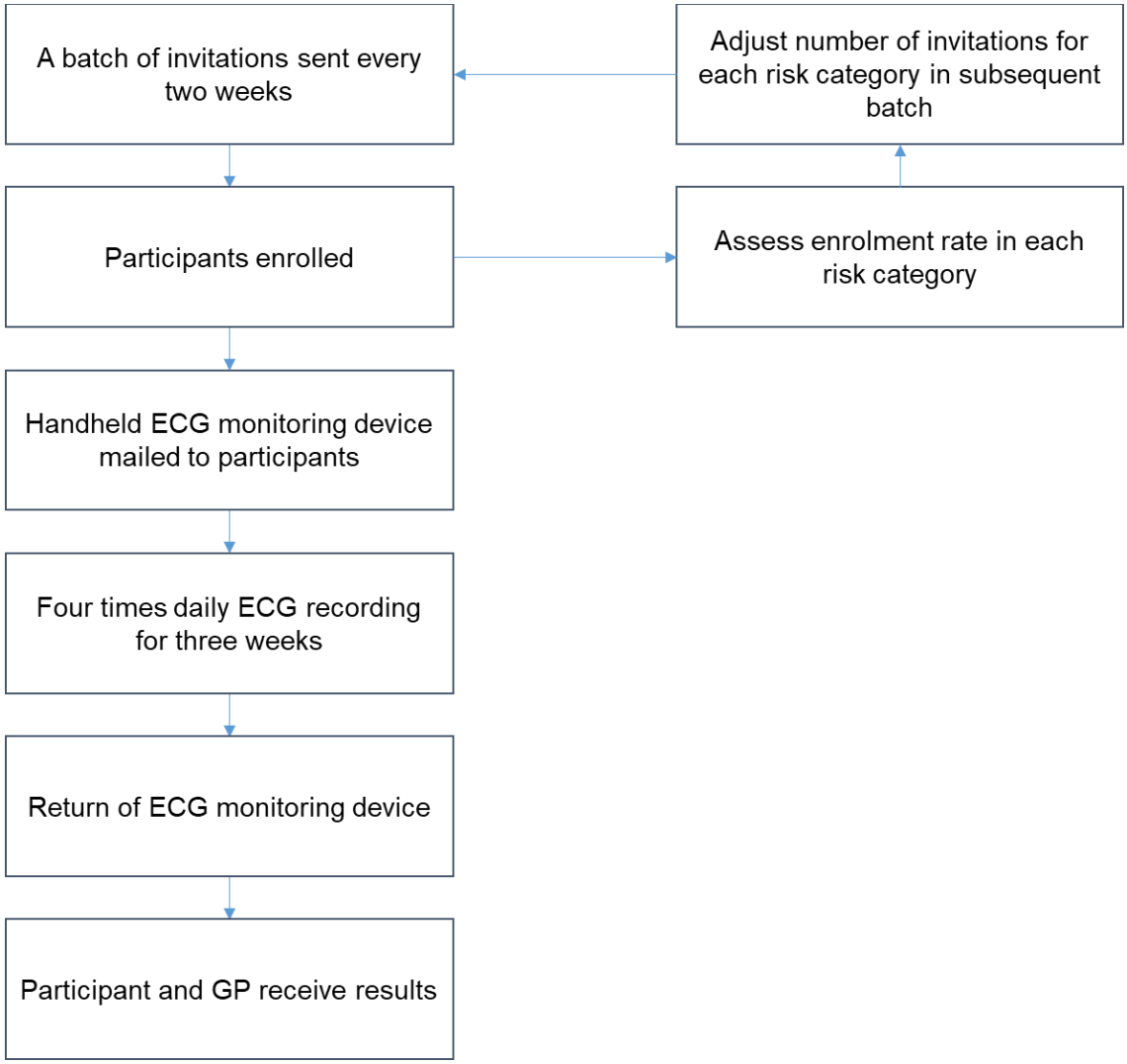


Abbreviations: EHR, electronic health record; FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation; GP, general practice

The study will enrol participants with higher and lower predicted AF risk in a 1:1 ratio. Study invitations will be sent to a random sample of eligible participants in each risk category in batches until the target sample size is reached. As participants are enrolled

in the study, the number of invitations for each risk category will be adjusted in the subsequent batches (Figure 2).

Figure 2 Batch enrolment, study intervention, and follow-up procedures



Abbreviations: ECG, electrocardiogram; GP, general practice

6.5.3 Enrolment method

Eligible participants will be identified by the primary care team via an electronic search of General Practice data (Figure 1). The EHRs for potential participants will be checked to ensure they meet study inclusion and exclusion criteria. Eligibility will be confirmed by a medical practitioner who will ensure it is appropriate to contact the potential participant.

All invitations to targeted screening will occur on site by members of the primary care site team. The invitation process consists of a text message followed by an information pack in the post including a participant information sheet, consent form, data protection leaflet and freepost return envelope (Figure 1). Potential participants will be supplied with a telephone number and email address to contact the study team if they wish to discuss the study and ask any questions they might have prior to providing written consent. All participants will be required to provide written informed consent by returning a completed consent form to the coordinating centre.

6.5.4 Intervention

All participants will undergo non-invasive ECG monitoring (Figure 2). Participants will receive a handheld Zenicor-ECG recorder via the mail with which they will be asked to record their ECG four times daily (morning, noon, afternoon and evening), or whenever they have palpitations, for three weeks.[251, 300] The Zenicor-ECG recorder is a single-lead ECG recorder that is CE-marked as a diagnostic device for AF.[301-303] ECG recordings from the Zenicor-ECG are not displayed on the recorder, but will be automatically and securely transmitted digitally via a 2G mobile network to a central secure Zenicor database.

The Zenicor database has an algorithm that classifies and tags each ECG trace as ‘no tag’, ‘possible AF’ or ‘poor quality’. The algorithm has been tested in 80,149 ECGs and the negative predictive value for ECGs classified as normal is 99.9%.[304] ECGs tagged as ‘possible AF’ will be reviewed by the research team on a weekly basis and a cardiologist will diagnose AF or any other important rhythm disturbances.

Once ECG monitoring reports have been reviewed, a standardised results letter will be sent to the participant and the General Practitioner. Results letters will be sent for all participants, irrespective of whether AF is diagnosed or not. The management of AF will be at the discretion of the GP, allowing doctor and participant to discuss the management strategy, including anticoagulation, independently and in line with how new cases of AF diagnosed in the community are managed in routine clinical practice. A diagnosis of AF does not require immediate action, but if there is a finding meeting the criteria for an emergent event per current clinical practice standards according to the National Institute of Health and Care Excellence, participants and appropriate clinicians will be notified. Actions will be taken following the same procedures as the established clinical workflow.

6.5.5 Baseline characteristics

Baseline participant characteristics will be drawn from their primary care EHRs (Table 2).

Table 2 Participant baseline characteristics

Participant characteristics
Age
Sex
Ethnicity
Medical History
Coronary artery disease
Chronic kidney disease
Heart Failure
Hypertension
Diabetes mellitus
Stroke / transient ischaemic attack
Valvular heart disease
CHA_2DS_2-VASC score
Medications
Aspirin
Angiotensin converting enzyme inhibitor or angiotensin receptor blocker
Beta blocker
Oral anticoagulant
Statin

Abbreviations: CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category

6.5.6 Outcomes

The primary outcome will be a new diagnosis of AF defined as at least one episode of completely irregular rhythm with no organised or regular atrial activity and a duration of 30 seconds on one-lead ECG during the Zenicor monitoring period.[251, 261] Enrolling participants at both higher and lower predicted risk of AF will provide data for the yield of AF across risk estimates, and allow the testing of the hypothesis that individuals identified at higher predicted AF risk are more likely to have AF diagnosed during ECG monitoring than individuals identified as lower predicted AF risk. Such a comparison will allow the assessment of whether predicted AF risk contributes to AF detection yield.

Secondary outcomes include:

- Number (%) of people who consent to participate compared to number of people who are invited.
- Characteristics of those who consent to participate and do not consent to participate.
- Number (%) of people who consent to participate but subsequently withdraw consent or decline ECG monitoring.
- Characteristics of those who participate and those that withdraw.
- Of those who conduct ECG monitoring, the number (%) of participants who record less than 50% of the stipulated amount of ECG recordings.
- Of those who conduct ECG monitoring, the day of first detection of AF.
- Number (%) of participants who are diagnosed with other arrhythmias during ECG monitoring in participants.
- Number (%) of participants who are diagnosed with AF during ECG monitoring who then receive a prescription of oral anticoagulation within 6 months.
- Number (%) of diagnoses of AF during routine practice outside of ECG monitoring (presence of an AF diagnostic code in their primary care EHR at 6 months after enrolment).

6.5.7 Sample size

For 977 participants completing the study protocol in each of the higher and lower risk groups, assuming 1.5% of the participants in the lower risk group have newly diagnosed AF,[67, 88] we will have 80% power to detect a significant difference if 6% of the higher risk group have newly diagnosed AF.[298, 299]

6.5.8 Statistical analysis

We will calculate the incidence rate ratio of AF detection during ECG monitoring between higher predicted AF risk and lower predicted AF risk participants. We will calculate positive predictive value, negative predictive value, sensitivity, specificity and area under the receiver operating characteristic curve (AUROC) for FIND-AF. We will explore other thresholds and report corresponding performance measures, which will inform whether the FIND-AF threshold to classify higher and lower risk should be altered for optimal yield.

6.5.9 Patient and public involvement

The FIND-AF patient and public involvement (PPI) group co-designed the study and co-drafted the consent forms and participant information sheets. Importantly, they designed the multi-modal invitation (text followed by letter) to screening as they concluded that the usual invitation approach, a letter alone, may lead to poorer participation from people of minority ethnicities and lower socioeconomic classifications.[70] The Arrhythmia Alliance will support dissemination activities.

6.6 Limitations

The FIND-AF Pilot is not a RCT. The Zenicor-ECG device is the only AF detection device that will be used in the study. Other studies have used a skin patch that can monitor the ECG rhythm continuously, for between 14 and 30 days.[298, 305] As the Zenicor device records a 30-second ECG and is being used only four times a day, it is possible that some cases of AF that would be diagnosed using continuous monitoring will be missed when using an intermittent monitoring approach. However the Zenicor-ECG was used for AF detection in the STROKESTOP RCT where treatment of screen-detected AF was associated with a 4% reduction in combined endpoint of ischaemic or haemorrhagic stroke, systemic embolism, bleeding leading to hospitalisation, and all-cause death.[70] Accordingly, we are reassured that treatment of AF detected during this study is clinically appropriate, whereas the optimal threshold of AF duration to be treated in continuous monitoring screen-detected AF is still to be established.[72] This pilot study is examining AF detection through targeted AF screening, but the effect of AF screening on clinical outcomes is subject to a conflicting evidence base,[92, 306] with further RCTs ongoing.[71]

6.7 Ethics and dissemination

The study will be performed in compliance with the articles of the Declaration of Helsinki (revised in October 2013). The study was approved by the North West – Greater Manchester South Research Ethics Committee, and the study was approved by the Health research Authority (IRAS project ID: 318197), and registered on Clinical Trials.gov (NCT05898165). Study results will be disseminated at relevant conferences and published in peer-reviewed journals. Authorship will be decided according to ICMJE guidelines as to qualifying contributions, and the primary results manuscript jointly drafted by the Co-Chief Investigators and the trial methodologists before circulating to remaining co-authors.

6.8 Discussion

Primary care EHRs provide a scalable approach to guide AF screening across a nation. Hitherto, AF screening interventions with non-invasive ECG devices have been hindered by low yields of newly detected AF. This pilot study will provide data for whether higher predicted AF risk identified by the FIND-AF machine learning algorithm using primary care EHRs is associated with higher yields of AF during ECG monitoring amongst a population eligible for oral anticoagulation.

Chapter 7 Risk calculator for incident atrial fibrillation across a range of prediction horizons

Jianhua Wu, Ramesh Nadarajah, Yoko M Nakao, Kazuhiro Nakao, Ronen Arbel, Moti Haim, Doron Zahger, Gregory Y H Lip, Campbell Cowan, Chris P Gale

*Jianhua Wu and Ramesh Nadarajah are joint first authors

7.1 Summary of the publication

- This analysis was performed using the methodology presented in Chapter 3 to develop and internally validate a parsimonious prediction model for incident AF using traditional regression and clinically-recognised risk factors for AF.
- This study developed models for incident AF at different prediction horizons using the same variables, so that the models could be used to predict both short and long-term AF risk and guide AF screening or primary prevention.
- This study showed that the association between known risk factors for AF and incident AF varies with different lengths of prediction horizon.
- The study found that the simple risk score showed stronger prediction performance at all prediction horizons than the the C_2HEST and CHA_2DS_2-VASc scores, and had excellent calibration performance.

7.2 Publication status

- Submitted, under review with JAMA Network Open

7.3 Abstract

7.3.1 Importance

The increasing burden of atrial fibrillation emphasizes the need to identify high-risk individuals for enrolment in clinical trials of AF screening and primary prevention.

7.3.2 Objective

To develop prediction models to identify individuals at high-risk of AF across prediction horizons from 6-months to 10-years.

7.3.3 Design

Retrospective cohort study between Jan 2, 1998 and Nov 30, 2018.

7.3.4 Setting

Anonymized secondary care-linked primary care electronic health records from Clinical Practice Research Datalink from a subset of the UK population.

7.3.5 Participants

Individuals aged ≥30 years without known AF, randomly divided into derivation (80%) and validation (20%) datasets.

7.3.6 Exposures

Demographics and clinical factors.

7.3.7 Main outcome and measures

AF and/or atrial flutter. Models were derived using logistic regression from known AF risk factors for incident AF in prediction periods of 6 months, 1-year, 2-years, 5-years and 10-years. Performance was evaluated using internal bootstrap validation with 200 samples, and compared against the CHA_2DS_2-VASc and C_2HEST scores.

7.3.8 Results

Of 2 081 139 individuals in the cohort (1 664 911 in the training dataset, 416 228 in testing dataset), the mean age was 49.9 (SD 15.4), 50.7% were women, and 86.7% were white. New cases of AF were 7 386 (0.4%) within 6 months, 15 349 (0.7%) in 1 year, 38 487 (1.8%) in 5 years, and 79 997 (3.8%) by 10 years. Valvular heart disease and heart failure were the strongest predictors, and association of hypertension with AF increased at longer prediction horizons. The optimal risk models incorporated age, sex, ethnicity and eight comorbidities. The models demonstrated good-to-excellent

discrimination and strong calibration across prediction horizons (AUROC, 95% CI, calibration slope: 6-months, 0.803, 0.789-0.821, 0.952; 1-year, 0.807, 0.794-0.819, 0.962; 2-years, 0.815, 0.807-0.823, 0.973; 5-years, 0.807, 0.803-0.812, 1.000; 10-years 0.780, 0.777-0.784, 1.010), and superior to the CHA_2DS_2-VASc and C_2HEST scores. The models are available as a web-based FIND-AF calculator.

7.3.9 Conclusions and relevance

The FIND-AF models demonstrate high discrimination and calibration across short- and long-term prediction horizons in 2 million individuals. Their utility to inform trial enrolment and clinical decisions for AF screening and primary prevention requires further study.

7.4 Introduction

Atrial fibrillation (AF) is a global epidemic affecting more than 37 million people worldwide,[9] and confers an increased risk of stroke, heart failure, cognitive decline and death.[1] It is estimated that up to 35% of disease burden remains undiagnosed,[67] and 15% of strokes occur in the context of undiagnosed AF.[248] Accordingly broader population-wide screening for AF has been the subject of randomised clinical trials (RCTs), but these are limited by low yields for newly detected AF.[67, 68, 88] Moreover, the number of new cases of AF diagnosed each year has risen by 72% over the last two decades,[307] which emphasises the need to consider primary prevention strategies.[38] However difficulties in identifying a group at sufficiently high risk impedes the conduct of primary prevention trials for AF.[151]

Comprehensive risk assessment will more reliably identify individuals at-risk of AF compared with any single risk factor,[183] but existing AF prediction models only predict either short- or long-term incident AF.[183, 197, 286] How multiple risk factors interact to confer AF risk may vary over different prediction horizons, emphasising the need to derive independent models for prediction of short- and long-term incident AF to be able optimise the targeting of AF screening and primary prevention interventions.

Thus, using a nationwide cohort we sought to i) quantify how the association of AF risk factors with incident AF varies at different prediction horizons; ii) develop models to estimate risk of incident AF in the general population at different prediction horizons utilising readily available data routinely recorded in primary care; iii) compare the new

prediction models against existing AF risk prediction models; and iv) develop a web-based tool (Future Innovations in Novel Detection of Atrial Fibrillation [FIND-AF] calculator) to enable individual-level estimation of AF risk over multiple prediction horizons.

7.5 Methods

7.5.1 Study design and population

We used primary care EHRs from the UK Clinical Practice Research Datalink (CPRD)-GOLD dataset. CPRD is one of the largest databases of longitudinal medical records from primary care worldwide and contains anonymised patient data from approximately 7% of the UK population,[202] and has been used to develop prediction models for AF.[286]

We included all adults registered at practices within CPRD who were ≥30 years of age at entry with no prior history of AF and at least one-year follow-up between January 2, 1998 and November 30, 2018. Individuals were censored to a diagnosis of AF (or atrial flutter [AFl], since it has similar thromboembolic risk and anticoagulation guidelines),[1] withdrawal from CPRD, or the prediction horizon over which the model was developed for (6-months through to 10 years), whichever came first. Individuals were randomly split 4:1 to establish a derivation dataset (80%) and validation dataset (20%) using the Mersenne twister pseudorandom number generator.

The sample size was calculated to be sufficient to develop prediction models (Chapter 3, Section 3.6.5). We followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline and the CODE-EHR best-practice framework for using structured electronic healthcare records in clinical research.[276, 278]

7.5.2 Outcomes

The primary outcome was a diagnosis of AF or AFl. All individuals in the CPRD dataset were linked to Hospital Episode Statistics (HES) Admitted Patient Care (APC) records to obtain comprehensive coverage of AF cases diagnosed in secondary care. Diagnoses of AF or AFl in primary care were identified using Read codes in CPRD and secondary care with the tenth revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes in HES-APC (Chapter 4, Table

3). Diagnostic coding for AF in CPRD has been shown to be consistent and valid, with a positive predictive value of 98%.[280]

7.5.3 Candidate predictors

To prioritise scalability of the prediction models, we restricted candidate variables to age, sex, comorbidities and ethnicity. We did not include observations and laboratory results were not included (such as systolic blood pressure, body mass index, and lipid profile) as these are often missing in primary care medical records.[252] Candidate predictor variables were selected a priori from known risk factors/associations as described in ESC guidelines: age, sex, heart failure, hypertension, diabetes mellitus, stroke/transient ischaemic attack (TIA)/systemic embolism, vascular disease (angina, myocardial infarction, peripheral arterial disease), chronic obstructive pulmonary disease (COPD), chronic kidney disease (CKD), valvular heart disease, rheumatoid arthritis, hyperthyroidism, and ethnicity.[1] Variables with low prevalence (< 2.0%) were not included as predictors in our model (obstructive sleep apnoea). Dyslipidaemia is associated with AF through vascular diseases, so it was excluded from this model. Ethnicity was treated as a dichotomous variable because non-white individuals comprised only a small proportion of the population (individuals with unrecorded or unknown ethnicity were included in the non-white group).[290] Diagnostic code lists only included the primary care coding system (Read codes), ensuring that only information readily available at point of care in a community setting could be incorporated within the models. Our entire analytical cohort had no missing data for any of the predictor variables and the models could be applied to all records.

7.5.4 Risk score derivation

Logistic regression models were fitted with a backward elimination approach to retain predictors of incident AF within distinct prediction horizons of 6 months, 1 year, 2 years, 5 years and 10 years to optimise model fit as assessed by minimising the Akaike information criterion and Bayesian information criterion. Age (the only continuous variables) was modelled with restricted cubic splines with 4 knots.

7.5.5 Statistical analyses

The baseline characteristics are summarised for the derivation and validation datasets. Continuous variables were reported as mean ± standard deviation (SD). Categorical variables were reported as frequencies with corresponding percentages.

The best fitting models from the derivation set were applied to the validation set by using the parameter coefficients obtained from the derivation sample to derive a weighted score for each individual. Model performance was determined using the full holdout validation set with internal bootstrap validation with 200 samples. We compared the best fitting prediction models to the CHA_2DS_2-VASc and C_2HEST scores over each prediction horizon. Both the CHA_2DS_2-VASc and C_2HEST scores have been tested in general population cohorts for AF risk prediction,[197] but the former was originally developed to predict stroke risk in individuals with AF, while the latter was developed as a simple clinical score for incident AF in a population without structural heart disease.[197, 198] For each of the FIND-AF prediction models and the CHA_2DS_2-VASc and C_2HEST scores, dichotomous covariates were assumed to persist and therefore carried forward if present during or prior to the ascertainment period, and were assumed to be absent if they were not present at the beginning of follow-up. The presence of variables was defined based on the presence of Read codes. The 6-month, 1 year, 2 year, 5 year and 10 year risk of AF based on the C_2HEST and CHA_2DS_2-VASc scores was estimated by using the baseline hazard and mean covariate estimates from the validation sample.

The area under the receiver operating characteristic curve (AUROC) was used to evaluate predictive ability (concordance index) with 95% confidence intervals (CIs) calculated using the DeLong method. AUROC values of <0.60, 0.60-0.70, 0.70-0.80, and >0.80 were defined a priori as inadequate, adequate, good and excellent based on prior publications.[195] Calibration was assessed by plotting predicted AF risk against observed AF incidence and the calibration slope. We calculated the Brier score, a measure of both discrimination and calibration, by taking the mean squared difference between predicted probabilities and the observed outcome. To assess the clinical impact of utilising the risk scores we conducted a decision curve analysis. We also investigated the performance of each risk score within relevant subgroups defined by sex (men vs women) and ethnicity (White vs. Non-White ethnic minorities), and conducted a sensitivity analysis where individuals with unrecorded or unknown ethnicity were excluded. We used R version 4·1·0 for all analyses.

Other AF risk prediction models were not selected for comparison, as they were not considered to implementable at scale in the community (Table 1).

Table 1 Prediction models for incident atrial fibrillation in the general population not selected for comparison in this study due to requirement for observations, examination findings or laboratory measurements, which may not be available at scale in the community

Prediction model	**Study**	**Variable frequently missing/unavailable in community records**			
		Observations	**ECG/echocardiogram parameters**	**Examination findings**	**Laboratory measurements**
ARIC-AF	Chamberlain 2011	Height, SBP	LAE, LVH	Significant murmur	
CHARGE-AF	Alonso 2013	Height, weight, SBP, DBP			
FHS-AF	Schnabel 2009	BMI, SBP	PR interval	Significant murmur	
$HARMS_2$-AF	Segan 2023	BMI, alcohol consumption in units, smoking status			
Maccabi Health System	Aronson 2018	BMI, SBP			
Mayo	Linker 2018			Significant murmur	
PuLSE-AI	Hill 2019	Height, weight, BMI, SBP, DBP			
PREVEND	Rienstra 2016	HR, BMI, height, weight, SBP, DBP	PR interval		eGFR, urine albumin secretion, serum lipids
Seirei	Hamada 2019	Waist circumference, DBP, HR		Significant murmur	
Suita	Kokubo 2017	SBP, BMI		Significant murmur	Serum lipids

Woman's health study	Everett 2013	Height, weight, SBP			

Abbreviations: AF, Atrial Fibrillation; AI, artificial intelligence; ARIC-AF, Atherosclerosis Risk In Communities score for Atrial Fibrillation; BMI, body mass index CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology; DBP, diastolic blood pressure; ECG, electrocardiogram; eGFR, estimated glomerular filtration rate; FHS-AF, Framingham Heart Study score for Atrial Fibrillation; HARMS2-AF, Hypertension, Age, Raised BMI, Male sex, Sleep apnoea, Smoking, Alcohol-AF score; HR, heart rate; LAE, left atrial enlargement; LVH, left ventricular hypertophy; MHS, Maccabi Healthcare Services; PREVEND, Prevention of Renal and Vascular End-stage Disease; SBP, systolic blood pressure

7.5.6 Clinical applicability

We developed a web-based FIND-AF calculator incorporating the prediction models for each prediction horizon (6-months, 1-year, 2-years, 5-years, 10 years) using RShiny. We also generated a FIND-AF risk score for 10-year AF risk by assigning points to each variable proportional to its regression coefficients rounded to the nearest integer. For age, a LOESS smoothing curve was fitted to incident AF against age to identify cut points of age that are associated with remarkable changes in the risk of incident AF.[308] These cut points divide age into several intervals, and a score was assigned to each interval.

7.6 Results

7.6.1 Patient population

There were 2 081 139 individuals registered in our UK primary care cohort (1 664 911 in the training dataset, 416 228 in testing dataset), with average age 49.9 (SD 15.4), 50.7% women, and 86.7% white. Baseline characteristics and clinical outcomes were similar in the development and validation datasets (Chapter 4, Table 6). 7 386 (0.4%) developed AF within 6 months, 15 349 (0.7%) in 1 year, 38 487 (1.8%) in 5 years, and 79 997 (3.8%) in 10 years.

7.6.2 Model development

The association of each candidate variable with incident AF in each prediction horizon are reported in Table 2. The strongest predictors in each prediction horizon were valvular heart disease and heart failure. The magnitude of association between predictors and AF decreased or remained the same as the prediction horizon lengthened, except for hypertension where the magnitude of the association increased (odds ratio: 6 months, 1.38, 95% CI 1.27-1.49; 10-years, 1.58, 95% CI 1.55-1.61). To ensure parsimonious final models, of the candidate predictors CKD and rheumatoid arthritis were excluded from the model owing to little association with incident AF in most of the prediction horizons. Accordingly, the final models included age, sex, ethnicity (white vs non-white), diabetes mellitus, heart failure, hypertension, stroke/TIA/systemic embolism, vascular disease, COPD, valvular heart disease and hyperthyroidism with varying coefficients across prediction horizons.

Table 2 Multivariate associations of candidate predictors with risk of incident atrial fibrillation in the derivation sample (n = 1 664 911) according to prediction horizon

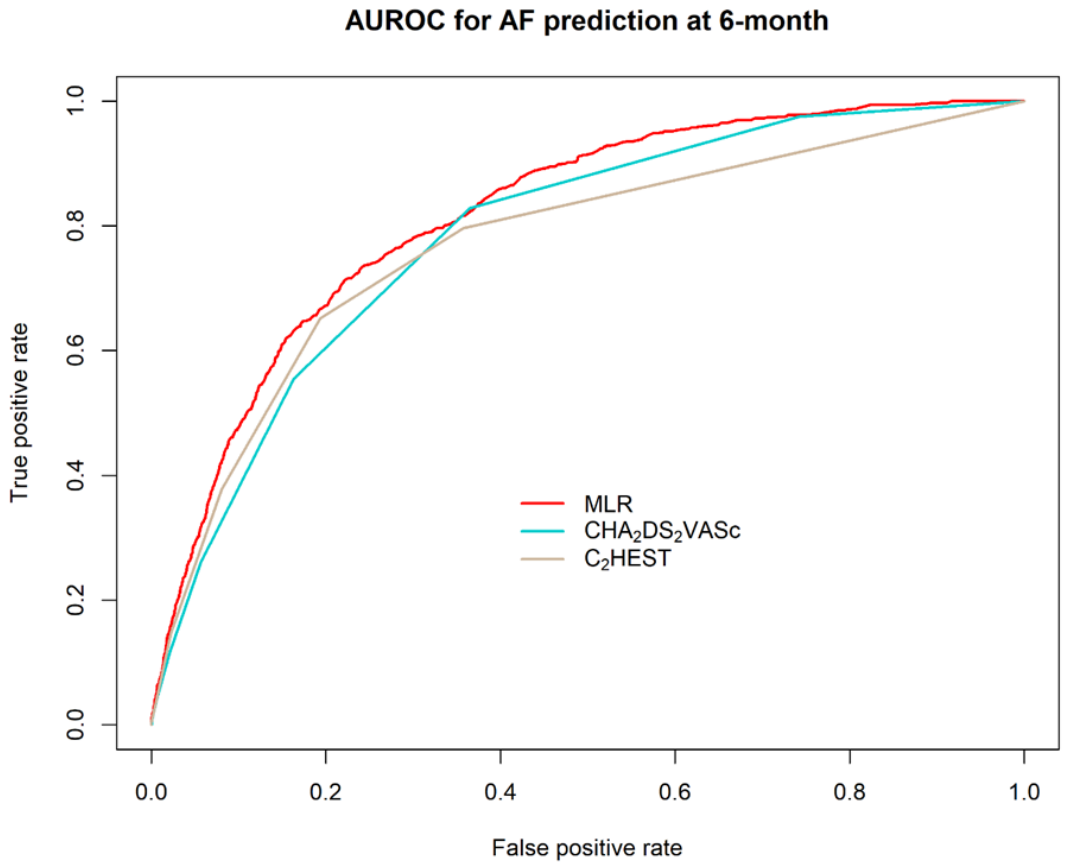
Prediction horizon	**6-months**	**1-year**	**2-years**	**5-years**	**10-years**
Predictors	**Odds ratio (95% CI)**	**Odds ratio (95% CI)**	**Odds ratio (95% CI)**	**Odds ratio (95% CI)**	**Odds ratio (95% CI)**
Age per year	1.08 (1.08-1.09)	1.09 (1.08-1.09)	1.09 (1.09-1.09)	1.09 (1.08-1.09)	1.08 (1.08-1.08)
Sex (woman vs man)	0.61 (0.56-0.65)	0.61 (0.58-0.65)	0.62 (0.60-0.65)	0.62 (0.61-0.64)	0.63 (0.62-0.64)
Ethnicity (won-White vs White)	0.47 (0.42-0.53)	0.46 (0.42-0.50)	0.44 (0.41-0.46)	0.39 (0.37-0.40)	0.30 (0.29-0.31)
Diabetes mellitus	1.31 (1.17-1.48)	1.31 (1.20-1.42)	1.32 (1.24-1.40)	1.30 (1.25-1.35)	1.20 (1.17-1.24)
Heart failure	2.37 (2.08-2.70)	2.13 (1.93-2.35)	2.27 (2.11-2.44)	2.14 (2.03-2.26)	1.82 (1.74-1.91)
Hypertension	1.38 (1.27-1.49)	1.45 (1.37-1.54)	1.46 (1.40-1.52)	1.54 (1.50-1.58)	1.58 (1.55-1.61)
Stroke / TIA / systemic embolism	1.49 (1.34-1.67)	1.43 (1.31-1.55)	1.40 (1.32-1.49)	1.29 (1.23-1.34)	1.12 (1.08-1.15)
Vascular disease (angina, myocardial infarction, peripheral arterial disease)	1.39 (1.26-1.53)	1.48 (1.38-1.59)	1.45 (1.38-1.53)	1.51 (1.46-1.56)	1.54 (1.50-1.58)
Chronic obstructive pulmonary disease	1.25 (1.07-1.46)	1.39 (1.25-1.56)	1.47 (1.36-1.59)	1.50 (1.43-1.58)	1.36 (1.30-1.42)
Valvular heart disease	3.62 (3.09-4.24)	3.23 (2.85-3.66)	3.23 (2.94-3.54)	3.24 (3.02-3.48)	3.19 (3.01-3.39)
Chronic kidney disease	0.82 (0.69-0.97)	1.02 (0.91-1.14)	1.04 (0.96-1.13)	0.94 (0.89-1.00)	0.76 (0.72-0.79)
Rheumatoid arthritis	1.07 (0.89-1.28)	1.16 (1.02-1.31)	1.10 (1.00-1.20)	1.10 (1.04-1.17)	1.12 (1.07-1.17)
Hyperthyroidism	1.64 (1.28-2.10)	1.67 (1.39-2.00)	1.60 (1.40-1.83)	1.45 (1.32-1.59)	1.34 (1.25-1.44)

Abbreviations: TIA, transient ischaemic attack

7.6.3 Model validation

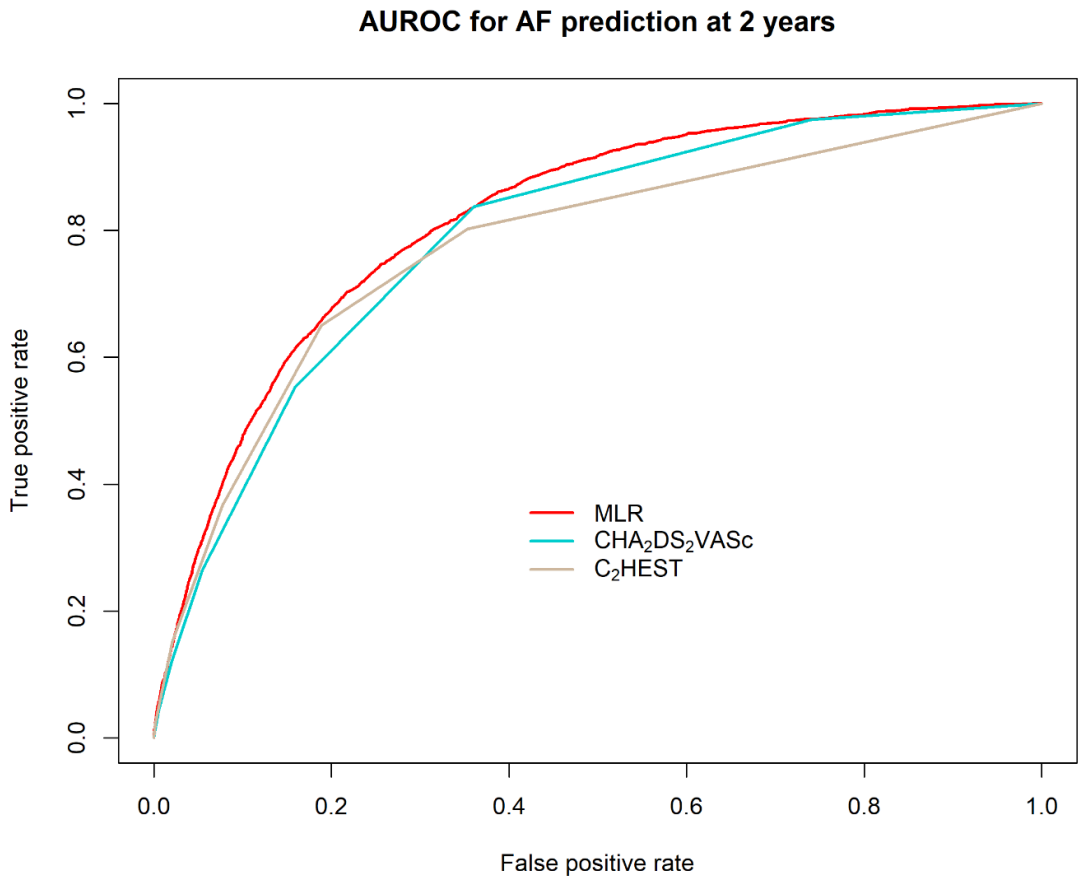
The FIND-AF prediction models had good-to-excellent discrimination in the internal validation cohorts across each prediction horizon (Table 3, Figure 1-4), with the highest performance at 2 years (AUROC 0.815, 95% CI 0.807-0.823) and the lowest performance at 10 years (AUROC 0.780, 95% CI 0.777-0.784).

Figure 1 Receiver operating characteristic curves for FIND-AF multivariable logistic regression model, CHA_2DS_2-VASc, and C_2HEST risk scores for prediction of incident AF in a 6-month prediction horizon



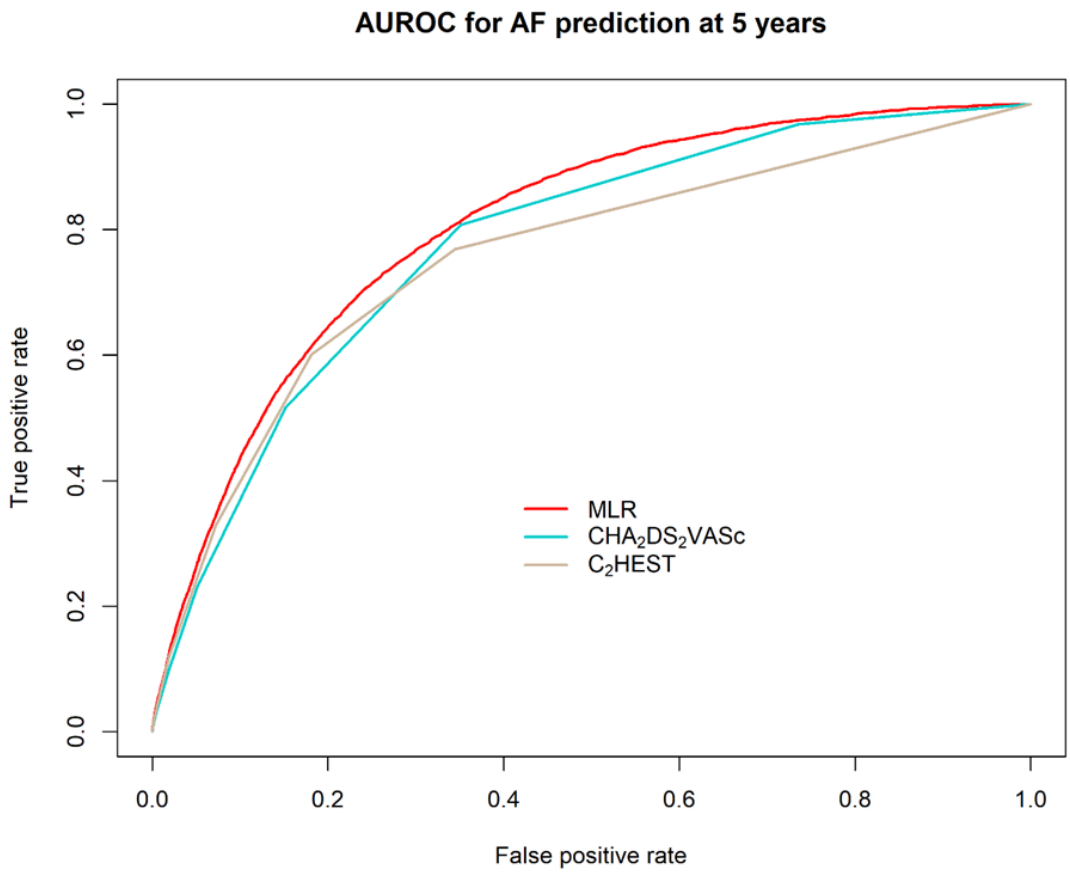
Abbreviations: AF, atrial fibrillation; AUROC, area under the receiver operating characteristic curve; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); CI, confidence interval. MLR, multivariable logistic regression

Figure 2 Receiver operating characteristic curves for FIND-AF, CHA_2DS_2-VASc, and C_2HEST risk scores for prediction of incident AF in a 2-year prediction horizon



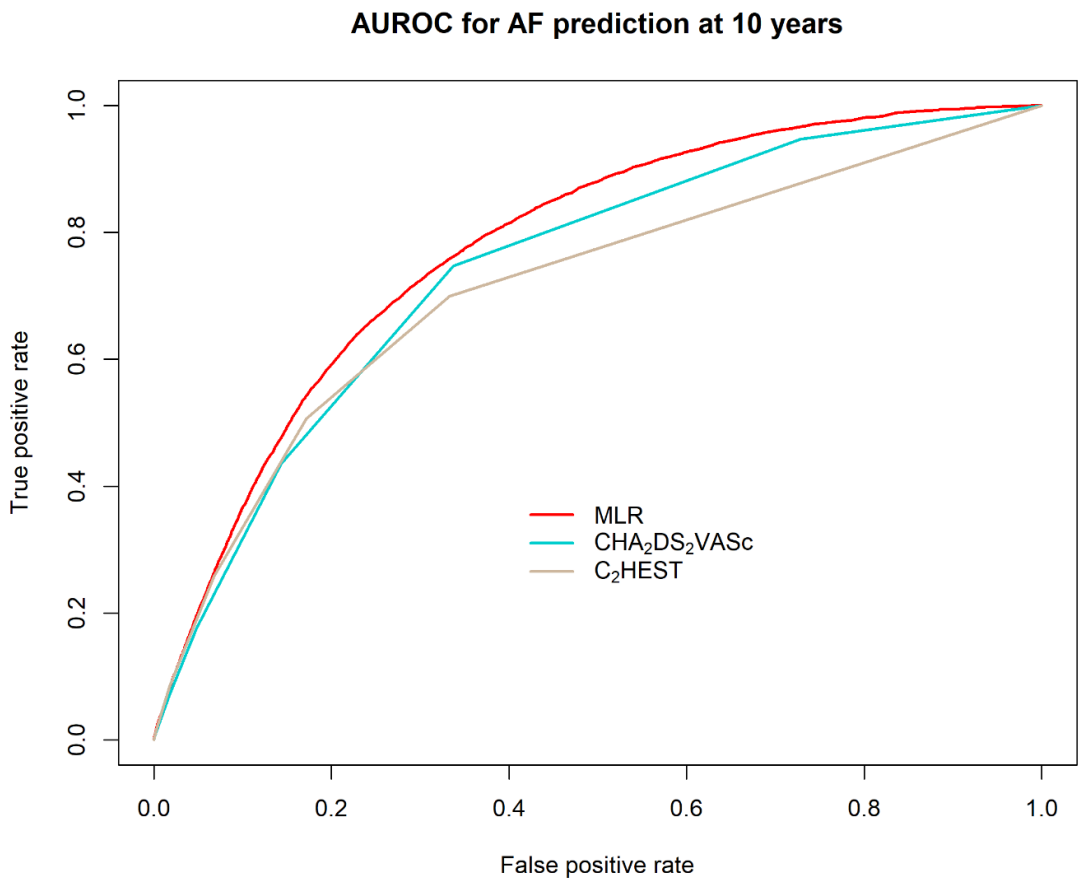
Abbreviations: AF, atrial fibrillation; AUROC, area under the receiver operating characteristic curve; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); CI, confidence interval. MLR, multivariable logistic regression

Figure 3 Receiver operating characteristic curves for FIND-AF, CHA_2DS_2-VASc, and C_2HEST risk scores for prediction of incident AF in a 5-year prediction horizon



Abbreviations: AF, atrial fibrillation; AUROC, area under the receiver operating characteristic curve; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); CI, confidence interval. MLR, multivariable logistic regression

Figure 4 Receiver operating characteristic curves for FIND-AF, CHA_2DS_2-VASc, and C_2HEST risk scores for prediction of incident AF in a 10-year prediction horizon



Abbreviations: AF, atrial fibrillation; AUROC, area under the receiver operating characteristic curve; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); CI, confidence interval. MLR, multivariable logistic regression

Table 3 Prediction performance of the FIND-AF, CHA_2DS_2-VASc, and C_2HEST risk scores for 6-months, 1-year, 2-years, 5-years, and 10-years incident AF

Prediction horizon	**6-months**	**1-year**	**2-years**	**5-years**	**10-years**
FIND-AF					
AUROC (95% CI)	0.803 (0.789 - 0.821)	0.807 (0.794 - 0.819)	0.815 (0.807 - 0.823)	0.807 (0.803 - 0.812)	0.780 (0.777 - 0.784)
Calibration slope* (95% CI)	0.952 (0.899 - 1.017)	0.962 (0.910 - 1.014)	0.973 (0.938 - 1.003)	1.000 (0.976 - 1.021)	1.010 (0.992 - 1.027)
Brier score	0.004 (0.003 - 0.004)	0.007 (0.007 - 0.007)	0.014 (0.014 - 0.015)	0.033 (0.032 - 0.034)	0.065 (0.064 - 0.066)
CHA_2DS_2-VASc					
AUROC (95% CI)	0.781 (0.758 - 0.802)	0.782 (0.769 - 0.794)	0.790 (0.781 - 0.800)	0.781 (0.776 - 0.789)	0.749 (0.745 - 0.754)
Calibration slope* (95% CI)	0.875 (0.804 - 0.941)	0.860 (0.819 - 0.901)	0.882 (0.849 - 0.916)	0.898 (0.879 - 0.922)	0.885 (0.869 - 0.900)
Brier score	0.002 (0.002 - 0.002)	0.004 (0.004 - 0.005)	0.009 (0.009 - 0.009)	0.021 (0.021 - 0.022)	0.043 (0.042 - 0.044)
C_2HEST					
AUROC (95% CI)	0.757 (0.739 - 0.775)	0.753 (0.737 - 0.767)	0.765 (0.755 - 0.775)	0.749 (0.743 - 0.756)	0.710 (0.706 - 0.715)
Calibration slope* (95% CI)	0.760 (0.703 - 0.823)	0.742 (0.694 - 0.789)	0.765 (0.733 - 0.795)	0.756 (0.735 - 0.778)	0.712 (0.696 - 0.727)
Brier score	0.003 (0.003 - 0.004)	0.007 (0.006 - 0.007)	0.013 (0.013 - 0.014)	0.031 (0.03 - 0.032)	0.062 (0.061 - 0.063)

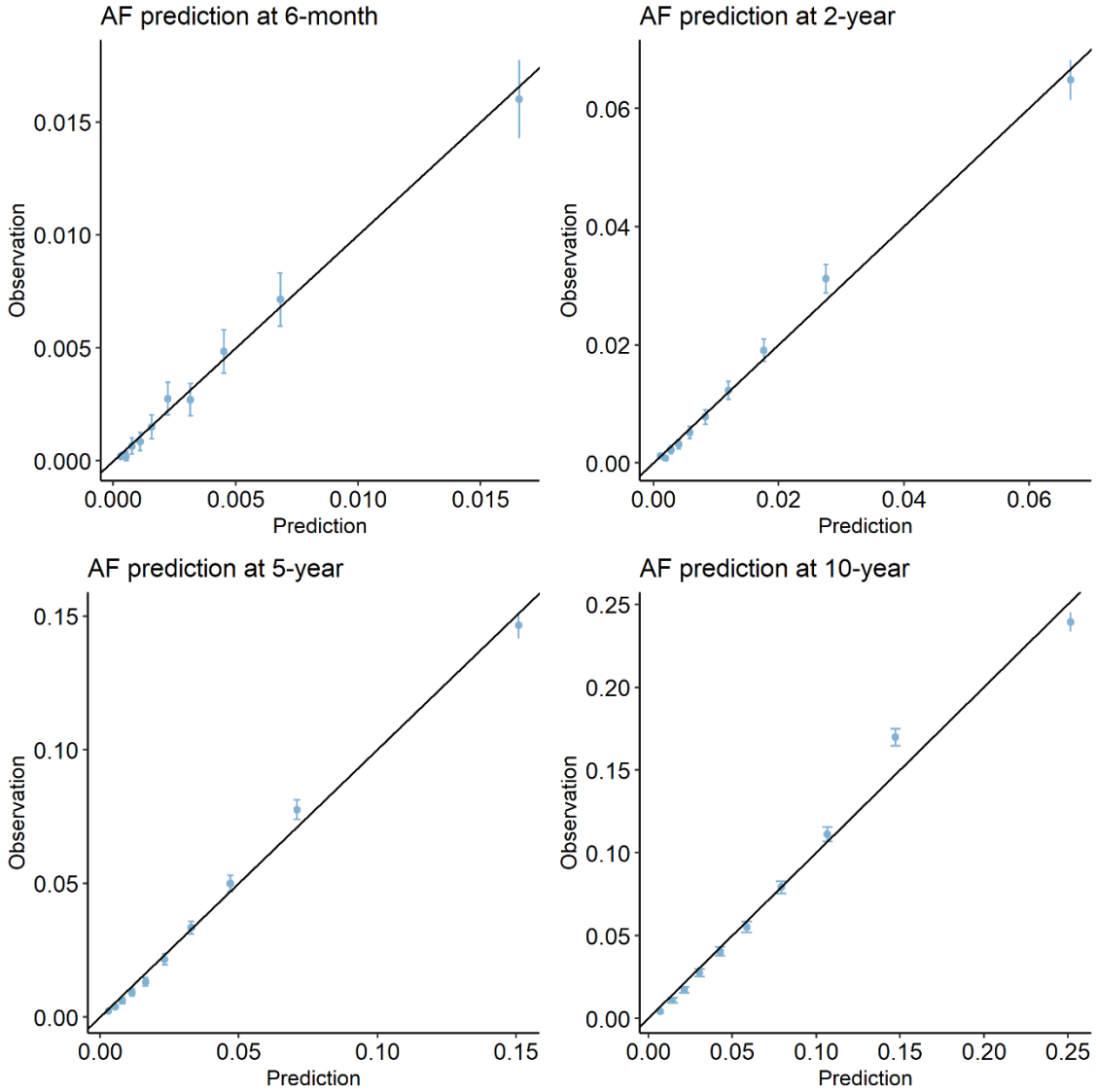
Abbreviations: AUROC, area under receiver operating characteristic; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; C_2HEST, Coronary artery disease /

Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); CI, confidence interval; FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation

N.B. *calibration slope was derived from linear regression models by forcing the intercept through origin (0,0).

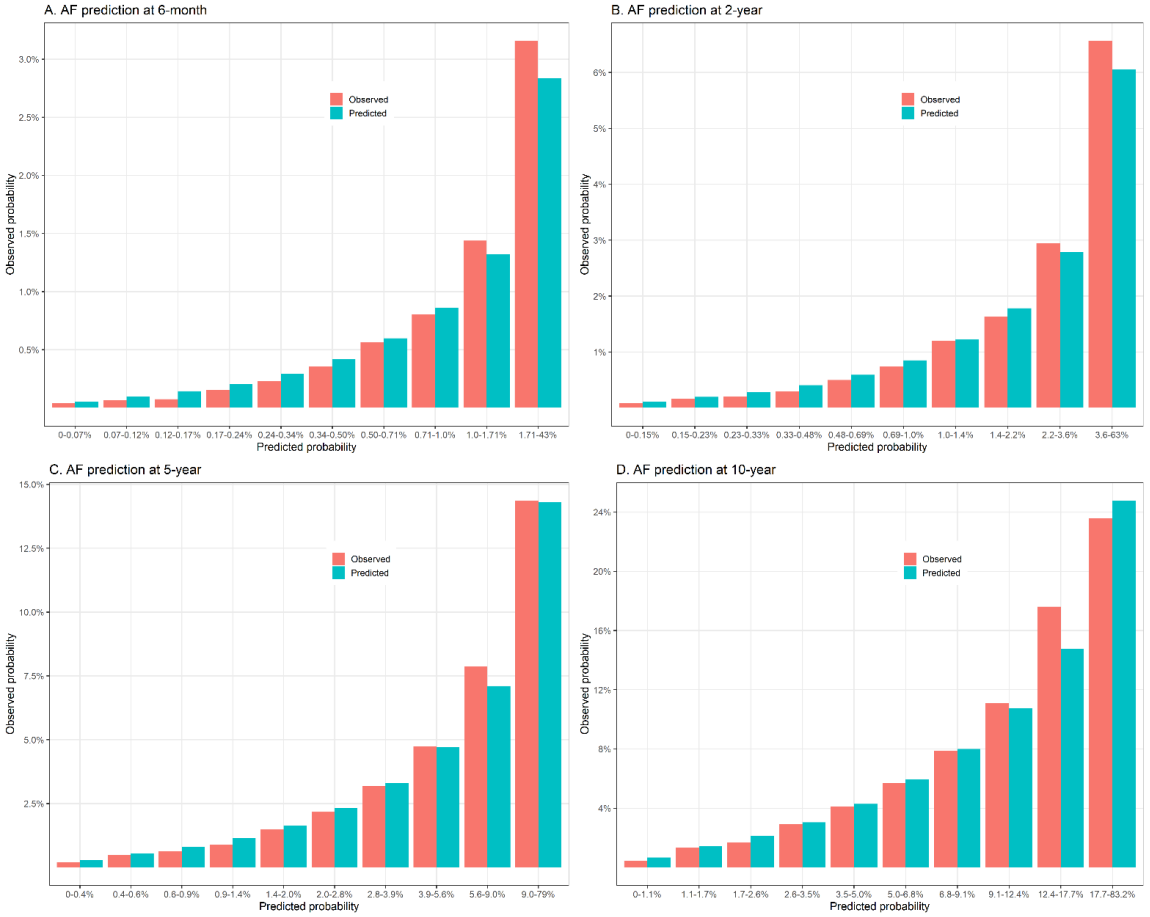
The FIND-AF prediction models were well calibrated across prediction horizons (calibration slope ranging from 0.952 to 1.010 across time horizons; Table 3, Figure 5-6).

Figure 5 Prediction horizon-specific calibration plots in validation sample for the FIND-AF prediction models. A) 6-months, B) 2-years, C) 5-years, D) 10-years



Abbreviations: AF, atrial fibrillation; FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation

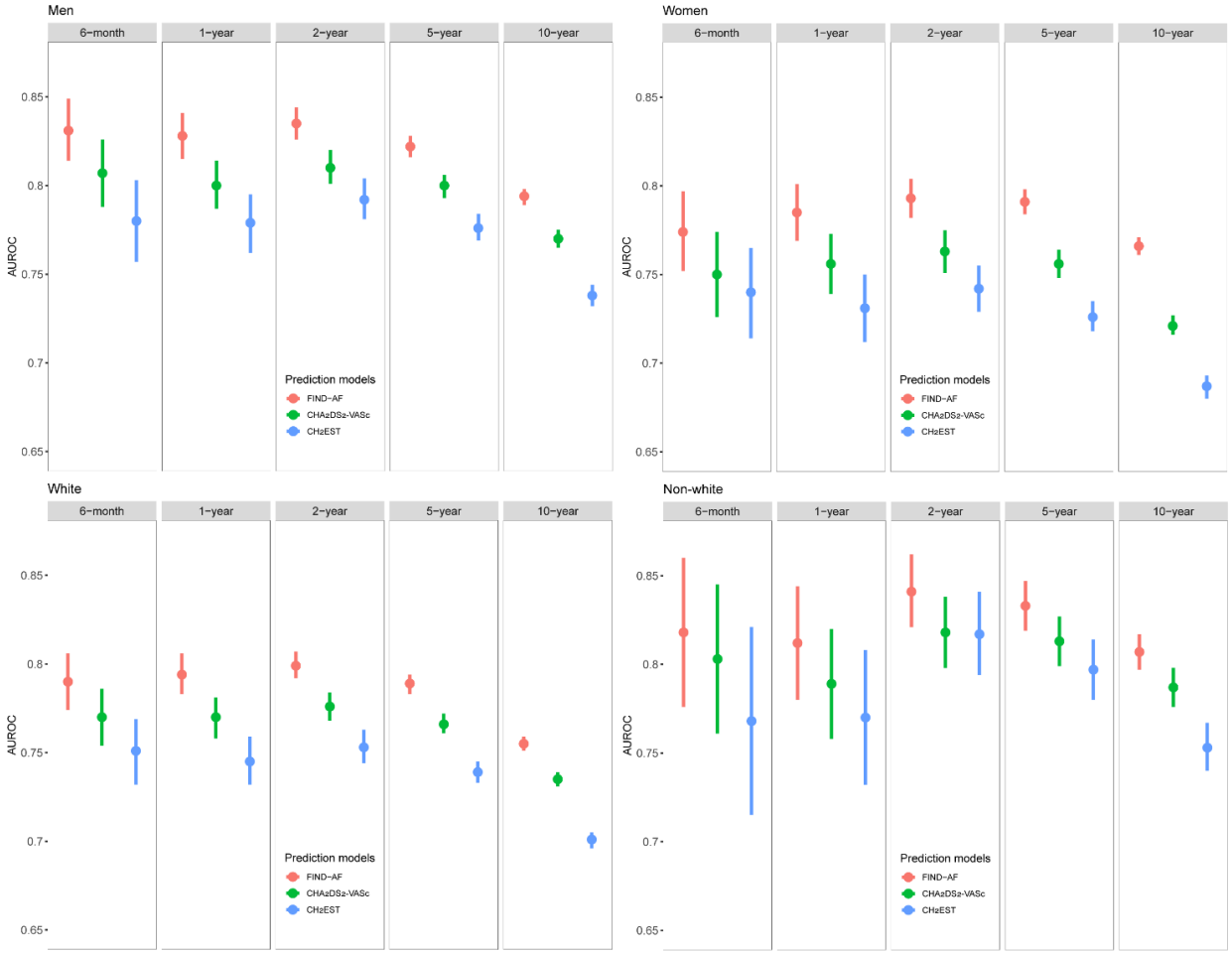
Figure 6 Prediction horizon-specific calibration plots in the validation sample of the FIND-AF prediction models, demonstrating observed (red) and predicted (blue) mean risk of atrial fibrillation, stratified by predicted risk. A) 6-months, B) 2-years, C) 5-years, D) 10-years



Abbreviations: AF, atrial fibrillation; FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation

Performance was better in men than women and non-White individuals compared with White individuals across prediction horizons (Figure 7).

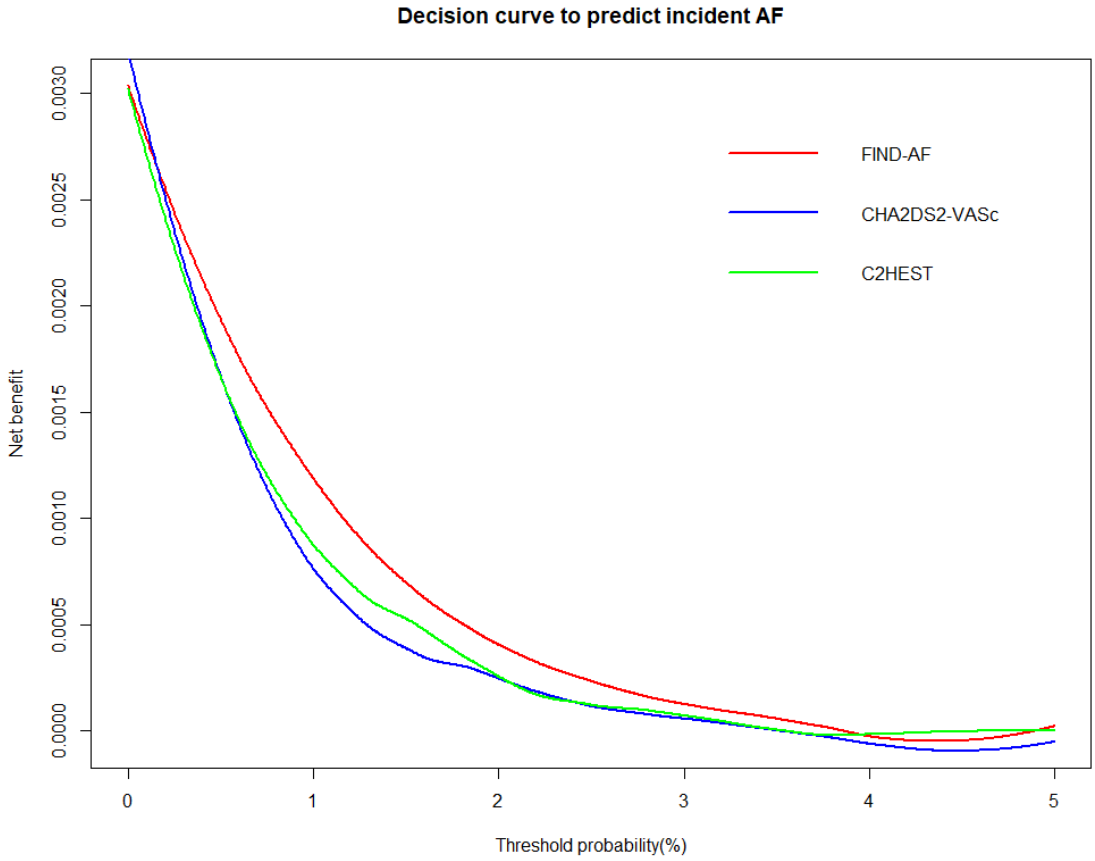
Figure 7 Prediction performance of FIND-AF, CHA_2DS_2-VASc, and C_2HEST prediction models across prediction horizons and clinical subgroups



Abbreviations: AUROC, area under receiver operating characteristic; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation

In a decision curve analysis there was net benefit from utilisation of the prediction models across all threshold probabilities (Figure 8).

Figure 8 Decision curve analysis for the FIND-AF, CHA_2DS_2-VASc, and C_2HEST risk scores

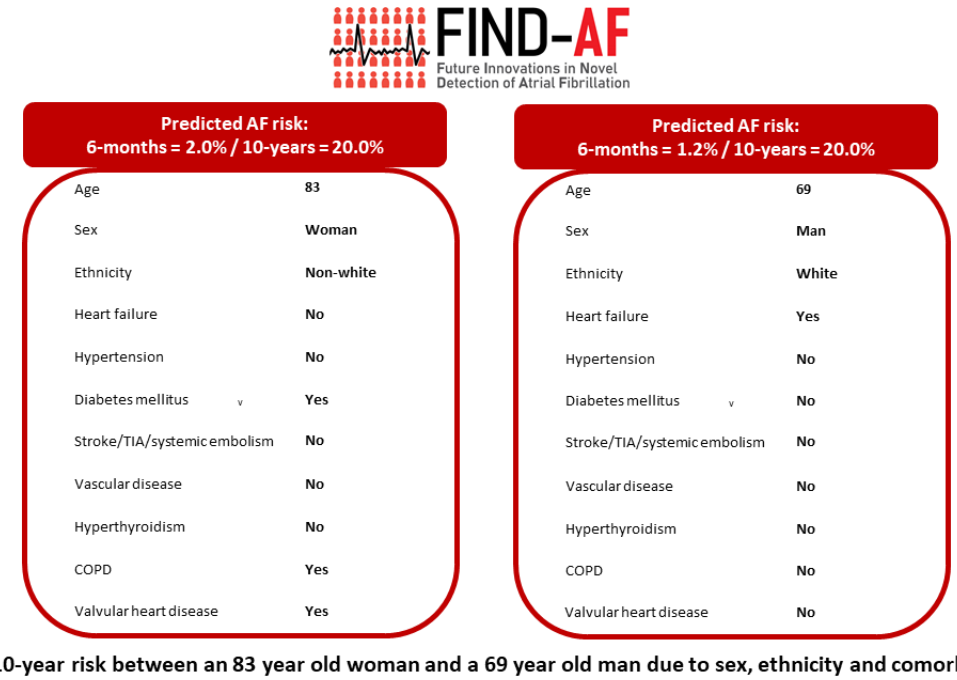


Abbreviations: CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation

Individuals in the highest group of predicted risk were significantly higher risk compared with other strata; at 5- and 10-years, 14.2% and 23.7% of these individuals had received a diagnosis of AF in routine practice (Figure 6). Varying risk factor profiles altered estimated absolute risk of AF. For example, estimated 6-month AF risk is 2.0% for both a 59-year-old White man with a high comorbidity burden (hypertension, ischaemic heart disease, valvular heart disease, and hyperthyroidism) and an 88-year-

old White man with hypertension. Furthermore, both an 83-year-old non-White woman with diabetes mellitus, COPD and valvular heart disease and a 69-year-old White man with heart failure have a 20.0% 10-year AF risk (Figure 9).

Figure 9 Example of predicted 6-month and 10-year atrial fibrillation risk estimates in a 69-year old White man and an 83-year old non-White woman with different comorbidity profiles using the FIND-AF calculator.

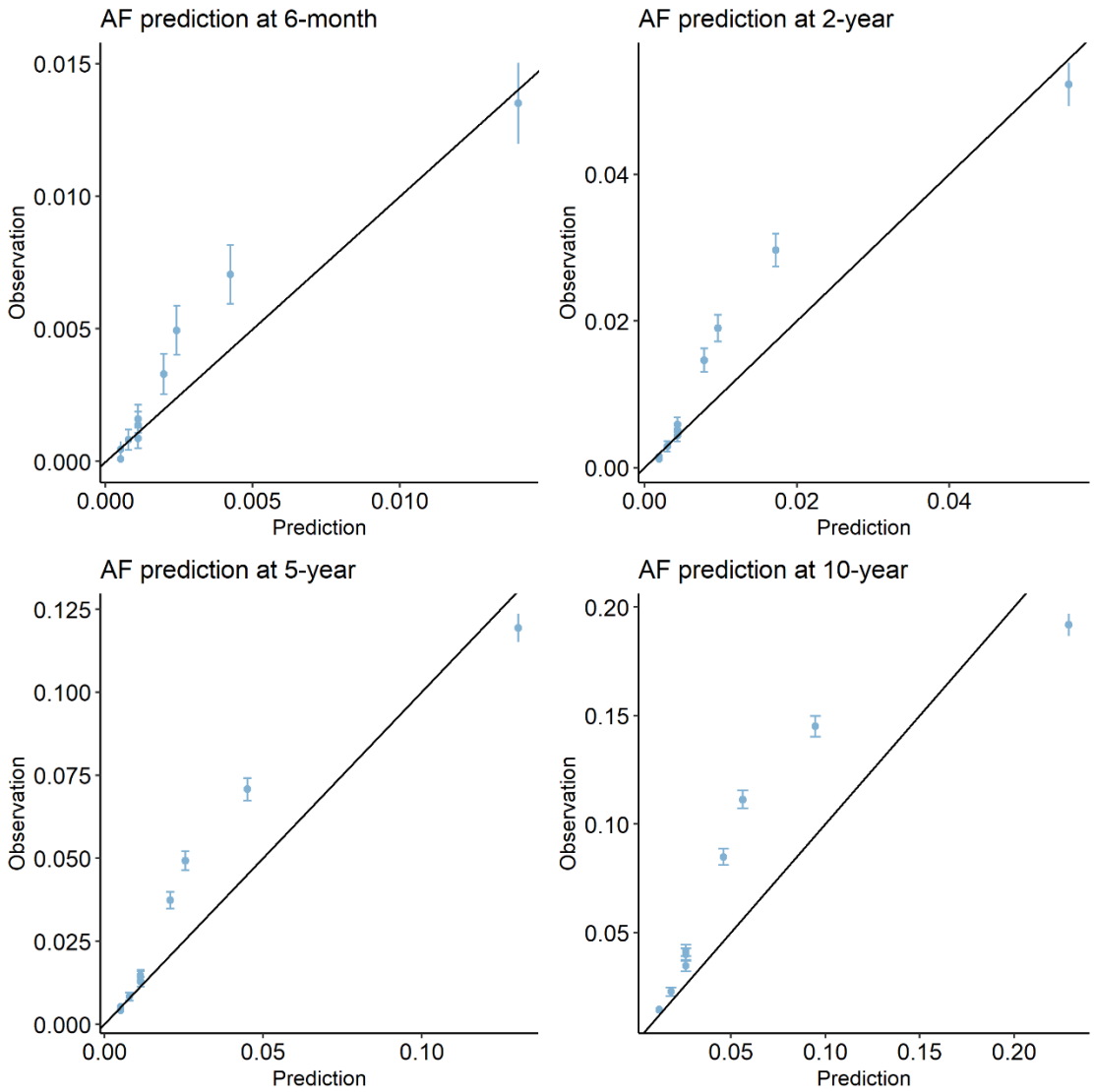


Abbreviations: AF, atrial fibrillation, COPD, chronic obstructive pulmonary disease; TIA, transient ischaemic attack; FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation

7.6.4 Comparison with CHA_2DS_2-VASc and C_2HEST scores for risk of incident AF

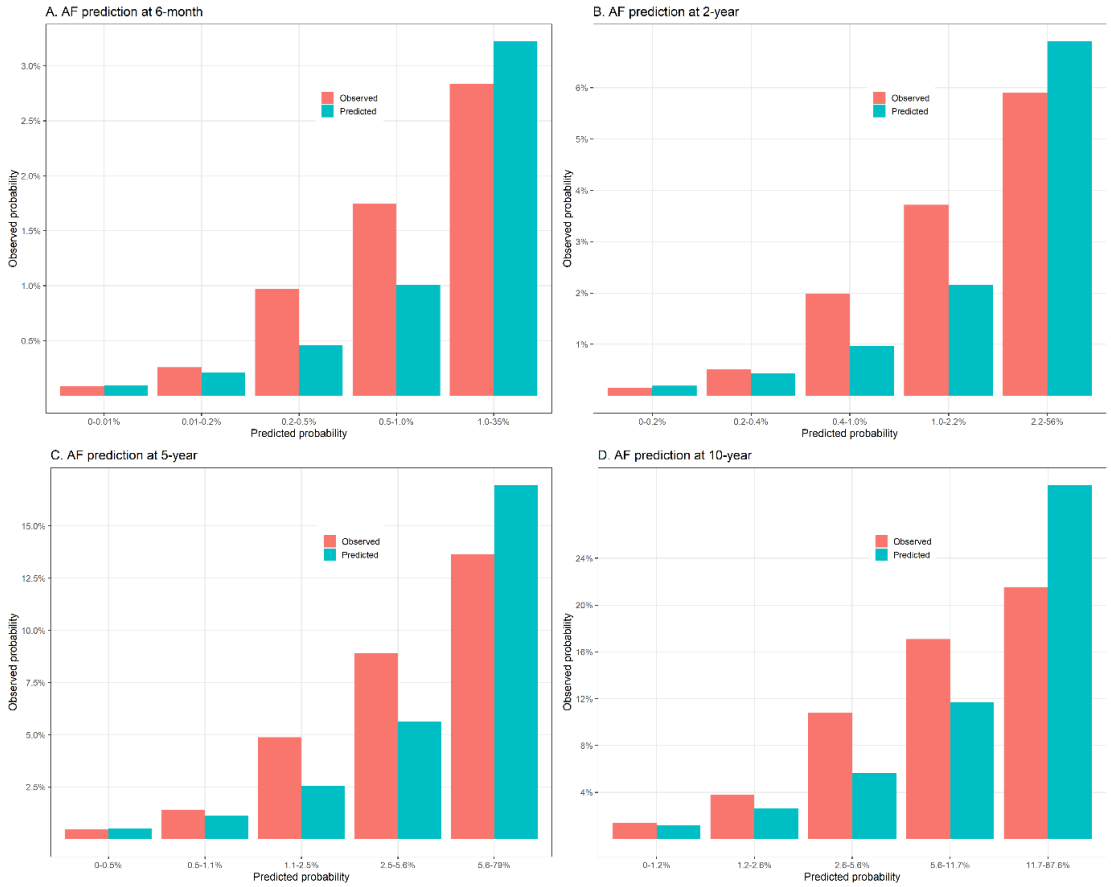
The CHA_2DS_2-VASc and C_2HEST scores showed good discrimination performance across all prediction horizons (AUROC ≥0.70), but were inferior in their performance to the FIND-AF prediction models (Table 3, Figure 1-4). Calibration was poorer for both the CHA_2DS_2-VASc and C_2HEST scores than the FIND-AF prediction models (Table 3). Both showed too much variation in predicted risk with underestimation of risk in the mid-range of predicted values, but over-estimation of risk at the highest range of predicted risk (Figure 10-13).

Figure 10 Prediction horizon-specific calibration plots in the validation sample of the CHA_2DS_2-VASc risk score. A) 6-months, B) 2-years, C) 5-years, D) 10-years



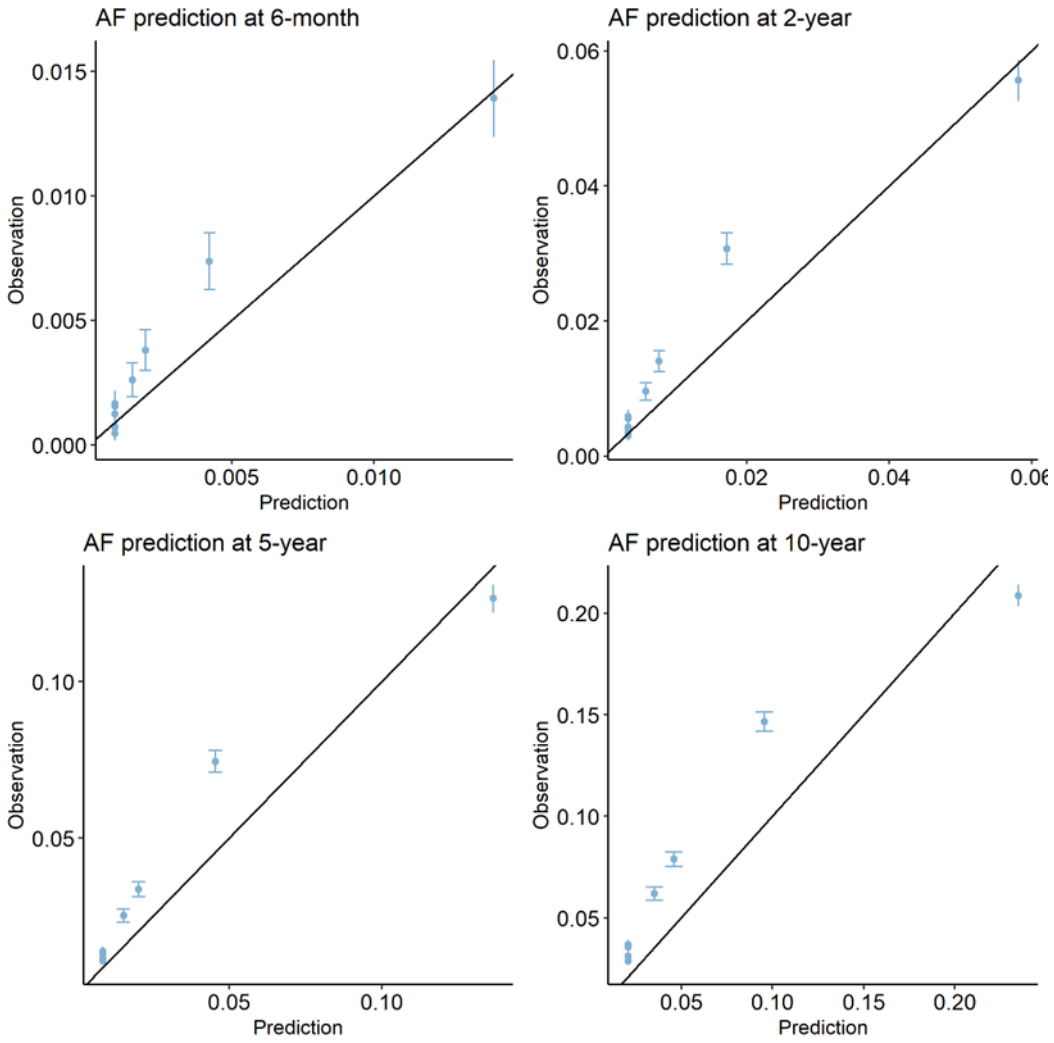
Abbreviations: AF, atrial fibrillation; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category

Figure 11 Prediction horizon-specific calibration plots in the validation sample of the CHA_2DS_2-VASc risk score, demonstrating observed (red) and predicted (blue) mean risk of AF, stratified by predicted risk. A) 6-months, B) 2-years, C) 5-years, D) 10-years



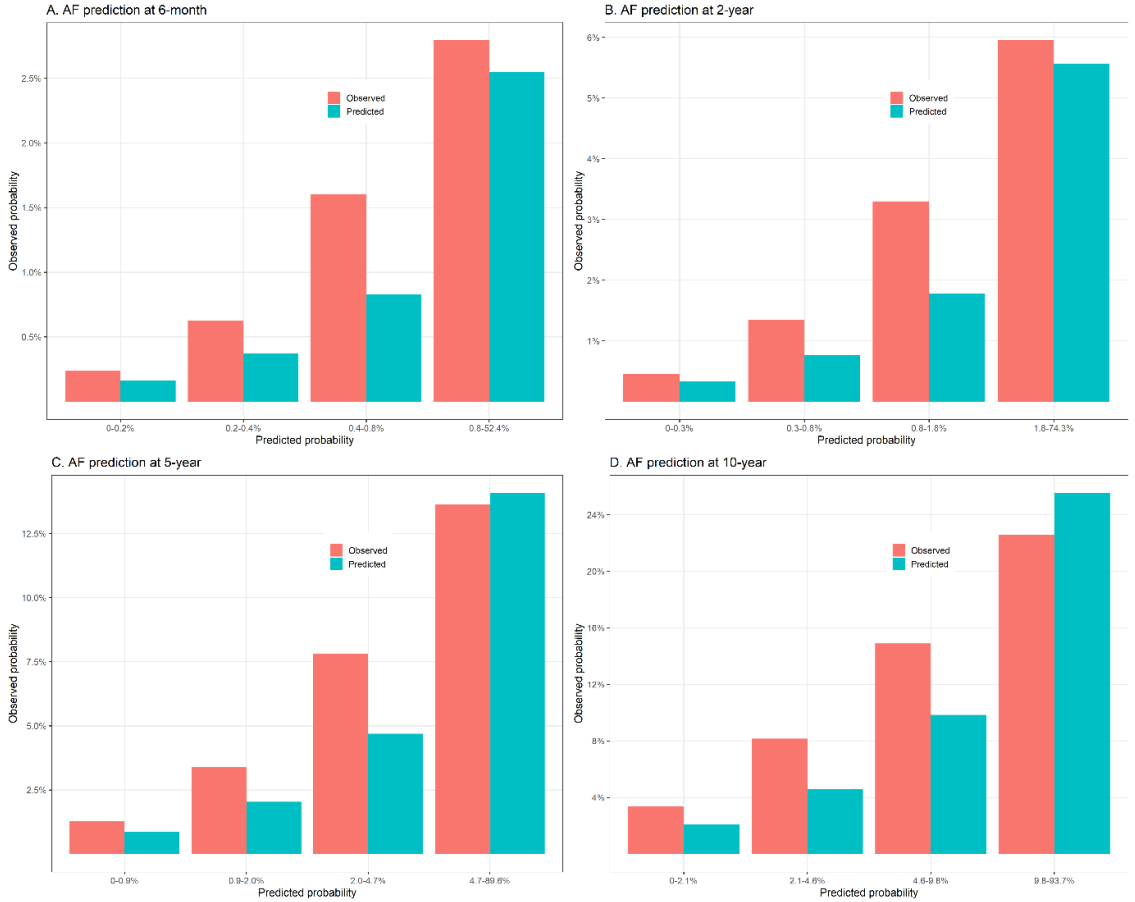
Abbreviations: AF, atrial fibrillation; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category

Figure 12 Prediction horizon-specific calibration plots in the validation sample of the C_2HEST risk score. A) 6-months, B) 2-years, C) 5-years, D) 10-years



Abbreviations: AF, atrial fibrillation; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism)

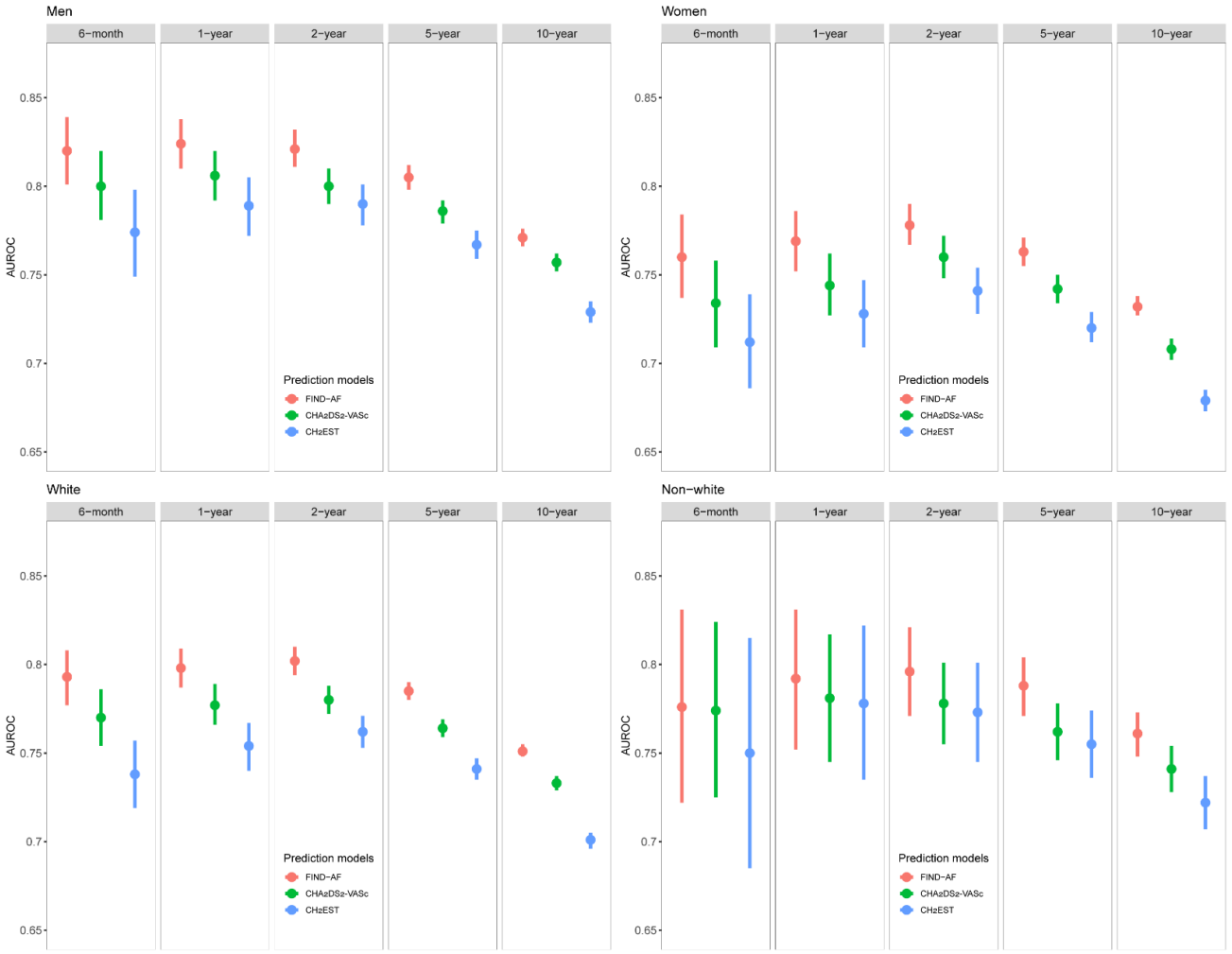
Figure 13 Prediction horizon-specific calibration plots in the validation sample of the C_2HEST risk score, demonstrating observed (red) and predicted (blue) mean risk of AF, stratified by predicted risk. A) 6-months, B) 2-years, C) 5-years, D) 10-years



Abbreviations: AF, atrial fibrillation; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); CI, confidence interval.

Both the CHA_2DS_2-VASc and C_2HEST scores showed better performance in men than women, and non-White individuals compared with White individuals across all prediction horizons (Figure 7). Across all prediction horizons, in both sexes and in White and non-White individuals, the FIND-AF prediction models had better discrimination performance for incident AF than the CHA_2DS_2-VASc and C_2HEST scores (Figure 7). The results were not altered when excluding individuals where ethnicity was unrecorded or unknown (Figure 14).

Figure 14 Prediction performance of FIND-AF prediction models, CHA_2DS_2-VASc, and C_2HEST risk scores across prediction horizons and clinical subgroups when individuals with unrecorded or unknown ethnicity are excluded



Abbreviations: AUROC, area under receiver operating characteristic; CHA_2DS_2-VASc, Congestive heart failure, Hypertension, Age >75 [2 points], Stroke/transient ischemic attack/thromboembolism [2 points], Vascular disease, Age 65-74, Sex Category; C_2HEST, Coronary artery disease / Chronic obstructive pulmonary disease [1 point each], Hypertension, Elderly (Age ≥75, 2 points), Systolic heart failure, Thyroid disease (hyperthyroidism); FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation

7.6.5 Clinical application

7.6.5.1 Web-based tool

To facilitate clinical application of the FIND-AF prediction models, a web-based FIND-AF calculator was developed to provide an estimated risk of AF for adults aged 30 years and older over 6-months, 1-year, 2-years, 5-years and 10-years depending on the user's requirements (https://minimization.shinyapps.io/FIND-AF-MLR/).

7.6.5.2 Risk score

We also developed a FIND-AF risk score to enable calculation of 10-year AF risk (Table 4). The total risk score ranged from a minimum value of 0 (lowest risk) to a maximum value of 14 (highest risk). Patients may be categorised into 3 risk groups based on their risk score (0-3: low, 4-6: high, 7-14: very high).

Table 4 Points assigned to atrial fibrillation risk factors in the FIND-AF 10-year atrial fibrillation risk score

Variable	**Score**
Demographics	
Age (years)	
<50	0
50-59	1
60-69	2
70-75	3
>75	4
Woman	1
White ethnicity	1
Comorbidities	
Heart failure	2
Hypertension	1
Diabetes	1
Stroke/TIA/systemic embolism	2

Valvular heart disease	2
Vascular disease	1
Hyperthyroidism	1

Abbreviations: AF, atrial fibrillation; FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation**;** TIA, transient ischaemic attack

N.B. * All the risk factors included in the risk score were included in the scoring system other than COPD, which had the weakest association with incident AF across the prediction horizons

The very high risk cohort constituted 3.1% of the validation set (13 111/416 228) and within 10 years 24.4% (3204/13111; Table 5) had received a diagnosis of AF during routine care.

Table 5 Occurrence of atrial fibrillation at 10 years by FIND-AF 10-year AF risk score and categorisation

FIND-AF risk category	**FIND-AF risk score**	**% of population**	**% who were diagnosed with AF at 10 years**
Low	0	8.71% (36259/416228)	0.05% (17/36259)
Low	1	33.04% (137527/416228)	0.25% (337/137527)
Low	2	24.62% (102467/416228)	1.15% (1179/102467)
Low	3	12.14% (50548/416228)	3.88% (1962/50548)
High	4	8.43% (35089/416228)	8.15% (2860/35089)
High	5	6.13% (25510/416228)	13.80% (3520/25510)

High	6	3.78% (15717/416228)	18.70% (2939/15717)
Very high	7	1.70% (7079/416228)	22.52% (1594/7079)
Very high	8	0.86% (3582/416228)	24.34% (872/3582)
Very high	9	0.37% (1553/416228)	29.62% (460/1553)
Very high	10-14	0.22% (897/416228)	30.99% (278/897)

Abbreviations: AF, atrial fibrillation; FIND-AF, Future Innovations in Novel detection of Atrial Fibrillation**;** TIA, transient ischaemic attack

7.7 Discussion

In this study we developed and internally validated parsimonious prediction models for incident AF (FIND-AF) at varying prediction horizons using clinically recognised risk factors in a cohort of over 2 million community-dwelling individuals. We demonstrate that the magnitude of association between risk factors and incident AF change over different prediction horizons, and so deriving risk scores with different coefficients for each prediction horizon is required to optimise prediction. The FIND-AF prediction models provide estimates of AF risk from 6-months to 10-years, only include age, sex, ethnicity, and eight comorbidities, and are available through a web tool (FIND-AF calculator). As such, the FIND-AF calculator could be used to identify potential participants for trial recruitment and inform clinical decisions for both screening and primary prevention for AF.

To our knowledge, this is the first analysis to demonstrate that the association of known risk factors with incident AF varies at different prediction horizons. In contrast to geographically distinct prospective cohorts, such as the Framingham Heart Study or Atherosclerosis Risk in the Community,[244, 309] the size of this nationwide sample provided a large enough sample of AF cases at both short- and long-term prediction horizons to derive meaningful estimates of association. Valvular heart disease and heart failure were the strongest predictors for AF, concordant with previous literature.[118]

However, the association between most comorbidities and incident AF weakened as the prediction horizon lengthened, potentially signalling the increasing importance of other factors at longer prediction horizons. In contrast to other comorbidities, the association between hypertension and AF increased at longer compared with shorter prediction horizons. This may be because the underlying pathophysiological relationship between hypertension and AF is that of progressive haemodynamic and structural changes related to left ventricular hypertrophy, LA remodelling accompanied by profibrotic changes, and renin-angiotensin-aldosterone system (RAAS) upregulation,[310] whereas with valvular heart disease and heart failure the AF substrate is already developed.[289, 311] By incorporating the variation in association between risk factors and incident AF over different prediction horizons through horizon-specific coefficients we were able to achieve only a 3% deterioration in prediction performance using the same variables at both 6-months and 10-years, whereas the performance of previous AF prediction models has been demonstrated to deteriorate by up to 12% between short- and long-term prediction.[266]

The predictive performance of the FIND-AF prediction models was statistically superior to the CHA_2DS_2-VASc score and C_2HEST scores for incident AF across different prediction horizons in this European population-based cohort. For the CHA_2DS_2-VASc score female compared to male sex gives a higher score but is associated with a lower risk of AF,[20] and the score does not include valvular heart disease, which was found to be the most important variable for incident AF risk in this cohort. The performance of the C_2HEST score here was similar to that described in its original study (AUROC 0.749)[199] and in a nationwide cohort of French hospital-based post-ischaemic stroke patients (0.734).[312] As it dichotomises age, and does not incorporate ethnicity and valvular heart disease, the ceiling of its performance may be limited. Several other risk scores have good discrimination for incident AF in large general population cohorts (CHARGE-AF, FHS-AF, ARIC-AF, $HARMS_2$-AF)[205, 244, 309, 313] but they require many instrumental, examination, and laboratory variables that might not be easily accessed in the community (Table 1).[183, 286] Accordingly, the FIND-AF prediction models represent an advance on previous tools to predict incident AF in being accurate, parsimonious, implementable at scale, and optimised for different prediction horizons. Importantly, the web-based calculator affords risk estimation to health care providers, trialists, and the general community, at a range of prediction horizons to suit their requirement.

AF screening at scale with non-invasive devices is feasible and acceptable to patients, but defining the eligible population by age- or stroke-risk has resulted in low yields for

newly detected AF (3.0-5.3%),[67, 68, 87, 88] which limits both the clinical- and cost-effectiveness of AF screening. The evidence base for primary prevention of AF predominantly relies on observational data and post-hoc analyses of clinical trial data where AF was not pre-specified as a primary or secondary endpoint, and occurrence was not systematically collected.[151] The FIND-AF models could be useful to identify reliably high-risk individuals for clinical trials of AF screening and primary prevention. For example, almost one in four of the 'very high' risk cohort received an AF diagnosis within 10 years in routine practice and this would likely be far higher with systematic endpoint collection. If trials were positive, the FIND-AF models could easily be automated at scale within medical records to guide population-level prevention and screening.

There are some limitations to our study. First, the CPRD database is routinely collected, retrospective primary care data. Underestimation of AF incidence is possible since there will have been individuals with unrecorded asymptomatic AF. Second, we did not have complete information on subtype of AF (e.g. paroxysmal, persistent or permanent), precluding interrogation of performance by subtype. However, the clinical utility of such a temporal pattern classification remains inconclusive.[314] Third, important predictor variables may have been 'missing by design'; nonetheless, we aimed to develop prediction models that used data routinely available in the community.[252] Fourth, the generalisability of the performance of the FIND-AF prediction models in other geographies remains uncertain, but evaluation is underway in the Israeli Clalit Health Services.

7.8 Conclusions

Using data from over 2 million people, we developed and validated prediction models to accurately estimate absolute risk of incident AF over a range of prediction horizons. These models are available as a web-based FIND-AF calculator and may serve to identify individuals for AF screening, primary prevention, and clinical trial enrolment.

Part III

Chapter 8 Discussion

In this Chapter, I will present the accomplishments of my PhD studies and critically discuss the weaknesses of the methodology used. Then I will discuss the challenges for translating the outputs of this PhD to clinical practice, and suggest future directions of research.

8.1 Accomplishments of the PhD studies

The thesis presents my research where I i) conducted a systematic review of prediction models for incident AF that have been developed and/or validated in community-based EHRs and performed quantitative synthesis of their performance metrics, ii) developed prediction models for incident AF that require only variables that are routinely collected in UK primary care EHRs, ii) quantified the association between risk of AF and the occurrence of non-AF outcomes, and iii) designed a pilot study to prospectively clinically validate the random forest (RF)-derived prediction model (FIND-AF).

My systematic review and meta-analysis demonstrated key shortcomings in prediction models for incident AF developed and/or validated in community-based EHRs. First, summary discrimination performance was only moderate. Most of the models had been developed using traditional regression techniques, but my analysis demonstrated that supervised ML techniques may offer incremental improvement in prediction performance. Second, prediction horizons were generally 5- or 10-years, which is less relevant when one is considering organising investigation for AF at the time of risk prediction. Third, the majority of models relied on the availability of laboratory investigations or observations to provide prediction, but this information has been shown to be missing in the majority of routinely-collected EHRs in the community.[52, 196] For this publication I conceived the idea, formulated the literature search, screened articles, extracted data, conducted statistical analysis, and then drafted and revised the manuscript. Professor Gale contributed to conception and Professor Wu supervised the statistical analysis. Dr Alsaeed screened articles and extracted data. Dr Hurdus, and Dr Aktaa screened articles. All co-authors reviewed the manuscript.

To address these shortcomings, for the development of FIND-AF, I prioritised clinical relevance, improved prediction, and applicability to existing healthcare EHR platforms in the UK NHS. I chose a prediction horizon of 6-months because this timescale is in keeping with the logistics of organising investigation as part of a screening programme, and may reflect individuals who have AF at the point of risk stratification but have yet to be diagnosed in routine care. I considered that supervised ML may offer improved prediction performance. Of ML methods, I chose a RF technique because it was the most frequently used technique and demonstrated robust performance across different EHR datasets in the systematic review, and because it is possible to show the importance of variables used in predictions.[315] I chose to limit candidate variables to age, sex, comorbidities and ethnicity (incorporating an 'ethnicity unrecorded' variable because a missing record of ethnicity is informative).[268] On analysis, FIND-AF demonstrated excellent prediction performance for incident AF in a large UK cohort, representative of the whole population in terms of age, sex, and ethnicity,[202] and the accuracy of predictions was robust across both sexes and in different ethnic groups. The accuracy of predictions was superior to previous models used for prediction of incident AF within community-based EHRs, and a model developed with traditional regression techniques in this cohort of individuals in the UK. Importantly decision curve analysis suggested use of the model would result in net clinical benefit across the threshold of probabilities. For this publication I conceived the idea, curated the codelists for variables and outcomes, designed the machine learning method and contributed to statistical analysis, and then drafted and revised the manuscript. Professor Gale contributed to conception of the idea and Professor Wu contributed to statistical analysis. All co-authors reviewed the manuscript.

I, alongside my supervisors Professor Gale and Professor Wu, applied and received funding (Bristol Myers Squibb British Heart Foundation Cardiovascular Catalyst Award - CC/22/250026) to conduct a prospective clinical validation of FIND-AF. I have formulated the protocol for the pilot, interventional, non-randomised, single arm, open label study. I chose this study design to test the hypothesis that the yield of AF diagnosis during ECG monitoring increases as predicted AF risk increases. By recruiting participants across a range of risk estimates I may be able to identify the threshold at which yield of new AF may be adequate to justify targeted screening. Furthermore, influenced by the design of RCTs of systematic AF screening such as SAFER and AMALFI, I have designed the study such that participants can participate remotely.[305, 316] Through enabling participants to consent to participate and undergo ECG monitoring without having to travel, I aim to increase the rates of participation and representativeness to the general population. I have received ethical approval for the study and it will begin enrolling participants in September 2023. For this publication I

conceived the idea, designed the study, completed the ethics submission, and then drafted and revised the manuscript. Professor Gale contributed to conception, and Professor Wu contributed to sample size calculations. All co-authors reviewed the manuscript.

I was also interested to understand how risk of AF, as identified by a ML algorithm, was associated with risk of other diseases and death. I chose to examine outcomes that are well-characterised to be associated with AF,[1, 32] and to compare the rate of events between individuals classified as higher and lower risk of AF. In the analysis conducted by Professor Wu and I, individuals with the ML-derived EHR phenotype of 'higher predicted AF risk' had an increased rate of cardio-renal-metabolic diseases and death, irrespective of subsequent occurrence of AF. Individuals at higher predicted AF risk constituted 70% of new HF diagnoses, 65% of new aortic stenosis diagnoses, and 74% of cardiovascular and cerebrovascular deaths over the decade following risk stratification. This demonstrates that the EHR phenotype identified by ML for AF has consequences beyond AF alone. For this publication I contributed to conception of the idea, identified the outcomes, designed the statistical analysis, and then drafted and revised the manuscript. Professor Gale contributed most to conception, Professor Wu contributed most to statistical analysis, and all authors reviewed the manuscript.

The RF FIND-AF model deliberately incorporated a large number of comorbidities and ethnicity categories, because I was seeking to maximise prediction performance with the expectation that in clinical practice the algorithm would be calculated automatically within the EHR platform. However, there are challenges to implementation of ML algorithms through EHR platforms, described below in Section 8.5.2, so I considered that it could be clinically useful to develop a simple risk prediction model for incident AF that could be accessed and used outside of an EHR platform. I chose to limit the variables to age, sex, ethnicity and comorbidities to enable the conduct of risk stratification to be remote, without the need for observations or laboratory measures. I also aimed to make the models understandable to clinicians by restricting variables to those well-characterised in international guidelines to be associated with AF.[1] I recommended to use a traditional regression technique in model development so that the association between each variable and AF could be quantified, and also to enable the transformation of the prediction model to a risk score. Because of these concessions, the prediction performance for AF within a 6-month prediction horizon of the parsimonious traditional regression model developed by Professor Wu and I is not quite as high as that of the RF FIND-AF model (AUROC 0.803 vs 0.824).

Over the course of the PhD studies I became increasingly interested in predicting long-term occurrence of AF using routinely-recorded data. This stemmed from work I was involved with, but was outwith of this thesis, which demonstrated that the incidence of AF in the UK is increasing rapidly (Chapter 1, Section 1.1.3). I now believe that to address the health and economic burden of AF on the UK NHS (Chapter 1, Section 1.5), it is not only imperative to identify the prevalent undiagnosed cases, but also to aim to delay/prevent cases in the future. However, formal pathways for primary prevention of AF are not available in the NHS. In the analysis conducted by Professor Wu and I presented in Chapter 7, using a small number of routinely-recorded variables could predict new-onset AF at both a short and long prediction horizon. These prediction models could be useful to identify individuals for recruitment into trials of primary prevention of AF using routinely-collected data, which I discuss below in Section 8.6.4. For this manuscript, which is under review for publication, I conceived the idea, curated the codelists for variables and outcomes, designed the statistical analysis, and then drafted and revised the manuscript. Professor Wu contributed most to the statistical analysis, and all authors reviewed the manuscript.

8.2 The gap that the PhD studies address

Opportunistic and systematic screening for AF with non-invasive devices has been the subject of several RCTs, as summarised in Chapter 1, Sections 1.3.8.1 to 1.3.8.3. Hitherto the eligibility for inclusion in these studies has been based on age or stroke risk. However yields from this approach to AF screening have been low.[67, 68, 87, 88] Low yields impact the clinical- and cost-effectiveness of AF screening. Consequently, systematic screening programmes at a comprehensive national healthcare level do not currently exist in any of the European countries or the USA.[317] As such there is the requirement to better refine the population to whom AF screening is offered with the aim to improve yield of newly detected AF. Comprehensive risk stratification using multiple variables associated with risk of AF may be able to better identify individuals likely to have undiagnosed AF than using age or stroke risk.[196]

In recognition of the importance of this research area the EU-funded AFFECT-EU project aims to develop a risk-based AF screening strategy using digital applications for rhythm monitoring to reduce the burden of stroke and other AF-related comorbidities.[318] A qualitative study including 24 healthcare professionals and regulators from 11 European countries explored opportunities and challenges for implementing AF screening.[317] There was a perceived need to implement AF screening and participants supported both opportunistic screening and systematic screening. Participants considered single time point opportunistic AF screening using a single-lead ECG

device as the most feasible approach on account of ease of implementation and lower costs, whereas the most effective approach was considered to be prolonged screening either with a continuous patch (as used in the mSToPS study)[87] or intermittent ECG monitoring over a 2-week period (as used in the STROKESTOP study)[67]. However prolonged monitoring was considered too expensive to implement, especially when the yield of new cases is low. Primary care was considered the most appropriate location for AF screening by the majority of participants Some participants stated that software systems in primary care had the potential to identify suitable individuals for screening but that algorithms did not currently exist for this.

A stated aim of the AFFECT-EU project is to develop clinical prediction models to permit refinement of AF screening and reduce the number needed to screen, though no prediction models have yet been published from this project. The VITAL-AF and D2AF RCTs demonstrated that opportunistic screening in individuals aged 65 years and older did not increase detection rates of AF compared to routine care (Chapter 1, Section 1.3.8.1.1). Targeting opportunistic screening to a reliably identified high-risk cohort may improve the yield of AF detection, and increase the chance of healthcare professionals in primary care adding to their usual daily work a task that may often be overlooked. The prediction models developed in the PhD studies have been designed to be implementable at scale in primary care EHR systems in the UK and provide accurate prediction in the UK population including in relevant subgroups. Thus the accomplishments presented in this thesis help address the gap for a model to refine AF screening in the primary care setting in the UK. The protocol for the pilot study (Chapter 6) could also be translated to a scalable digital detection pathway aspired to in the AFFECT-EU project. Furthermore the models deliver absolute risk estimation and so can enable better communication of AF risk to individuals which may improve concordance with investigations and treatments.

8.3 Addition to existing knowledge

Whilst multivariable prediction models for incident AF have been developed using community-based EHRs (Chapter 2), these PhD studies offer novel advances.

Prediction models for incident AF have mostly predicted AF over a long prediction horizon. In the case of models originally developed using prospective cohorts this may have been because their limited sample size meant they did not have sufficient AF cases within a short prediction horizon (6-months or 1-year) to derive meaningful associations and develop a robust prediction model.[205, 309] When the prediction horizon

extends to up to ten years individuals classified as higher risk may not have manifested AF within the next few months, which is the expected timescale for organisation of investigation after risk stratification. By contrast, in this PhD, incident AF within the next six months was predicted to ensure immediate relevance of the prediction for targeting investigation. Additionally, it was demonstrated that individuals at higher risk of AF within six months continue to have an increased rate of AF occurrence over the next ten years compared to individuals with lower predicted AF risk. Thus, individuals identified as higher predicted risk of AF may not just be suitable for a one-off screening event, the approach tested in previous AF screening trials,[67, 87, 88] but also repeat screening at a later time point (longitudinal screening),[319] akin to how breast cancer and cervical cancer screening is delivered.[48]

Furthermore, from inception I considered the issue of implementation of the prediction model. Whilst previous reports have used large EHR databases for deriving prediction models for AF, it is not clear that the authors have considered the practicalities of implementing their models in clinical practice. For risk-guided AF screening to be efficient minimal resources should be required to adequately perform risk stratification. Previous models for incident AF often require observations such as blood pressure, height, weight and/or BMI (Chapter 2, Table 8 and Chapter 7, Table 1). All of these variables are associated with AF risk but have been shown to be incomplete in real-world primary care data,[320] with selective reporting favouring those with higher comorbidity rates.[321] When the CHARGE-AF model was evaluated in a primary care EHR database representative in the Netherlands only one in six individuals aged 40 years and older without prevalent AF had the complete baseline data for risk stratification.[196] Individuals with complete baseline data were systematically different to individuals without complete baseline data; they were older, had a ten-fold higher prevalence of DM, and a three-fold higher prevalence of hypertension and COPD. Data completeness for observations in primary care EHRs varies by age, sex, ethnicity and deprivation index,[322, 323] so the implementation of a prediction model only applicable to people with complete data in a screening programme may entrench health inequalities. Higher completeness for data in primary care EHRs is to be aspired to, but until this is achieved models that do not rely on measurement variables may be the most suitable choice for remote, automatic AF risk assessment in primary care settings.[196] I decided to limit the variables included in the FIND-AF prediction models to variables that have high completeness in routinely-collected records, and incorporated an 'ethnicity unrecorded' category. Accordingly, the FIND-AF prediction models could be applied to all individuals in the large dataset of routinely-collected primary care records. The CHA_2DS_2-VaSC algorithm can be called for each patient record through EHR systems because it only requires age and the presence of structured codes for comorbidities.

The FIND-AF algorithm, by design, should be similarly implementable at scale in primary care EHRs, which cover 98% of the UK population,[202] to guide nationwide population AF screening.

Models that leverage supervised ML to predict incident AF using community-based EHR data have been reported (Chapter 2), but they have failed to report calibration, how performance varies by sex and ethnic group, or clinical utility analysis.[194, 236, 241] These metrics are important when one considers implementing a model in clinical practice. Notably in a systematic review and meta-analysis of prediction models for HF applicable in the community that I conducted, but is outwith of this thesis, I also found that calibration or clinical utility analysis had not been reported for ML models.[209] Adherence to reporting guidelines designed specifically for risk prediction ML studies, currently under development,[324] would improve the quality of reporting and increase confidence in the translatability of these models to clinical practice.[325] This PhD advanced the field of reporting for ML algorithms to predict AF by considering these factors within the analysis.

This PhD has also extended the concept of AF risk beyond previous reports. Some reports have considered the association between elevated AF risk and occurrence of stroke, but from the perspective of whether AF was diagnosed before or after the stroke event.[326] This reflects a narrow interest in prediction of AF risk for stroke prophylaxis. However AF is increasingly understood as a manifestation, expression, and symptom of underlying disease.[11, 327] After diagnosis AF is associated with increased risk of cardio-renal diseases beyond stroke,[32] but the importance of the arrhythmia itself on these occurrences is unclear. Accordingly, if individuals with the same characteristics as those who have AF can be found, they may be also at risk of adverse events.

Machine learning can uncover meaningful non-linear associations not apparent to physicians.[114] After adjustment for age, sex, ethnicity, and the presence of other cardio-renal-metabolic comorbidity at baseline, elevated AF risk was still associated with an excess risk for all outcomes, but particularly aortic stenosis (64%), HF (63%), CKD (46%), and stroke or TIA (40%). This may suggest that data-derived associations between the large number of morbidities included in the model (including diseases such as gout, rheumatological diseases, pulmonary hypertension and inflammatory bowel disease) may mimic underlying pathological changes. At the least, the association illustrates that cardio-renal-metabolic diseases commonly co-exist, and have shared risk factors and pathological pathways.[328, 329]

The association between AF risk and non-AF cardio-renal-metabolic diseases and death may also have important clinical implications. It is increasingly recognised that multimorbidity is very common in AF patients. In the UK, 93.5% of AF patients have at least one comorbidity, and amongst those aged 65 years and older the mean number of comorbidities is 5.[327] Work that I was involved in, but is outwith of this thesis, has demonstrated that stroke only contributes a small proportion to hospitalisation and mortality within one year of AF diagnosis.[330] The importance of managing comorbidities in patients with AF is highlighted in the 2020 ESC guidelines.[1] In this PhD, it was found that individuals at elevated risk of AF have adverse outcomes beyond AF and stroke, and make up three quarters of cardiovascular deaths over the following decade. Thus, one may consider that optimisation of comorbidities recommended in the ESC guidelines should be extended to those at elevated risk of AF before the manifestation of the arrhythmia.

8.4 Appraisal of the used methodology

Whilst this study and the outputs are genuinely novel, and the questions this thesis has addressed are of clinical importance, the limitations and alternative approaches that could have been utilised will now be discussed in the context of the literature.

8.4.1 Alternative machine learning techniques in electronic health record data

In these PhD studies I chose the RF method for the development of the FIND-AF. It is possible that other supervised ML techniques may have performed differently in predicting AF from the EHR dataset.[331] However, many studies have shown that amongst commonly used high-performing ML algorithms, there is minimal difference in performance.[114, 194, 331]

Feature extraction in supervised ML is based on domain knowledge; that is, I pre-selected the variables I considered important based on literature review, and the ML technique takes a data-driven approach to identify the strength of association between these variables and the outcome. By contrast, deep learning utilizes more advanced techniques to learn the representations directly from the raw data to generate abstract concept and patient representations (unsupervised learning), which may then be used for prediction.[109] Deep learning is commonly performed using artificial neural networks (ANNs) that simulate the neuronal activity of the human brain in the processing of

information.[332] ANNs consist of multiple layers of interconnected nodes (analogous to neurons in the brain), which help in learning from data. Typically, an ANN has an input layer, multiple hidden layers, and an output layer. These additional hidden layers allow the model to learn hierarchical representations of the raw data, with each layer learning to represent the data at a different level of abstraction. Due to these non-linear transformations, one is able to model many non-linear and heterogeneous effects. Several ANN architectures have demonstrated exceptional discriminative performance for disease prediction in EHRs.[333, 334] Furthermore, different variants of ANNs - convolutional neural networks, multi-layer recurrent neural networks and Transformers – are well-suited to capture information on the sequential order of visits and inter-visit duration,[333, 334] which may better model the temporality of EHR data, a person's evolving health status and disease pathogenesis.

However, there are drawbacks for using deep learning models for risk prediction. Constructing and training deep learning models is often time-consuming and dependent on computational resources. They can suffer from a lack of transferability;[335] a model trained in a particular dataset often cannot be reused for other tasks without significant retraining, and may become outdated if the feature space, distribution, or training dataset changes. Most importantly, ANNs are commonly a 'black box' model where, due to their multi-layer non-linear structure, their predictions are not traceable by humans.[336] By contrast with the RF method the importance of variables used in predicting incident AF could be demonstrated, which may make FIND-AF more likely to be 'trusted' by healthcare professionals and explainable to people when implemented at scale within a screening programme.[315]

8.4.2 Strategies to overcome the problem of missing data in routinely-collected electronic health records

In these PhD studies I chose to design the prediction models to overcome the issue of missing data in real-world practice. However, there are practical approaches to implement prediction models when there are missing values for variables, though many have been shown to be problematic.[337, 338] Some prediction models enforce valid values for all predictors, for example implementations of the Framingham model for cardiovascular disease and the Seattle Heart Failure model.[339, 340] Alternatively, some models allow for missing data on a limited set of variables and use simple imputation procedures. For example, the QRISK3 model uses i) the average value from the development study for a measure of deprivation when geographical region is unknown (mean imputation); ii) a conditional average based on ethnicity, age, and sex for

missing values of cholesterol/high density lipoprotein ratio, blood pressure and BMI (conditional mean imputation); and iii) zero imputation when the SD of the last two blood pressure readings is missing.[341]

Prediction models that intend to allow for missing data in practice implies that they need to be developed with missing data methods that transfer to real-life application. If model development data is available at the time of practical application, then particular statistical methods for handling missing data are possible. During a simulation study, the most accurate method, in terms of corrected c-statistic and root mean squared prediction error, was the use of the 2^k submodels (which requires estimated regression coefficients for all submodels of the prediction model) and use of fixed chained equations (which requires the vector of parameter estimates for each of the fully conditional models derived in the development dataset, as well as the mean of each variable in the development data).[337] However there are practical limiting factors with these approaches. First they computationally very expensive, because each new prediction requires imputation data. Second the development data has to be available at the time of prediction, which is often not possible due to privacy regulations. Pragmatic imputation of real-world missing values is possible, as well as the reduced model methods, hybrid model method, and the naïve approach.[342, 343] Performance in real-life practice may not deteriorate far below the level seen in research datasets using these methods as long as the extent of missingness in the variables that contribute the most to prediction is small.[344]

8.4.3 External validation in other electronic health record datasets in the United Kingdom

The predictive performance of a model in the development dataset is often optimistic, related to the association between predictors and outcomes often being stable within the same sample. Prediction models may correspond too closely or accidentally be fitted to idiosyncrasies in the development dataset, which is called overfitting.[345] This will result in predicted risks that are too extreme when used in new patients.[262] Ideally, a newly developed prognostic model needs to prove reproducibility and generalisability.

Reproducibility pertains to whether a prediction model is valid in new individuals that are similar to the development population. The internal validation of FIND-AF, in a large dataset of 400,000 individuals not included in the development of the model, can give an indication that the model performs satisfactorily in new patients that are similar to the development cohort.[346] As the CPRD dataset is representative of the UK population

in terms of age, sex, and ethnicity,[202] the data in this PhD suggests that the performance of FIND-AF will be robust when used in the UK population.

However, in different EHR systems code usage and prevalence may vary depending on different browsers, types of data entry templates and incentivised coding. I hoped to externally validate the FIND-AF model in the ResearchOne database of TPP, but unfortunately the University and TPP were unable to reach agreement on the terms of a material transfer agreement to allow the external validation to take place. Professor Gale and I are currently working to progress an external validation with Egton Medical Information Systems (EMIS).

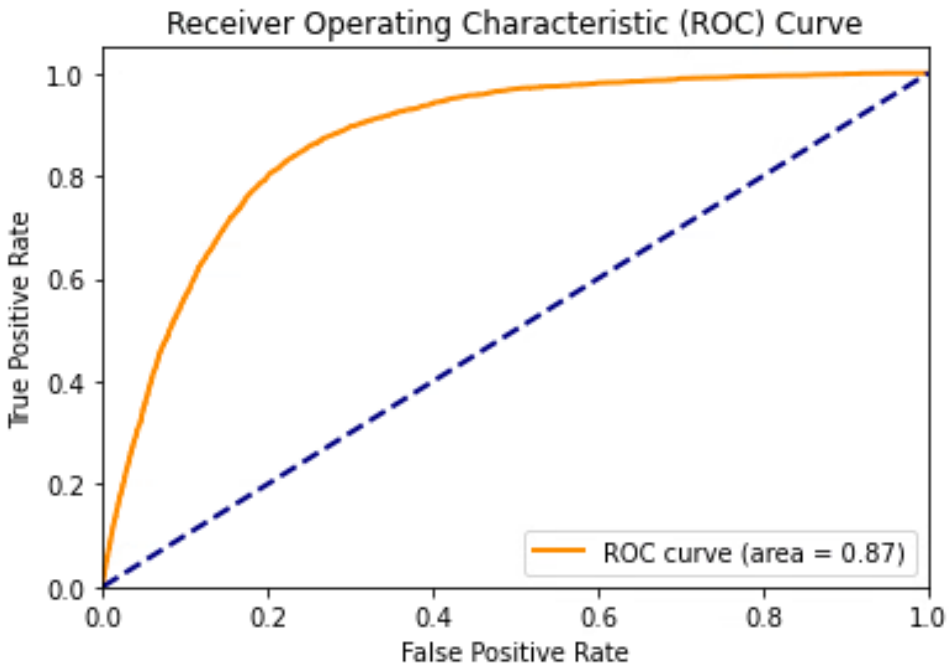
8.4.4 External validation in electronic health record datasets in external geographies

Generalizability (also called transportability) involves exploring whether the prediction model is transportable to a separate population with different patient characteristics. In Chapter 4 the predictive performance of FIND-AF was compared to the CHA_2DS_2-VASC and C_2HEST models. FIND-AF showed superior performance, but that is not surprising as it was the comparison of performance between the internal validation of one model and external validation of another model. In such a case it was likely that FIND-AF would appear superior, as it is optimally designed to fit the data.[345] The ideal direct comparison of performance between two prediction models should be done in an external validation dataset that is independent of both model development cohorts.[347] Ideally, external validation is performed in a separate study by different researchers to prevent the temptation of fine-tuning the model formula based on external validation results.[348] As demonstrated in Chapter 3, our research group have developed a collaboration with a research group who are conducting the external validation of FIND-AF in the Clalit Health Services dataset in Israel.

The FIND-AF model incorporates a large number of variables, including variables such as ethnicity, that may have a different meaning in an external geographic context. These variables may have to be substituted with other surrogates (e.g. socioeconomic status). Furthermore the Read codes used in the FIND-AF development have been translated into ICD-9 codes, the system used in Israeli community practice, but the prevalence of each variable may vary between the UK and Israel. There may also be heterogeneity of predictor effects, that is, the same predictor may have different prognostic value in varying populations. There may also be differences in case mix between the development and validation cohorts, that is, the distribution of predictor

values (e.g. differences in baseline characteristics such as prevalence of hypertension).[349] All of these factors could lead to a deterioration in the performance of the FIND-AF model, but preliminary results suggest that it still shows excellent discrimination performance, with an AUROC of 0.87 (Figure 1).

Figure 1 Receiver operating characteristic curve for FIND-AF in the Clalit Health Services dataset in Israel



Abbreviations: FIND-AF, Future Innovations in Novel Detection of Atrial Fibrillation

If the rest of the performance metrics are also excellent then this may provide evidence that that the model is generalizable for use in geographies external to the UK and interoperable with reference to the definitions of variables. However, given the variability in population characteristics and AF incidence between different countries,[350] a validation is only relevant to populations with the same characteristics as the validation population.

8.4.5 Under-recording of outcomes and predictors in electronic health records

When I defined the diagnostic codelists for the outcome and variables for the development of the model I had to assume that absence of a recording of a disease in the primary care EHR equates to absence of the event. However it has been shown

that recording of diseases is incomplete in primary care EHRs.[351] For example, the absence of a code for AF from a patient's EHR does not mean that AF is absent from the patient. It is possible that the diagnosis was recorded incorrectly in the primary care record, or was entered as free-text which is not available in the research database, or that AF is phenotypically manifest but has not yet been diagnosed. Incorporating diagnostic codes for AF used in primary and secondary care records reduces the possibility of missing AF cases, but a significant burden of AF is undiagnosed in usual care in the UK NHS (Chapter 1, Section 1.3.2). This may have led to the discrimination performance of FIND-AF being under-estimated as some of the false positives (individuals at higher AF risk but without a recorded AF diagnosis during follow-up) may have had undiagnosed AF. In the pilot study I will have data from a population with a range of risk estimates who have all undergone intermittent ECG monitoring, from which I can calculate positive predictive value, negative predictive value, sensitivity, specificity and AUROC for FIND-AF.

For the variables included in the model I limited the diagnostic codes used to only those available in primary care records (Read codes). This was to ensure only information that was available in the primary care record at point of care was included in the model, to align as closely as possible to the circumstance when the model would be used in clinical practice. The frequency of disease codes in EHRs is not random but rather indicates that the subject is ill and leads to the possibility of informed presence bias,[352] whereby more frequent interactions with healthcare professionals may give more opportunities for illnesses to be identified. This may have led to differences in the magnitude of association between certain variables and the outcome compared with a prospective cohort or trial. In these study designs the occurrence of a clinical event or comorbidity is actively sought for each participant at baseline and at set time intervals, so that the recording incidence is not contingent on the participant's engagement with the health sector. A study evaluating the degree of agreement between a community-based prospective cohort and an EHR database found good agreement for background characteristics but differences in cardiovascular risk factors and events.[353]

8.4.6 Study population

In these PhD studies I included a cohort of individuals aged 30 years or older without known AF. I chose to include individuals younger than 65 years of age in the cohort because previous work that I have contributed to, but is outwith of this thesis, has demonstrated that individuals in the UK from the most deprived socioeconomic quintile are diagnosed with AF at a younger age than individuals in the most affluent quintile.[11] It was also demonstrated in Chapter 4 that one in five new AF diagnoses within a 6-

month prediction horizon were under the age of 65 years. Moreover, stroke risk prevention in younger compared with older individuals could have greater potential long term benefits on an individual and societal level.[354] I also decided not to restrict the eligible population to individuals with an elevated CHA_2DS_2-VASc score, because it is possible that, as new OACs become available,[355] guideline recommendations on which patients with AF are eligible for oral anticoagulation may change.

However. one could argue that it is in the older population where AF screening is most justified as the incidence and prevalence of AF is strongly correlated with age.[11, 356] Including the cohort of individuals aged younger than 65 years led to age being, by far, the most important variable in the FIND-AF model, and prediction performance deteriorated in older populations. Furthermore, for older individuals with elevated CHA_2DS_2-VASc scores, a new AF diagnosis would lead to an immediate change in treatment. In a younger patient with a CHA_2DS_2-VASc score of 0 or 1, being informed of elevated AF risk may lead to health anxiety, extra investigations, and excess healthcare costs but diagnosis may not result in prescription of oral anticoagulation. Only 4.2% of individuals stratified as higher risk of AF by FIND-AF were aged younger than 65 years, so it may be appropriate to consider developing in future a version of FIND-AF specifically for individuals aged 65 years and older.

8.5 Challenges of translating an electronic health record-based prediction model for atrial fibrillation to clinical practice

To the best of my knowledge, this PhD thesis reports the first example of developing a ML algorithm in routinely-collected primary care EHR data that can be implemented at scale in UK primary care to stratify an individual's risk of AF within the next six months. In the following section I will explore some of the challenges to translate the algorithm to use in clinical practice.

8.5.1 Regulatory compliance

Though risk prediction models have been utilised in clinical practice for a number of years, they are now considered 'software as a medical device'.[357] Medical Devices have been regulated by three EU Directives since the early 1990s. This legislation is implemented and enforced in each EU member state by a Competent Authority. In the UK, the competent authority is the Medicines and Healthcare products Regulatory Agency (MHRA), and the Medical Devices Directive (93/42/EEC) classifies products according to their level of risk.[358] Researchers proposing for their risk prediction model

to be used in clinical practice must make sure the model complies with all the legislation's relevant essential requirements in order for the prediction model to be certified as a medical device. This includes submitting a technical file with the user requirements specification, software requirements specification, testing documents, device version history, clinical evidence report, privacy policy, manual tests, instructions for use, and terms and conditions.

In addition, novel software as a medical device aiming to be used in the NHS in the UK is required to meet the Digital Technology Assessment Criteria (DTAC) and DCB 0129 standard.[359, 360] The DTAC aims to give staff, patients, and citizens confidence that the digital health tools they use meet clinical safety, data protection, technical security, interoperability and usability and accessibility standards. DTAC is administered by NHSX, and is the baseline criteria required to gain a listing on the NHS Apps Library from January 2021. The assessment criteria focus on clinical safety, data protection, technical assurance, interoperability, and usability and accessibility. The DCB 0129 standard is issued by NHS Digital, and is mandatory under the Health and Social care Act 2012. Included are the requirement to nominate a clinical safety officer, design and document clinical risk management processes, and carry out risk assessment that is documented in a hazard log and safety case.

I, alongside my supervisor Professor Gale, successfully applied for National Institute for Health and Care Research (NIHR) i4iFAST funding (NIHR204580) to procure the services of Ethos regulatory consultants with the aim to surmount these regulatory barriers. Working with Ethos we understood that FIND-AF would be considered a Class I medical device, in that it uses data from individuals to predict a risk score in healthy populations for a chronic disease that can be managed effectively with interventions that are normally non-invasive.[357] After completing all requirements FIND-AF is now registered with the MHRA as a Class I medical device, and the DCB 0129 standard has also been met.

8.5.2 Implementation within electronic health record systems

An underappreciated barrier to the adoption of prediction models in clinical practice is the lack of integration with EHRs. Some models have been converted to online tools and made available through web-interfaces or mobile applications. To use the models however, a healthcare professional is required to access a website or open an app and manually complete data fields with the patients' details to receive a risk estimation.[361] Though this task may seem trivial compared with the potential added benefit of greater

quality decision-making, the practicalities and time constraints of clinical practice form a significant barrier to usage.[361] This is compounded with the potential of manual transcription errors to lead to incorrect risk estimates.[361] Usability barriers may be mitigated if a healthcare professional can access a prediction model within their local EHR and have a risk score presented automatically as fields are populated with relevant data from within the system. An example of successful prediction model integration is the QRISK models, which have been embedded within UK primary care EHR systems to calculate individual cardiovascular risk based on existing data.[362]

However, the value to EHR providers to cover the costs and risks of integrating prediction models into their EHR system is currently not there, especially as clinical stakeholders do not yet expect such functionalities in EHRs.[363] Thus, the incentives for researchers, EHR providers, and healthcare professionals are unaligned. One proposed solution to align incentives of these different stakeholders is to use Blockchain, a form of distributed ledger technology.[363] Blockchain is an open network of distributed data stored in secure blocks, which are available to all participants (known as 'nodes') on a network.[364] By distributing blocks across all nodes, the data in the network is difficult to hack, change or corrupt, creating a traceable, immutable and secure record of transactions between nodes.[365] Prediction models could be published by researchers to the national marketplace on the blockchain and EHR providers could integrate it into their interface through an application programming interface (API). An API act as a software intermediary to allow the input of data to a prediction model (request), and return of risk prediction (response) to an external application. Clinical data from EHR systems could be entered securely to receive results with a micro-payment triggered at every use, via smart contracts.[363] This would provide a monetary incentive for researchers and Universities and minimise costs for EHR providers of having to formally integrate a prediction model within their system architecture. In the first step to enable interoperability amongst EHR systems I have created the diagnostic codelists relating to each variable in the algorithm in the two 'languages' used across primary care in EHRs: Read codes and Systematised Nomenclature of Medicine Clinical Terms (SNOMED CT).

8.5.3 Acceptability to end users

A systematic review and meta-analysis of 108 studies reporting the absolute improvements in care achieved by computerised clinical decision support systems found only a small increase of 5.8% in the proportion of patients receiving desired

care.[366] The small effect sizes typically achieved by clinical decision support systems may be related to an incomplete recognition of the rich sociotechnical interactions that shape the effectiveness of these solutions,[367] including the human-computer interface, hardware and software computing infrastructure, clinical content, people, workflow and communication, internal organisational culture, external regulations, and system measurement.[366]

User experience is a significant part of successfully implementing a prediction model through EHRs.[368] If a risk score is presented as an alert, healthcare professionals can experience 'alert fatigue',[366] that is, become less responsive to the information provided. Previous reports have demonstrated concerns from healthcare professionals around the impact of the use of prediction models on clinical workflow.[369] If the FIND-AF prediction model reaches the stage of integration within an EHR then the interface through which a healthcare professional would interact with it, and how this would affect their workflow, needs to be carefully designed and tested.[370] To better understand this, I am supervising a qualitative study of semi-structured interviews with healthcare professionals at participating sites in the FIND-AF pilot study to identify obstacles and opportunities of EHR-risk guided AF screening in primary care.

8.5.4 Clinical and cost-effectiveness of risk-guided atrial fibrillation screening

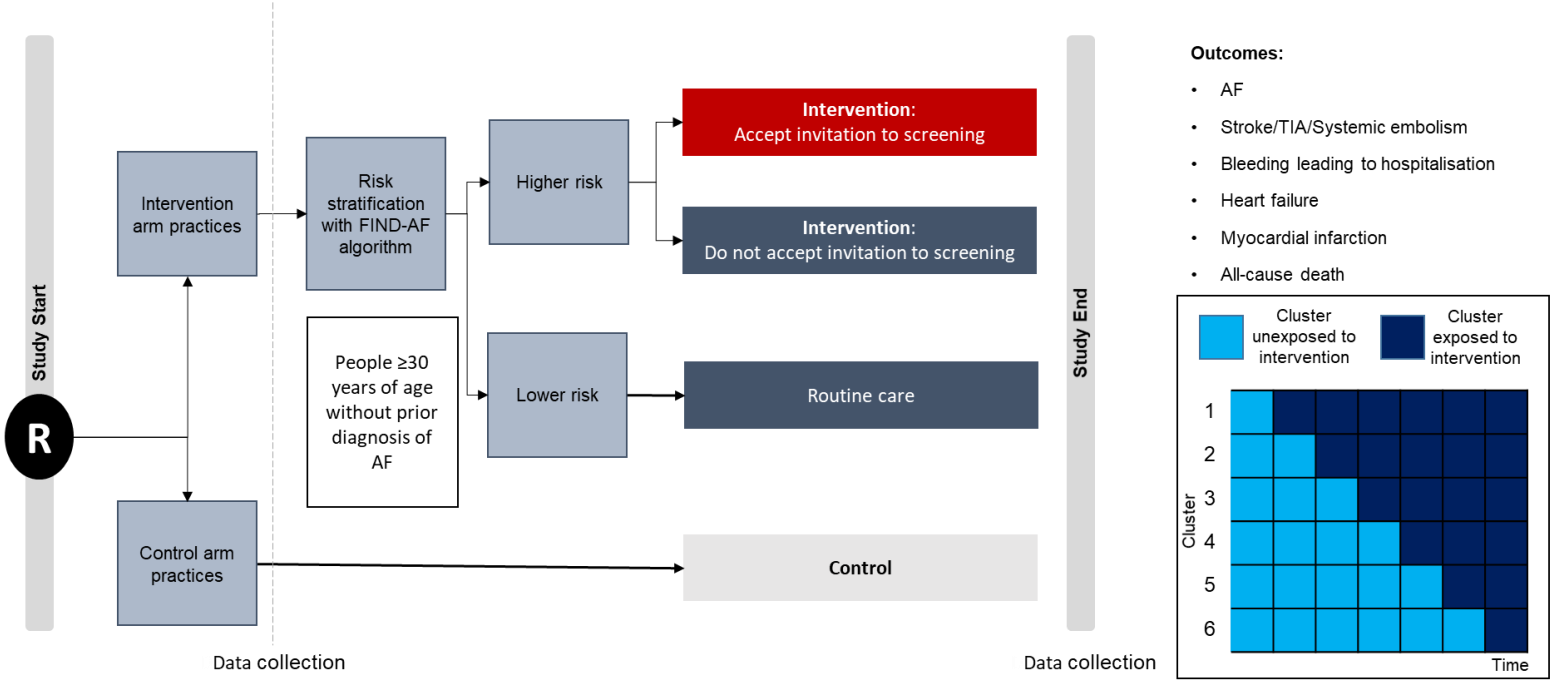
For a prediction model to be used in clinical practice, there needs to be a clear clinical rationale for its use, evidence for clinical and cost-effectiveness of its use, and support from leading professionals in the field or recommendation in guidelines.[105]

A potential barrier for use of FIND-AF in the UK is that screening for AF was not recommended when the UK National Screening Committee last considered the topic in 2019.[99] Further ongoing trials will inform the merits of population screening for AF (Chapter 1, Table 7), and may lead to an alteration in this recommendation. The NHS Long Team Plan aims for early detection and treatment of AF,[98] and improving AF detection is part of the new Investment and Impact Fund incentives in primary care provided by NHS England. Primary care healthcare professionals on the FIND-AF Scientific Advisory Board have stated that they believe the technology would be useful and relevant in their daily practice, irrespective of centrally-mandated screening pathways. Professor Gale and I have presented the FIND-AF prediction model and pilot study protocol to national and regional stakeholders, including the Yorkshire and Humber Cardiovascular Network, Yorkshire and Humber Academic Health Sciences

Network, West Yorkshire Integrated Stroke Delivery Network, the North East and Yorkshire NHS England and NHS Improvement Cardiac Network, the British Heart Rhythm Society, and the Primary Care Cardiovascular Society, and received support and endorsement.

The use of a prediction model can be seen as an intervention that requires preclinical evaluation of its impact on health outcomes and cost effectiveness of care.[105] If the pilot study demonstrates favourable results then a comparative study could be conducted, with one group where usual care is provided and another group in which model predictions guide AF screening.[371] When I, Professor Wu, and Professor Gale have previously discussed this with Professor Amanda Farrin, Professor of Clinical Trials and Evaluation of Complex Interventions at the University of Leeds, a stepped-wedge cluster randomised trial design was considered potentially a suitable design (Figure 2).

Figure 2 Possible design for a FIND-AF stepped-wedge cluster randomised clinical trial



Abbreviations: AF, atrial fibrillation; TIA, transient ischaemic attack

The intervention is the implementation of the RF FIND-AF algorithm followed by systematic screening of people identified as higher risk. Control arm practices would provide routine care (no screening for AF). After an initial period where no clusters are exposed to the intervention at regular intervals one cluster/group of clusters is randomised to cross to intervention under evaluation. At the end of the study there will be a period where all clusters are exposed. Each cluster contributes observations under both control and intervention observation periods. In the short-term the effect on diagnosis

of AF could be assessed, and long-term EHR follow-up could assess the effect on diagnosis of stroke/TIA/systemic embolism, bleeding leading to hospitalisation, and all-cause death.

I and Professor Gale have also discussed with the University of Leeds Academic Unit of Health Economics regarding how a cost-effectiveness analysis may be conducted. They recommended a cost utility analysis and budget impact analysis.[372] They would update and adapt the model used in the UK National Screening Committee evidence review in 2019 for the cost-effectiveness of AF screening in the UK,[96, 373] which consists of an initial decision-tree which captures the screening and diagnosis process followed by a Markov model which captures the expected lifetime costs and benefits associated with treatment. The key comparators in the model would include: 1) no AF screening (current practice), 2) opportunistic AF screening and 3) systematic targeted screening for AF facilitated by FIND-AF. A NHS and Personal Social Services perspective on costs would be adopted and the recommended annual discount rate (currently 3.5%) would be applied to costs and health effects. The comparative life-time costs and benefits would be presented in terms of incremental cost-effectiveness ratios and net health benefit.

8.6 Future directions

In my opinion, key future research questions have emerged during this PhD thesis, which I detail below and intend to explore further as part of future grant applications.

8.6.1 Cluster analysis of the higher predicted atrial fibrillation risk cohort

Individuals with the machine learning-derived EHR phenotype of higher predicted AF risk experience high rates of incident cardio-renal-metabolic diseases and death irrespective of whether they receive AF diagnosis. These individuals represent a heterogeneous group, reflecting a variety of combinations of underlying diseases, which may limit the effectiveness of any one primary prevention or targeted diagnostic strategy to improve outcomes.

Cluster analysis is a ML technique that can be used to classify subjects from heterogeneous populations into cohesive groups based on clinical information.[374] It has been applied to improve characterisation of subphenotypes amongst patients with diagnosed AF.[375-378] However, these groupings may not be applicable to directing upstream interventions before the occurrence of the arrhythmia. Furthermore previous studies have incorporated variables for clustering that are either not available in the

community (e.g. echocardiographic parameters),[375, 376, 379] or highly likely to be missing in routinely-collected records,[376-378] which limits their clinical utility in the primary prevention setting.

Using the existing dataset I could investigate the use of two unsupervised machine learning techniques, hierarchical and K-prototype clustering algorithms, on the higher predicted AF risk cohort. For hierarchical cluster techniques, the Ward minimum variance method of clustering is used to identify patient clusters, given the mixture of binary and continuous variables.[380] The K-prototype clustering method combines the K-means of numerical variables and K-modes of categorical variables to cluster a mixture of continuous and categorical data.[381] Similar to the FIND-AF prediction model I could limit variables for clustering to age, sex, ethnicity and comorbidities, to ensure that the clusters were meaningful and applicable in the community. The association between chosen clusters and non-AF clinical diseases and death investigated in Chapter 5 could be assessed using the unadjusted and adjusted (by CHA_2DS_2-VASc score) Cox proportional hazards models. Kaplan-Meier curves could be plotted for the cumulative incidence curves of events, and the log-rank test used to compare the differences in each cluster. The hypothesis would be that amongst the higher predicted AF risk cohort, there are distinct groups of individuals with shared characteristics and outcomes that could be targeted for specific interventions.

8.6.2 Clinical phenotyping of individuals at higher predicted atrial fibrillation risk

Following on from the research planned in 8.6.1, I wish to determine the extent to which higher predicted AF risk individuals could be suitable for targeted cardiovascular preventive interventions to reduce future cardiovascular events. As a sub-study of the FIND-AF pilot study, consenting participants could be approached and recruited to attend a research phenotyping clinic. At a phenotyping appointment individuals could undergo assessment for cardiovascular risk factors (including BMI, smoking and alcohol consumption), cardio-renal-metabolic workup (including HBA1c, lipid profile, NT-proBNP, urea and electrolytes, and urine:albumin creatinine ratio, and an echocardiogram).

The hypothesis would be that individuals at higher predicted AF risk have lifestyle factors, undiagnosed comorbidities, and sub-optimally treated comorbidities that present opportunities to reduce the risk of future cardiovascular events. An outcome from the research phenotyping clinic appointment would be the documentation of

hitherto undiagnosed cardio-renal-metabolic disease, and advice to the general practitioner regarding treatment optimisation. If it was demonstrated that there was a significant burden of undiagnosed or sub-optimally treated cardiovascular risk factors and diseases, then a trial could be planned to determine the effect on the occurrence of cardio-renal-metabolic disease and death of a multi-modal intervention compared to usual care amongst individuals at higher predicted AF risk.

8.6.3 Mechanistic studies to assess if the higher predicted risk of atrial fibrillation electronic health record phenotype has a pathological correlate

In some patients pulmonary vein triggers may be the predominant pathway leading to AF, but for others AF may also represent a secondary manifestation of a progressive fibrotic atrial cardiomyopathy (FACM).[382] Different expressions can be found categorised as mild (FACM I), moderate (FACM II), or excessive fibrosis (FACM III).[383] The presence of interstitial fibrosis leads to changes in cellular coupling and spatial 'non-uniform anisotropic' impulse propagation, and is a potential cause of atrial activation abnormalities that may underlie the initiation and perpetuation of AF.[384] Atrial fibrosis can be detected, quantified, and localised using delayed-enhancement MRI including four categories of structural changes (Utah Stages: I, 0-5% enhancement; II, >5-20% enhancement; III, >20-35% enhancement; IV >35% enhancement).[385] Atrial fibrosis does not appear to be an age-related process,[386] but atrial remodelling has been observed in individuals with conditions predisposing to AF, but before manifestation of AF, including for mitral stenosis and hypertension.[147, 387]

Moreover, there is an independent correlation between atrial fibrosis and stroke.[388] Studies hypothesise that the underlying atrial myopathy that causes AF can also affect thrombosis risk by modulating the atrial blood flow and/or the haemostatic profile, thereby increasing thromboembolic risk even in the absence of AF.[389] Furthermore, LA enlargement is related to stroke as well as AF. In a meta-analysis of 66,007 participants with 3,549 stroke events, LA enlargement was associated with a 1.68-fold (95% CI 1.36-2.07) increased risk of stroke independent of AF and other comorbidities, with each 1-cm increase in LA diameter increasing the risk of stroke by 24%.[390]

A substudy of the FIND-AF pilot study could recruit age- and sex-matched individuals at lower and higher predicted AF risk, with and without detected AF, to compare the

presence and extent of atrial dilatation and fibrosis by risk score and by AF status. The hypothesis would be that higher risk individuals have a greater degree of atrial remodelling than lower risk individuals. This would provide mechanistic insights into how the EHR phenotype of predicted AF risk may translate to the AF pathological substrate. Furthermore, treatment of the underlying aetiology has been demonstrated to be associated with a significant increase in atrial voltage in individuals with FACM,[391] raising the potential that identification atrial remodelling – dilatation, dysfunction and fibrosis – before the advent of AF may enable intervention to reduce the subsequent risk of AF.

8.6.4 Risk-guided recruitment for trials of primary prevention of atrial fibrillation

The strain the coronavirus disease 2019 (COVID-19) pandemic has placed on healthcare services and resources underscores the importance of pivoting the focus of healthcare to prevention of major adverse events.[392] The rising burden of AF will lead to escalating morbidity, mortality and healthcare use and cost.[11] Therefore I believe that strategies to lower the risk of AF development are urgently needed. Hitherto primary prevention of AF has focussed primarily on reversing modifiable risk factors for AF, and specific upstream therapies have demonstrated disappointing results.[393] International guidelines recommend an angiotensin converting enzyme inhibitor or angiotensin receptor blocker for primary prevention of new-onset AF in patients with HFrEF (Class IIa, level of evidence B), and that they may be considered for patients with hypertension (Class IIb, level of evidence B).[394] Weight loss combined with risk factor modification is recommended for overweight and obese patients with AF (Class I, level of evidence B-Randomised)[395] but lifestyle modifications to address modifiable risk factors for AF before arrhythmia onset remain potential targets.[37]

Preventative methods aimed at high-risk individuals might reduce the burden of AF. Interventions that were ineffective for unselected populations or for individuals with a single morbidity may be effective for individuals who are objectively at high risk of AF when considering multiple factors. Conduct of primary prevention trials for AF has been limited by difficulties in identifying groups at sufficiently high risk.[151] The prediction models developed Chapter 7 in these PhD studies could be used identify individuals for recruitment into primary prevention trials. For example, almost a quarter of individuals classified as very high risk developed clinically-diagnosed AF within 10 years. Moreover, as they are scalable in routinely-collected primary care EHRs in the UK, they could facilitate innovative recruitment strategies.[396, 397]

8.7 Conclusion

This PhD investigated the use of supervised ML in UK primary care EHRs to predict risk of AF within the next six months. In a dataset of over two million individuals, the RF FIND-AF algorithm was highly accurate, and its performance was robust in both sexes and across ethnic groups. In contrast to previously developed prediction models for incident AF, FIND-AF was designed to be scalable in primary care EHRs at the point of implementation. Funding has been successfully applied for to conduct a pilot clinical study of the FIND-AF algorithm and I have formulated the protocol to determine if remote ECG monitoring is associated with higher rates of new AF detection in individuals at higher predicted risk of AF compared to individuals at lower predicted risk of AF. Hitherto, RCTs of systematic population AF screening have resulted in low yields of newly detected AF, which limits clinical- and cost-effectiveness. As such, the outputs of this PhD could address an important knowledge gap to make AF screening more efficient and effective.

This PhD also demonstrates that the ML derived EHR phenotype of higher predicted risk of AF in the short-term is also associated with elevated AF occurrence in the long-term, as well as increased occurrence of incident cardio-renal-metabolic diseases and death. This could inform novel targeted treatment strategies for individuals at risk of AF, rather than just for those who have apparent manifestation of the arrhythmia. Furthermore, this PhD has demonstrated that it is possible to predict AF over both short and long prediction horizons with a small number of routinely-recorded variables, which could be used to recruit participants to trials of primary prevention of AF.

There is strong evidence from these PhD studies that individuals at elevated risk of AF can be identified using data that is routinely-collected at scale in the UK. Future work will be able to establish the external validity of the prediction models in external geographies. Pending positive results from the pilot study a RCT is required to investigate whether risk-guided systematic AF screening is more effective than usual care at detecting AF, and whether this has an impact on adverse events. Furthermore greater characterisation of the higher predicted AF risk cohort, within the existing dataset or with biomarkers and imaging, could uncover insights into how an EHR phenotype translates into pathological and clinical characteristics, and establish whether the EHR phenotype is an actionable target to improve patient outcomes.

References

1. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C*, et al.* 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association 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. European heart journal 2021;**42**(5):373-498.
2. Arvanitis P. Clinical evaluation and implications of left atrial remodeling in atrial fibrillation: From silent cerebral lesions and atrial stunning to novel electrocardiographic tools for prediction of arrhythmia outcome. Acta Universitatis Upsaliensis 2023.
3. Heijman J, Voigt N, Nattel S, Dobrev D. Cellular and molecular electrophysiology of atrial fibrillation initiation, maintenance, and progression. Circulation research 2014;**114**(9):1483-1499.
4. Jansen HJ, Bohne LJ, Gillis AM, Rose RA. Atrial remodeling and atrial fibrillation in acquired forms of cardiovascular disease. Heart Rhythm O2 2020;**1**(2):147-159.
5. Haissaguerre M, Jaïs P, Shah DC, Takahashi A, Hocini M, Quiniou G*, et al.* Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. New England Journal of Medicine 1998;**339**(10):659-666.
6. Jalife J, Berenfeld O, Mansour M. Mother rotors and fibrillatory conduction: a mechanism of atrial fibrillation. Cardiovascular research 2002;**54**(2):204-216.
7. Csepe TA, Hansen BJ, Fedorov VV. Atrial fibrillation driver mechanisms: insight from the isolated human heart. Trends in cardiovascular medicine 2017;**27**(1):1-11.
8. Chen YC, Voskoboinik A, La Gerche A, Marwick TH, McMullen JR. Prevention of pathological atrial remodeling and atrial fibrillation: JACC state-of-the-art review. Journal of the American College of Cardiology 2021;**77**(22):2846-2864.
9. Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: an increasing epidemic and public health challenge. International Journal of Stroke 2021;**16**(2):217-221.
10. Krijthe BP, Kunst A, Benjamin EJ, Lip GY, Franco OH, Hofman A*, et al.* Projections on the number of individuals with atrial fibrillation in the European Union, from 2000 to 2060. European heart journal 2013;**34**(35):2746-2751.
11. Wu J, Nadarajah R, Nakao YM, Nakao K, Wilkinson C, Mamas MA*, et al.* Temporal trends and patterns in atrial fibrillation incidence: A population-based study of 3· 4 million individuals. The Lancet Regional Health-Europe 2022:100386.
12. Cancer Research UK. *Cancer incidence for common cancers*. https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared#heading-Zero.
13. Wolf PA, Abbott RD, Kannel WBJs. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. 1991;**22**(8):983-988.
14. Freedman B, Potpara TS, Lip GY. Stroke prevention in atrial fibrillation. The Lancet 2016;**388**(10046):806-817.
15. Friberg L, Rosenqvist M, Lindgren A, Terént A, Norrving B, Asplund K. High prevalence of atrial fibrillation among patients with ischemic stroke. Stroke 2014;**45**(9):2599-2605.
16. Kishore A, Vail A, Majid A, Dawson J, Lees KR, Tyrrell PJ*, et al.* Detection of atrial fibrillation after ischemic stroke or transient ischemic attack: a systematic review and meta-analysis. Stroke 2014;**45**(2):520-526.
17. Violi F, Pastori D, Pignatelli P. Mechanisms and management of thrombo-embolism in atrial fibrillation. Journal of atrial fibrillation 2014;**7**(3).
18. Watson T, Shantsila E, Lip GY. Mechanisms of thrombogenesis in atrial fibrillation: Virchow's triad revisited. The Lancet 2009;**373**(9658):155-166.
19. Ntaios G, Hart RG. Embolic stroke. Circulation 2017;**136**(25):2403-2405.

20. Staerk L, Sherer JA, Ko D, Benjamin EJ, Helm RH. Atrial fibrillation: epidemiology, pathophysiology, and clinical outcomes. Circulation research 2017;**120**(9):1501-1517.
21. Bekwelem W, Connolly SJ, Halperin JL, Adabag S, Duval S, Chrolavicius S*, et al.* Extracranial systemic embolic events in patients with nonvalvular atrial fibrillation: incidence, risk factors, and outcomes. Circulation 2015;**132**(9):796-803.
22. Diener H-C, Hart RG, Koudstaal PJ, Lane DA, Lip GY. Atrial fibrillation and cognitive function: JACC review topic of the week. Journal of the American College of Cardiology 2019;**73**(5):612-619.
23. Marzona I, O'Donnell M, Teo K, Gao P, Anderson C, Bosch J*, et al.* Increased risk of cognitive and functional decline in patients with atrial fibrillation: results of the ONTARGET and TRANSCEND studies. Cmaj 2012;**184**(6):E329-E336.
24. Rastas S, Verkkoniemi A, Polvikoski T, Juva K, Niinistö L, Mattila K*, et al.* Atrial fibrillation, stroke, and cognition: a longitudinal population-based study of people aged 85 and older. Stroke 2007;**38**(5):1454-1460.
25. de Bruijn RF, Heeringa J, Wolters FJ, Franco OH, Stricker BH, Hofman A*, et al.* Association between atrial fibrillation and dementia in the general population. JAMA neurology 2015;**72**(11):1288-1294.
26. Koh YH, Lew LZ, Franke KB, Elliott AD, Lau DH, Thiyagarajah A*, et al.* Predictive role of atrial fibrillation in cognitive decline: a systematic review and meta-analysis of 2.8 million individuals. EP Europace 2022;**24**(8):1229-1239.
27. Santhanakrishnan R, Wang N, Larson MG, Magnani JW, McManus DD, Lubitz SA*, et al.* Atrial fibrillation begets heart failure and vice versa: temporal associations and differences in preserved versus reduced ejection fraction. Circulation 2016;**133**(5):484-492.
28. Pieske B, Tschöpe C, De Boer RA, Fraser AG, Anker SD, Donal E*, et al.* How to diagnose heart failure with preserved ejection fraction: the HFA–PEFF diagnostic algorithm: a consensus recommendation from the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). European heart journal 2019;**40**(40):3297-3317.
29. Soliman EZ, Safford MM, Muntner P, Khodneva Y, Dawood FZ, Zakai NA*, et al.* Atrial fibrillation and the risk of myocardial infarction. JAMA internal medicine 2014;**174**(1):107-114.
30. Guo Y, Lip GY, Apostolakis S. Inflammation in atrial fibrillation. Journal of the American College of Cardiology 2012;**60**(22):2263-2270.
31. Akar JG, Jeske W, Wilber DJ. Acute onset human atrial fibrillation is associated with local cardiac platelet activation and endothelial dysfunction. Journal of the American College of Cardiology 2008;**51**(18):1790-1793.
32. Odutayo A, Wong CX, Hsiao AJ, Hopewell S, Altman DG, Emdin CA. Atrial fibrillation and risks of cardiovascular disease, renal disease, and death: systematic review and meta-analysis. bmj 2016;**354**.
33. Alkhouli M, Friedman PA. Ischemic stroke risk in patients with nonvalvular atrial fibrillation: JACC review topic of the week. Journal of the American College of Cardiology 2019;**74**(24):3050-3065.
34. Laupacis A, Boysen G, Connolly S, Ezekowitz M, Hart R, James K*, et al.* Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Analysis of pooled data from five randomized controlled trials. Arch Intern Med 1994;**154**(13):1449-1457.
35. Ruff CT, Giugliano RP, Braunwald E, Hoffman EB, Deenadayalu N, Ezekowitz MD*, et al.* Comparison of the efficacy and safety of new oral anticoagulants with warfarin in patients with atrial fibrillation: a meta-analysis of randomised trials. The Lancet 2014;**383**(9921):955-962.
36. National Institute for Health and Care Excellence. *Atrial fibrillation: diagnosis and management NICE guideline [NG196]*. https://www.nice.org.uk/guidance/ng196/chapter/Recommendations#assessment-of-stroke-and-bleeding-risks (30 June 2021; date last accessed).

37. Lau DH, Nattel S, Kalman JM, Sanders P. Modifiable risk factors and atrial fibrillation. Circulation 2017;**136**(6):583-596.
38. Chung MK, Eckhardt LL, Chen LY, Ahmed HM, Gopinathannair R, Joglar JA*, et al.* Lifestyle and risk factor modification for reduction of atrial fibrillation: a scientific statement from the American Heart Association. Circulation 2020;**141**(16):e750-e772.
39. Abed HS, Wittert GA, Leong DP, Shirazi MG, Bahrami B, Middeldorp ME*, et al.* Effect of weight reduction and cardiometabolic risk factor management on symptom burden and severity in patients with atrial fibrillation: a randomized clinical trial. Jama 2013;**310**(19):2050-2060.
40. Pathak RK, Middeldorp ME, Lau DH, Mehta AB, Mahajan R, Twomey D*, et al.* Aggressive risk factor reduction study for atrial fibrillation and implications for the outcome of ablation: the ARREST-AF cohort study. Journal of the American College of Cardiology 2014;**64**(21):2222-2231.
41. Pathak RK, Middeldorp ME, Meredith M, Mehta AB, Mahajan R, Wong CX*, et al.* Long-term effect of goal-directed weight management in an atrial fibrillation cohort: a long-term follow-up study (LEGACY). Journal of the American College of Cardiology 2015;**65**(20):2159-2169.
42. Kim MH, Johnston SS, Chu B-C, Dalal MR, Schulman KL. Estimation of total incremental health care costs in patients with atrial fibrillation in the United States. Circulation: Cardiovascular Quality and Outcomes 2011;**4**(3):313-320.
43. Burdett P, Lip GY. Atrial fibrillation in the UK: predicting costs of an emerging epidemic recognizing and forecasting the cost drivers of atrial fibrillation-related costs. European Heart Journal-Quality of Care and Clinical Outcomes 2022;**8**(2):187-194.
44. Kalarus Z, Mairesse GH, Sokal A, Boriani G, Średniawa B, Casado-Arroyo R*, et al.* Searching for atrial fibrillation: looking harder, looking longer, and in increasingly sophisticated ways. An EHRA position paper. Europace 2023;**25**(1):185-198.
45. Jones NR, Taylor CJ, Hobbs FR, Bowman L, Casadei B. Screening for atrial fibrillation: a call for evidence. European heart journal 2020;**41**(10):1075-1085.
46. Wilson JMG, Jungner G, Organization WH. Principles and practice of screening for disease. 1968.
47. Raffle AE, Gray JM. *Screening: evidence and practice*: Oxford University Press, USA; 2019.
48. UK National Screening Committee. *Population screening explained*. https://www.gov.uk/guidance/population-screening-explained (20 January 2023; date last accessed).
49. Clua-Espuny JL, Lechuga-Duran I, Bosch-Princep R, Roso-Llorach A, Panisello-Tafalla A, Lucas-Noll J*, et al.* Prevalence of undiagnosed atrial fibrillation and of that not being treated with anticoagulant drugs: the AFABE study. Revista Española de Cardiología (English Edition) 2013;**66**(7):545-552.
50. Claes N, Van Laethem C, Goethals M, Goethals P, Mairesse G, Schwagten B*, et al.* Prevalence of atrial fibrillation in adults participating in a large-scale voluntary screening programme in Belgium. Acta cardiologica 2012;**67**(3):273-278.
51. Turakhia MP, Shafrin J, Bognar K, Trocio J, Abdulsattar Y, Wiederkehr D*, et al.* Estimated prevalence of undiagnosed atrial fibrillation in the United States. PloS one 2018;**13**(4):e0195088.
52. Szymanski T, Ashton R, Sekelj S, Petrungaro B, Pollock KG, Sandler B*, et al.* Budget impact analysis of a machine learning algorithm to predict high risk of atrial fibrillation among primary care patients. EP Europace 2022.
53. Tsang T, Barnes M, Pellikka P, Gin K, Miyasaka Y, Seward J*, et al.* 173 Silent atrial fibrillation in olmsted county: a community-based study. Canadian Journal of Cardiology 2011;**27**(5):S122.
54. Siontis KC, Gersh BJ, Killian JM, Noseworthy PA, McCabe P, Weston SA*, et al.* Typical, atypical, and asymptomatic presentations of new-onset atrial fibrillation in the community: characteristics and prognostic implications. Heart Rhythm 2016;**13**(7):1418-1424.

55. Martinez C, Katholing A, Freedman SB. Adverse prognosis of incidentally detected ambulatory atrial fibrillation. Thrombosis and haemostasis 2014;**112**(08):276-286.
56. Freedman B, Martinez C, Katholing A, Rietbrock S. Residual risk of stroke and death in anticoagulant-treated patients with atrial fibrillation. JAMA cardiology 2016;**1**(3):366-368.
57. Freedman B, Camm J, Calkins H, Healey JS, Rosenqvist M, Wang J*, et al.* Screening for atrial fibrillation: a report of the AF-SCREEN International Collaboration. Circulation 2017;**135**(19):1851-1867.
58. Kirchhof P. The future of atrial fibrillation management: integrated care and stratified therapy. The Lancet 2017;**390**(10105):1873-1887.
59. National Institute for Health and Care Excellence. *Atrial fibrillation: diagnosis and management NICE guideline [NG196]*. https://www.nice.org.uk/guidance/ng196/chapter/Recommendations#stroke-prevention (30 June 2021; date last accessed).
60. Harte E, MacLure C, Martin A, Saunders CL, Meads C, Walter FM*, et al.* Reasons why people do not attend NHS Health Checks: a systematic review and qualitative synthesis. British Journal of General Practice 2018;**68**(666):e28-e35.
61. Ghazal F, Theobald H, Rosenqvist M, Al-Khalili F. Validity of daily self-pulse palpation for atrial fibrillation screening in patients 65 years and older: A cross-sectional study. PLoS medicine 2020;**17**(3):e1003063.
62. Clark CE, McDonagh ST, McManus RJ. Accuracy of automated blood pressure measurements in the presence of atrial fibrillation: systematic review and meta-analysis. Journal of human hypertension 2019;**33**(5):352-364.
63. Kyriakoulis KG, Kollias A, Anagnostopoulos I, Gravvani A, Kalogeropoulos P, Destounis A*, et al.* Diagnostic accuracy of a novel cuffless self-blood pressure monitor for atrial fibrillation screening in the elderly. The Journal of Clinical Hypertension 2019;**21**(12):1797-1802.
64. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T*, et al.* Large-scale assessment of a smartwatch to identify atrial fibrillation. New England Journal of Medicine 2019;**381**(20):1909-1917.
65. Guo Y, Wang H, Zhang H, Chen Y, Lip GY. Population-based screening or targeted screening based on initial clinical risk assessment for atrial fibrillation: a report from the Huawei Heart Study. Journal of Clinical Medicine 2020;**9**(5):1493.
66. Rajakariar K, Koshy AN, Sajeev JK, Nair S, Roberts L, Teh AW. Accuracy of a smartwatch based single-lead electrocardiogram device in detection of atrial fibrillation. Heart 2020;**106**(9):665-670.
67. Svennberg E, Engdahl J, Al-Khalili F, Friberg L, Frykman V, Rosenqvist M. Mass screening for untreated atrial fibrillation: the STROKESTOP study. Circulation 2015;**131**(25):2176-2184.
68. Steinhubl SR, Waalen J, Edwards AM, Ariniello LM, Mehta RR, Ebner GS*, et al.* Effect of a home-based wearable continuous ECG monitoring patch on detection of undiagnosed atrial fibrillation: the mSToPS randomized clinical trial. Jama 2018;**320**(2):146-155.
69. Lumikari TJ, Pirinen J, Putaala J, Sibolt G, Kerola A, Pakarinen S*, et al.* Prolonged ECG with a novel recorder utilizing electrode belt and mobile device in patients with recent embolic stroke of undetermined source: A pilot study. Annals of Noninvasive Electrocardiology 2020;**25**(6):e12802.
70. Svennberg E, Friberg L, Frykman V, Al-Khalili F, Engdahl J, Rosenqvist M. Clinical outcomes in systematic screening for atrial fibrillation (STROKESTOP): a multicentre, parallel group, unmasked, randomised controlled trial. Lancet 2021;**398**(10310):1498-1506.
71. Engdahl J, Rosenqvist M. Large-scale screening studies for atrial fibrillation–is it worth the effort? Journal of Internal Medicine 2021;**289**(4):474-492.
72. Svendsen JH, Diederichsen SZ, Højberg S, Krieger DW, Graff C, Kronborg C*, et al.* Implantable loop recorder detection of atrial fibrillation to prevent stroke (The LOOP Study): a randomised controlled trial. The Lancet 2021.

73. Harris K, Edwards D, Mant J. How can we best detect atrial fibrillation? The journal of the Royal College of Physicians of Edinburgh 2012;**42**:5-22.
74. Stergiou G, Karpettas N, Protogerou A, Nasothimiou E, Kyriakidis M. Diagnostic accuracy of a home blood pressure monitor to detect atrial fibrillation. Journal of human hypertension 2009;**23**(10):654-658.
75. Wiesel J, Wiesel D, Suri R, Messineo FC. The use of a modified sphygmomanometer to detect atrial fibrillation in outpatients. Pacing and clinical electrophysiology 2004;**27**(5):639-643.
76. Desteghe L, Raymaekers Z, Lutin M, Vijgen J, Dilling-Boer D, Koopman P*, et al.* Performance of handheld electrocardiogram devices to detect atrial fibrillation in a cardiology and geriatric ward setting. Ep Europace 2017;**19**(1):29-39.
77. Kaasenbrood F, Hollander M, Rutten FH, Gerhards LJ, Hoes AW, Tieleman RG. Yield of screening for atrial fibrillation in primary care with a hand-held, single-lead electrocardiogram device during influenza vaccination. EP Europace 2016;**18**(10):1514-1520.
78. William AD, Kanbour M, Callahan T, Bhargava M, Varma N, Rickard J*, et al.* Assessing the accuracy of an automated atrial fibrillation detection algorithm using smartphone technology: The iREAD Study. Heart rhythm 2018;**15**(10):1561-1565.
79. Brasier N, Raichle CJ, Dörr M, Becke A, Nohturfft V, Weber S*, et al.* Detection of atrial fibrillation with a smartphone camera: first prospective, international, two-centre, clinical validation study (DETECT AF PRO). Ep Europace 2019;**21**(1):41-47.
80. Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ*, et al.* Passive detection of atrial fibrillation using a commercially available smartwatch. JAMA cardiology 2018;**3**(5):409-416.
81. Bumgarner JM, Lambert CT, Hussein AA, Cantillon DJ, Baranowski B, Wolski K*, et al.* Smartwatch algorithm for automated detection of atrial fibrillation. Journal of the American College of Cardiology 2018;**71**(21):2381-2388.
82. NHS. *When you'll be invited for breast screening and who should go*. https://www.nhs.uk/conditions/breast-screening-mammogram/when-youll-be-invited-and-who-should-go/ (10 September 2021; date last accessed).
83. Ghazal F, Theobald H, Rosenqvist M, Al-Khalili F. Feasibility and outcomes of atrial fibrillation screening using intermittent electrocardiography in a primary healthcare setting: A cross-sectional study. PloS one 2018;**13**(5):e0198069.
84. Mant J, Fitzmaurice DA, Hobbs FR, Jowett S, Murray ET, Holder R*, et al.* Accuracy of diagnosing atrial fibrillation on electrocardiogram by primary care practitioners and interpretative diagnostic software: analysis of data from screening for atrial fibrillation in the elderly (SAFE) trial. Bmj 2007;**335**(7616):380.
85. Guo Y, Wang H, Zhang H, Liu T, Liang Z, Xia Y*, et al.* Mobile photoplethysmographic technology to detect atrial fibrillation. Journal of the American College of Cardiology 2019;**74**(19):2365-2375.
86. Hobbs FR, Fitzmaurice D, Mant J, Murray E, Jowett S, Bryan S*, et al.* A randomised controlled trial and cost-effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in people aged 65 and over. The SAFE study. 2005.
87. Gladstone DJ, Wachter R, Schmalstieg-Bahr K, Quinn FR, Hummers E, Ivers N*, et al.* Screening for atrial fibrillation in the older population: a randomized clinical trial. JAMA cardiology 2021;**6**(5):558-567.
88. Halcox JP, Wareham K, Cardew A, Gilmore M, Barry JP, Phillips C*, et al.* Assessment of remote heart rhythm sampling using the AliveCor heart monitor to screen for atrial fibrillation: the REHEARSE-AF study. Circulation 2017;**136**(19):1784-1794.
89. Kahwati LC, Asher GN, Kadro ZO, Keen S, Ali R, Coker-Schwimmer E*, et al.* Screening for atrial fibrillation: updated evidence report and systematic review for the US preventive services task force. JAMA 2022;**327**(4):368-383.
90. Friberg L, Engdahl J, Frykman V, Svennberg E, Levin L-Å, Rosenqvist M. Population screening of 75-and 76-year-old men and women for silent atrial fibrillation (STROKESTOP). Europace 2013;**15**(1):135-140.

91. Van Gelder IC, Healey JS, Crijns HJ, Wang J, Hohnloser SH, Gold MR*, et al.* Duration of device-detected subclinical atrial fibrillation and occurrence of stroke in ASSERT. European heart journal 2017;**38**(17):1339-1344.
92. McIntyre WF, Diederichsen SZ, Freedman B, Schnabel RB, Svennberg E, Healey JS. Screening for atrial fibrillation to prevent stroke: a meta-analysis. European Heart Journal Open 2022;**2**(4):oeac044.
93. Hoare S, Powell A, Modi RN, Armstrong N, Griffin SJ, Mant J*, et al.* Why do people take part in atrial fibrillation screening? Qualitative interview study in English primary care. BMJ open 2022;**12**(3):e051703.
94. Fitzmaurice DA, Hobbs FR, 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**(7616):383.
95. Moran PS, Teljeur C, Ryan M, Smith SM. Systematic screening for the detection of atrial fibrillation. Cochrane database of systematic reviews 2016(6).
96. Welton NJ, McAleenan A, Thom HH, Davies P, Hollingworth W, Higgins JP*, et al.* Screening strategies for atrial fibrillation: a systematic review and cost-effectiveness analysis. Health technology assessment 2017;**21**(29).
97. Lyth J, Svennberg E, Bernfort L, Aronsson M, Frykman V, Al-Khalili F*, et al.* Cost-effectiveness of population screening for atrial fibrillation: the STROKESTOP study. European Heart Journal 2022.
98. NHS. *Cardiovascular disease*. https://www.longtermplan.nhs.uk/areas-of-work/cardiovascular-disease/ (21 August 2019; date last accessed).
99. UK National Screening Committee. *The UK NSC recommendation on Atrial Fibrillation screening in adults*. https://legacyscreening.phe.org.uk/atrialfibrillation#:~:text=The%20UK%20NSC%20does%20not,in%20people%20found%20through%20screening.
100. Svennberg E, Braunschweig F. Screening for atrial fibrillation: less is more? European Heart Journal Open 2021;**1**(3):oeab038.
101. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ*, et al.* Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. Circulation 2014;**129**(8):837-847.
102. Proietti M, Mairesse GH, Goethals P, Scavee C, Vijgen J, Blankoff I*, et al.* A population screening programme for atrial fibrillation: a report from the Belgian Heart Rhythm Week screening programme. Ep Europace 2016;**18**(12):1779-1786.
103. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G*, et al.* Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nature medicine 2019;**25**(1):70-74.
104. Hygrell T, Viberg F, Dahlberg E, Charlton PH, Kemp Gudmundsdottir K, Mant J*, et al.* An artificial intelligence–based model for prediction of atrial fibrillation from single-lead sinus rhythm electrocardiograms facilitating screening. Europace 2023;**25**(4):1332-1338.
105. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S*, et al.* Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS medicine 2013;**10**(2):e1001381.
106. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I*, et al.* Prediction models for cardiovascular disease risk in the general population: systematic review. bmj 2016;**353**.
107. Wyatt JC, Altman DG. Commentary: Prognostic models: clinically useful or quickly forgotten? Bmj 1995;**311**(7019):1539-1541.
108. Riley RD, van der Windt D, Croft P, Moons KG. *Prognosis research in healthcare: concepts, methods, and impact*: Oxford University Press; 2019.
109. Deo RC. Machine learning in medicine. Circulation 2015;**132**(20):1920-1930.
110. Krittanawong C, Johnson KW, Rosenson RS, Wang Z, Aydar M, Baber U*, et al.* Deep learning for cardiovascular medicine: a practical primer. European heart journal 2019;**40**(25):2058-2073.
111. Hagan MT, Demuth HB, Beale M. *Neural network design*: PWS Publishing Co.; 1997.

112. Trayanova NA, Popescu DM, Shade JK. Machine learning in arrhythmia and electrophysiology. Circulation Research 2021;**128**(4):544-566.
113. Sidey-Gibbons JA, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC medical research methodology 2019;**19**(1):1-18.
114. Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. European heart journal 2017;**38**(23):1805-1814.
115. Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV*, et al.* Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. Jama 2001;**285**(18):2370-2375.
116. Kistler PM, Sanders P, Fynn SP, Stevenson IH, Spence SJ, Vohra JK*, et al.* Electrophysiologic and electroanatomic changes in the human atrium associated with age. Journal of the American College of Cardiology 2004;**44**(1):109-116.
117. Psaty BM, Manolio TA, Kuller LH, Kronmal RA, Cushman M, Fried LP*, et al.* Incidence of and risk factors for atrial fibrillation in older adults. Circulation 1997;**96**(7):2455-2461.
118. Benjamin EJ, Levy D, Vaziri SM, D'Agostino RB, Belanger AJ, Wolf PA. Independent risk factors for atrial fibrillation in a population-based cohort: the Framingham Heart Study. Jama 1994;**271**(11):840-844.
119. Heeringa J, van der Kuip DA, Hofman A, Kors JA, van Herpen G, Stricker BHC*, et al.* Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. European heart journal 2006;**27**(8):949-953.
120. Miyasaka Y, Barnes ME, Gersh BJ, Cha SS, Bailey KR, Abhayaratna WP*, et al.* Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. Circulation 2006;**114**(2):119-125.
121. Rodriguez CJ, Soliman EZ, Alonso A, Swett K, Okin PM, Goff Jr DC*, et al.* Atrial fibrillation incidence and risk factors in relation to race-ethnicity and the population attributable fraction of atrial fibrillation risk factors: the Multi-Ethnic Study of Atherosclerosis. Annals of epidemiology 2015;**25**(2):71-76. e1.
122. Noubiap Nzeale JJ. Screening, risk stratification, and management of atrial fibrillation 2023.
123. Vaziri SM, Larson MG, Benjamin EJ, Levy D. Echocardiographic predictors of nonrheumatic atrial fibrillation. The Framingham Heart Study. Circulation 1994;**89**(2):724-730.
124. Soliman EZ, Alonso A, Goff DC. Atrial fibrillation and ethnicity: the known, the unknown and the paradox. Future cardiology 2009;**5**(6):547-556.
125. Marcus GM, Alonso A, Peralta CA, Lettre G, Vittinghoff E, Lubitz SA*, et al.* European ancestry as a risk factor for atrial fibrillation in African Americans. Circulation 2010;**122**(20):2009-2015.
126. Roberts JD, Hu D, Heckbert SR, Alonso A, Dewland TA, Vittinghoff E*, et al.* Genetic investigation into the differential risk of atrial fibrillation among black and white individuals. JAMA cardiology 2016;**1**(4):442-450.
127. Christophersen IE, Ravn LS, Budtz-Joergensen E, Skytthe A, Haunsoe S, Svendsen JH*, et al.* Familial aggregation of atrial fibrillation: a study in Danish twins. Circulation: Arrhythmia and Electrophysiology 2009;**2**(4):378-383.
128. Fox CS, Parise H, D'Agostino Sr RB, Lloyd-Jones DM, Vasan RS, Wang TJ*, et al.* Parental atrial fibrillation as a risk factor for atrial fibrillation in offspring. Jama 2004;**291**(23):2851-2855.
129. Roselli C, Rienstra M, Ellinor PT. Genetics of atrial fibrillation in 2020: GWAS, genome sequencing, polygenic risk, and beyond. Circulation research 2020;**127**(1):21-33.
130. Azarbal F, Stefanick ML, Salmoirago-Blotcher E, Manson JE, Albert CM, LaMonte MJ*, et al.* Obesity, physical activity, and their interaction in incident atrial fibrillation in postmenopausal women. Journal of the American Heart Association 2014;**3**(4):e001127.

131. Aizer A, Gaziano JM, Cook NR, Manson JE, Buring JE, Albert CM. Relation of vigorous exercise to risk of atrial fibrillation. The American journal of cardiology 2009;**103**(11):1572-1577.
132. Pelliccia A, Maron BJ, Di Paolo FM, Biffi A, Quattrini FM, Pisicchio C*, et al.* Prevalence and clinical significance of left atrial remodeling in competitive athletes. Journal of the American College of Cardiology 2005;**46**(4):690-696.
133. Huang J-L, Wen Z-C, Lee W-L, Chang M-S, Chen S-A. Changes of autonomic tone before the onset of paroxysmal atrial fibrillation. International journal of cardiology 1998;**66**(3):275-283.
134. Heeringa J, Kors JA, Hofman A, van Rooij FJ, Witteman JC. Cigarette smoking and risk of atrial fibrillation: the Rotterdam Study. American heart journal 2008;**156**(6):1163-1169.
135. Fröhlich M, Sund M, Löwel H, Imhof A, Hoffmeister A, Koenig W. Independent association of various smoking characteristics with markers of systemic inflammation in men: results from a representative sample of the general population (MONICA Augsburg Survey 1994/95). European heart journal 2003;**24**(14):1365-1372.
136. Stewart P, Catterall J. Chronic nicotine ingestion and atrial fibrillation. British heart journal 1985;**54**(2):222.
137. Tu SJ, Gallagher C, Elliott AD, Linz D, Pitman BM, Hendriks JM*, et al.* Risk thresholds for total and beverage-specific alcohol consumption and incident atrial fibrillation. Clinical Electrophysiology 2021;**7**(12):1561-1569.
138. Preedy V, Siddiq T, Why H, Richardson P. The deleterious effects of alcohol on the heart: involvement of protein turnover. Alcohol and Alcoholism 1994;**29**(2):141-147.
139. Wong CX, Sullivan T, Sun MT, Mahajan R, Pathak RK, Middeldorp M*, et al.* Obesity and the risk of incident, post-operative, and post-ablation atrial fibrillation: a meta-analysis of 626,603 individuals in 51 studies. JACC: clinical electrophysiology 2015;**1**(3):139-152.
140. Abed HS, Samuel CS, Lau DH, Kelly DJ, Royce SG, Alasady M*, et al.* Obesity results in progressive atrial structural and electrical remodeling: implications for atrial fibrillation. Heart rhythm 2013;**10**(1):90-100.
141. Mahajan R, Lau DH, Brooks AG, Shipp NJ, Manavis J, Wood JP*, et al.* Electrophysiological, electroanatomical, and structural remodeling of the atria as consequences of sustained obesity. Journal of the American College of Cardiology 2015;**66**(1):1-11.
142. Mahajan R, Lau DH, Brooks AG, Shipp NJ, Wood JP, Manavis J*, et al.* Atrial fibrillation and obesity: reverse remodeling of atrial substrate with weight reduction. Clinical Electrophysiology 2021;**7**(5):630-641.
143. Lau DH, Schotten U, Mahajan R, Antic NA, Hatem SN, Pathak RK*, et al.* Novel mechanisms in the pathogenesis of atrial fibrillation: practical applications. European heart journal 2016;**37**(20):1573-1581.
144. Hatem SN, Sanders P. Epicardial adipose tissue and atrial fibrillation. Cardiovascular research 2014;**102**(2):205-213.
145. Kannel WB, Abbott RD, Savage DD, McNamara PM. Epidemiologic features of chronic atrial fibrillation. The Framingham study. New England Journal of Medicine 1982;**306**(17):1018-1022.
146. Verdecchia P, Reboldi G, Gattobigio R, Bentivoglio M, Borgioni C, Angeli F*, et al.* Atrial fibrillation in hypertension: predictors and outcome. hypertension 2003;**41**(2):218-223.
147. Medi C, Kalman JM, Spence SJ, Teh AW, Lee G, Bader I*, et al.* Atrial electrical and structural changes associated with longstanding hypertension in humans: implications for the substrate for atrial fibrillation. Journal of cardiovascular electrophysiology 2011;**22**(12):1317-1324.
148. Yiu K, Tse H. Hypertension and cardiac arrhythmias: a review of the epidemiology, pathophysiology and clinical implications. Journal of human hypertension 2008;**22**(6):380-388.
149. Wachtell K, Lehto M, Gerdts E, Olsen MH, Hornestam B, Dahlöf B*, et al.* Angiotensin II receptor blockade reduces new-onset atrial fibrillation and subsequent

stroke compared to atenolol: the Losartan Intervention For End Point Reduction in Hypertension (LIFE) study. Journal of the American College of Cardiology 2005;**45**(5):712-719.
150. Okin PM, Devereux RB, Jern S, Kjeldsen SE, Julius S, Nieminen MS*, et al.* Regression of electrocardiographic left ventricular hypertrophy during antihypertensive treatment and the prediction of major cardiovascular events. Jama 2004;**292**(19):2343-2349.
151. Benjamin EJ, Chen P-S, Bild DE, Mascette AM, Albert CM, Alonso A*, et al.* Prevention of atrial fibrillation: report from a national heart, lung, and blood institute workshop. Circulation 2009;**119**(4):606-618.
152. Stevenson WG, Stevenson LW. Atrial fibrillation in heart failure. In: Mass Medical Soc; 1999, 910-911.
153. Wang TJ, Larson MG, Levy D, Vasan RS, Leip EP, Wolf PA*, et al.* Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: the Framingham Heart Study. Circulation 2003;**107**(23):2920-2925.
154. Huxley RR, Lopez FL, Folsom AR, Agarwal SK, Loehr LR, Soliman EZ*, et al.* Absolute and attributable risks of atrial fibrillation in relation to optimal and borderline risk factors: the Atherosclerosis Risk in Communities (ARIC) study. Circulation 2011;**123**(14):1501-1508.
155. Kotecha D, Piccini JP. Atrial fibrillation in heart failure: what should we do? European heart journal 2015;**36**(46):3250-3257.
156. Lau DH, Psaltis PJ, Mackenzie L, Kelly DJ, Carbone A, Worthington M*, et al.* Atrial Remodeling in an Ovine Model of Anthracycline-Induced Nonischemic Cardiomyopathy: Remodeling of the Same Sort. Journal of cardiovascular electrophysiology 2011;**22**(2):175-182.
157. Ohtani K, Yutani C, Nagata S, Koretsune Y, Hori M, Kamada T. High prevalence of atrial fibrosis in patients with dilated cardiomyopathy. Journal of the American College of Cardiology 1995;**25**(5):1162-1169.
158. Lee JZ, Cha Y-M. Atrial fibrillation and heart failure: A contemporary review of current management approaches. Heart Rhythm O2 2021;**2**(6):762-770.
159. Lau DH, Alasady M, Brooks AG, Sanders P. New-onset atrial fibrillation and acute coronary syndrome. Expert review of cardiovascular therapy 2010;**8**(7):941-948.
160. Amar D, Shi W, Hogue CW, Zhang H, Passman RS, Thomas B*, et al.* Clinical prediction rule for atrial fibrillation after coronary artery bypass grafting. Journal of the American College of Cardiology 2004;**44**(6):1248-1253.
161. Kinjo K, Sato H, Sato H, Ohnishi Y, Hishida E, Nakatani D*, et al.* Prognostic significance of atrial fibrillation/atrial flutter in patients with acute myocardial infarction treated with percutaneous coronary intervention. The American journal of cardiology 2003;**92**(10):1150-1154.
162. Hirayama A, Goto T, Shimada YJ, Faridi MK, Camargo Jr CA, Hasegawa K. Acute exacerbation of chronic obstructive pulmonary disease and subsequent risk of emergency department visits and hospitalizations for atrial fibrillation. Circulation: Arrhythmia and Electrophysiology 2018;**11**(9):e006322.
163. Agustí A, Hogg JC. Update on the pathogenesis of chronic obstructive pulmonary disease. New England Journal of Medicine 2019;**381**(13):1248-1256.
164. Hiram R, Naud P, Xiong F, Al-u'datt Da, Algalarrondo V, Sirois MG*, et al.* Right atrial mechanisms of atrial fibrillation in a rat model of right heart disease. Journal of the American College of Cardiology 2019;**74**(10):1332-1347.
165. Lip GY, Fauchier L, Freedman SB, Van Gelder I, Natale A, Gianni C*, et al.* Atrial fibrillation (Primer). Nature Reviews: Disease Primers 2016;**2**(1).
166. Linz D, McEvoy RD, Cowie MR, Somers VK, Nattel S, Lévy P*, et al.* Associations of obstructive sleep apnea with atrial fibrillation and continuous positive airway pressure treatment: a review. JAMA cardiology 2018;**3**(6):532-540.
167. Aune D, Feng T, Schlesinger S, Janszky I, Norat T, Riboli E. Diabetes mellitus, blood glucose and the risk of atrial fibrillation: a systematic review and meta-analysis of cohort studies. Journal of Diabetes and its Complications 2018;**32**(5):501-511.

168. Rutter MK, Parise H, Benjamin EJ, Levy D, Larson MG, Meigs JB*, et al.* Impact of glucose intolerance and insulin resistance on cardiac structure and function: sex-related differences in the Framingham Heart Study. Circulation 2003;**107**(3):448-454.
169. Odegaard AO, Jacobs DR, Sanchez OA, Goff DC, Reiner AP, Gross MD. Oxidative stress, inflammation, endothelial dysfunction and incidence of type 2 diabetes. Cardiovascular diabetology 2016;**15**(1):1-12.
170. Liu C, Fu H, Li J, Yang W, Cheng L, Liu T*, et al.* Hyperglycemia aggravates atrial interstitial fibrosis, ionic remodeling and vulnerability to atrial fibrillation in diabetic rabbits/Hiperglisemi diyabetik tavsanlarda atriyal interstisiyel fibrosis, iyonik remodeling ve atriyal fibrilasyon duyarliligini arttirmaktadir. The Anatolian Journal of Cardiology 2012;**12**(7):543.
171. Bernier M, Marelli AJ, Pilote L, Bouchardy J, Bottega N, Martucci G*, et al.* Atrial arrhythmias in adult patients with right-versus left-sided congenital heart disease anomalies. The American journal of cardiology 2010;**106**(4):547-551.
172. Rangnekar G. The characterisation of risk factors, substrate and management strategies for atrial fibrillation 2015.
173. Kirsh JA, Walsh EP, Triedman JK. Prevalence of and risk factors for atrial fibrillation and intra-atrial reentrant tachycardia among patients with congenital heart disease. American Journal of Cardiology 2002;**90**(3):338-340.
174. Iung B, Vahanian A. Epidemiology of valvular heart disease in the adult. Nature Reviews Cardiology 2011;**8**(3):162-172.
175. Diker E, Aydogdu S, Özdemir M, Kural T, Polat K, Cehreli S*, et al.* Prevalence and predictors of atrial fibrillation in rheumatic valvular heart disease. The American journal of cardiology 1996;**77**(1):96-98.
176. Bailey GW, Braniff BA, Hancock EW, Cohn KE. Relation of left atrial pathology to atrial fibrillation in mitral valvular disease. Annals of internal medicine 1968;**69**(1):13-20.
177. Widgren V, Dencker M, Juhlin T, Platonov P, Willenheimer R. Aortic stenosis and mitral regurgitation as predictors of atrial fibrillation during 11 years of follow-up. BMC cardiovascular disorders 2012;**12**(1):1-8.
178. Aronow WS. Valvular aortic stenosis in the elderly. Cardiology in review 2007;**15**(5):217-225.
179. Dalsgaard M, Egstrup K, Wachtell K, Gerdts E, Cramariuc D, Kjaergaard J*, et al.* Left atrial volume in patients with asymptomatic aortic valve stenosis (the Simvastatin and Ezetimibe in Aortic Stenosis study). The American journal of cardiology 2008;**101**(7):1030-1034.
180. Bansal N, Zelnick LR, Alonso A, Benjamin EJ, de Boer IH, Deo R*, et al.* eGFR and albuminuria in relation to risk of incident atrial fibrillation: a meta-analysis of the Jackson Heart Study, the Multi-Ethnic Study of Atherosclerosis, and the Cardiovascular Health Study. Clinical Journal of the American Society of Nephrology 2017;**12**(9):1386-1398.
181. Martin FL, McKie PM, Cataliotti A, Sangaralingham SJ, Korinek J, Huntley BK*, et al.* Experimental mild renal insufficiency mediates early cardiac apoptosis, fibrosis, and diastolic dysfunction: a kidney-heart connection. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 2012;**302**(2):R292-R299.
182. Huang S-Y, Chen Y-C, Kao Y-H, Hsieh M-H, Chen Y-A, Chen W-P*, et al.* Renal failure induces atrial arrhythmogenesis from discrepant electrophysiological remodeling and calcium regulation in pulmonary veins, sinoatrial node, and atria. International journal of cardiology 2016;**202**:846-857.
183. Himmelreich JC, Veelers L, Lucassen WA, Schnabel RB, Rienstra M, van Weert HC*, et al.* Prediction models for atrial fibrillation applicable in the community: a systematic review and meta-analysis. EP Europace 2020;**22**(5):684-694.
184. 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**.
185. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. Bmj 2011;**342**.

186. Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, Hooft L*, et al.* A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Statistical methods in medical research 2019;**28**(9):2768-2786.
187. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. JAMA Cardiol. 2001;**285**(22):2864-2870.
188. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. Chest 2010;**137**(2):263-272.
189. de Vos CB, Pisters R, Nieuwlaat R, Prins MH, Tieleman RG, Coelen R-JS*, et al.* Progression from paroxysmal to persistent atrial fibrillation: clinical correlates and prognosis. Journal of the American College of Cardiology 2010;**55**(8):725-731.
190. 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 Medicine 2014;**11**(10):e1001744.
191. Linker DT, Murphy TB, Mokdad AH. Selective screening for atrial fibrillation using multivariable risk models. Heart 2018;**104**(18):1492-1499.
192. 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**(29):2243-51.
193. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS*, et al.* PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Annals of internal medicine 2019;**170**(1):51-58.
194. Hill NR, Ayoubkhani D, McEwan P, Sugrue DM, Farooqui U, Lister S*, et al.* Predicting atrial fibrillation in primary care using machine learning. PLoS One 2019;**14**(11):e0224582.
195. Lloyd-Jones DM. Cardiovascular risk prediction: basic concepts, current status, and future directions. Circulation 2010;**121**(15):1768-1777.
196. Himmelreich JC, Lucassen WA, Harskamp RE, Aussems C, van Weert HC, Nielen MM. 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**(1):e001459.
197. Nadarajah R, Alsaeed E, Hurdus B, Aktaa S, Hogg D, Bates MG*, et al.* Prediction of incident atrial fibrillation in community-based electronic health records: a systematic review with meta-analysis. Heart 2021.
198. Nielsen JC, Lin Y-J, de Oliveira Figueiredo MJ, Sepehri Shamloo A, Alfie A, Boveda S*, et al.* European Heart Rhythm Association (EHRA)/Heart Rhythm Society (HRS)/Asia Pacific Heart Rhythm Society (APHRS)/Latin American Heart Rhythm Society (LAHRS) expert consensus on risk assessment in cardiac arrhythmias: use the right tool for the right outcome, in the right population. EP Europace 2020;**22**(8):1147-1148.
199. Li YG, Pastori D, Farcomeni A, Yang PS, Jang E, Joung B*, et al.* A Simple Clinical Risk Score (C2HEST) 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**(3):510-518.
200. Lip GY, Skjøth F, Nielsen PB, Larsen TB. Evaluation of the C2HEST risk score as a possible opportunistic screening tool for incident atrial fibrillation in a healthy population (from a nationwide Danish cohort study). The American Journal of Cardiology 2020;**125**(1):48-54.
201. Digital N. *Appointments in General Practice January, 2019*. https://digital.nhs.uk/data-and-information/publications/statistical/appointments-in-general-practice/january-2019.
202. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Van Staa T*, et al.* Data resource profile: clinical practice research datalink (CPRD). International journal of epidemiology 2015;**44**(3):827-836.

203. European Commission. *eHealth adoption in primary healthcare in the EU is on the rise*. https://digital-strategy.ec.europa.eu/en/library/ehealth-adoption-primary-healthcare-eu-rise#:~:text=The%20study%20analysed%20different%20eHealth,use%20it%20in%20their%20practice. (18 June 2019; date last accessed).
204. Kolek MJ, Graves AJ, Xu M, Bian A, Teixeira PL, Shoemaker MB*, et al.* Evaluation of a prediction model for the development of atrial fibrillation in a repository of electronic medical records. JAMA cardiology 2016;**1**(9):1007-1013.
205. 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**(2):e000102.
206. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC medical research methodology 2014;**14**(1):1-13.
207. Mahmoudi E, Kamdar N, Kim N, Gonzales G, Singh K, Waljee AK. Use of electronic medical records in development and validation of risk prediction models of hospital readmission: systematic review. BMJ 2020;**369**.
208. Clegg A, Bates C, Young J, Ryan R, Nichols L, Ann Teale E*, et al.* Development and validation of an electronic frailty index using routine primary care electronic health record data. Age and ageing 2016;**45**(3):353-360.
209. National Institute for Health and Care Excellence. *Multimorbidity: clinical assessment and management*. https://www.nice.org.uk/guidance/ng56.
210. tpp. *UK National Frailty Index*. https://tpp-uk.com/news/uk-national-frailty-index/.
211. Kozieł M, Teutsch C, Halperin JL, Rothman KJ, Diener H-C, Ma C-S*, et al.* Atrial fibrillation and comorbidities: Clinical characteristics and antithrombotic treatment in GLORIA-AF. PloS one 2021;**16**(4):e0249524.
212. Piccini JP, Hammill BG, Sinner MF, Hernandez AF, Walkey AJ, Benjamin EJ*, et al.* Clinical course of atrial fibrillation in older adults: the importance of cardiovascular events beyond stroke. European heart journal 2014;**35**(4):250-256.
213. Haroun MK, Jaar BG, Hoffman SC, Comstock GW, Klag MJ, Coresh J. Risk factors for chronic kidney disease: a prospective study of 23,534 men and women in Washington County, Maryland. Journal of the American Society of Nephrology 2003;**14**(11):2934-2941.
214. Smith Jr SC. Multiple risk factors for cardiovascular disease and diabetes mellitus. The American journal of medicine 2007;**120**(3):S3-S11.
215. Cipolletta E, Tata LJ, Nakafero G, Avery AJ, Mamas MA, Abhishek A. Association between gout flare and subsequent cardiovascular events among patients with gout. JAMA 2022;**328**(5):440-450.
216. Camm AJ, Amarenco P, Haas S, Hess S, Kirchhof P, Kuhls S*, et al.* XANTUS: a real-world, prospective, observational study of patients treated with rivaroxaban for stroke prevention in atrial fibrillation. European heart journal 2016;**37**(14):1145-1153.
217. Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. Annals of internal medicine 2007;**146**(12):857-867.
218. Public Health England. *Technical document for subnational English atrial fibrillation prevalence estimates: Application of age–sex rates in a Swedish region to the English population*. https://www.gov.uk/government/publications/atrial-fibrillation-prevalence-estimates-forlocal-populations.
219. United Kingdom National Screening Committee. *The UK NSC recommendation on Atrial Fibrillation screening in adults*. https://legacyscreening.phe.org.uk/atrialfibrillation.
220. Page M, McKenzie J, Bossuyt P, Boutron I, Hoffmann T, Mulrow C. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. MetaArXiv. 2020. In; 2020.

221. Poorthuis MH, Jones NR, Sherliker P, Clack R, de Borst GJ, Clarke R*, et al.* Utility of risk prediction models to detect atrial fibrillation in screened participants. European Journal of Preventive Cardiology 2020;**28**(6):586-595.
222. Ammenwerth E, Neyer S, Hörbst A, Mueller G, Siebert U, Schnell-Inderst PJ. Adult patient access to electronic health records. Cochrane Database of Systematic Reviews 2021(2).
223. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Systematic reviews 2016;**5**(1):1-10.
224. He Y, Ong Y, Li X, Din FV, Brown E, Timofeeva M*, et al.* Performance of prediction models on survival outcomes of colorectal cancer with surgical resection: a systematic review and meta-analysis. Surgical oncology 2019;**29**:196-202.
225. Snell KI, Ensor J, Debray TP, Moons KG, Riley RD. Meta-analysis of prediction model performance across multiple studies: Which scale helps ensure between-study normality for the C-statistic and calibration measures? Statistical Methods in Medical Research 2018;**27**(11):3505-3522.
226. Debray T, Jong Vd. metamisc: Meta-Analysis of Diagnosis and Prognosis Research Studies. In; 2020.
227. R Core Team. R: A Language and Environment for Statistical Computing. In: R Foundation for Statistical Computing; 2019.
228. Viechtbauer W. Conducting meta-analyses in R with the metafor package. Journal of Statistical Software 2010;**36**:1–48.
229. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P*, et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;**336**(7650):924-926.
230. Van Remoortel H, Scheers H, De Buck E, Haenen W, Vandekerckhove P. Prediction modelling studies for medical usage rates in mass gatherings: A systematic review. PLOS One 2020;**15**(6):e0234977.
231. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. Journal of clinical epidemiology 1995;**48**(12):1503-1510.
232. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Statistics in medicine 2016;**35**(2):214-226.
233. Pencina MJ, D'Agostino RB. Evaluating discrimination of risk prediction models: the C statistic. JAMA 2015;**314**(10):1063-1064.
234. 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**(9):1556-1563.
235. Chao TF, Liu CJ, Chen SJ, Wang KL, Lin YJ, Chang SL*, et al.* CHADS2 score and risk of new-onset atrial fibrillation: a nationwide cohort study in Taiwan. Int J Cardiol 2013;**168**(2):1360-3.
236. Hu W-S, Hsieh M-H, Lin C-L. A novel atrial fibrillation prediction model for Chinese subjects: a nationwide cohort investigation of 682 237 study participants with random forest model. EP Europace 2019;**21**(9):1307-1312.
237. Hu W-S, Lin C-L. Prediction of new-onset atrial fibrillation for general population in Asia: A comparison of C2HEST and HATCH scores. International journal of cardiology 2020;**313**:60-63.
238. Saliba W, Gronich N, Barnett-Griness O, Rennert 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**(8):843-9.
239. Sekelj S, Sandler B, Johnston E, Pollock KG, Hill NR, Gordon J*, et al.* Detecting undiagnosed atrial fibrillation in UK primary care: Validation of a machine learning prediction algorithm in a retrospective cohort study. Eur J Prev Cardiol 2020:2047487320942338.

240. 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**(1):e5597.
241. Kim I-S, Yang P-S, Jang E, Jung H, You SC, Yu HT*, et al.* Long-term PM 2.5 exposure and the clinical application of machine learning for predicting incident atrial fibrillation. Scientific reports 2020;**10**(1):1-11.
242. 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. European Journal of Preventive Cardiology 2015;**22**(7):932-9.
243. 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). Journal of the American Heart Association 2016;**5**(2):e003077.
244. Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB*, et al.* Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. The Lancet 2009;**373**(9665):739-745.
245. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient-versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Statistics in medicine 2002;**21**(3):371-387.
246. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A*, et al.* External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC medical research methodology 2014;**14**(1):1-11.
247. Morrison A, Polisena J, Husereau 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. International journal of technology assessment in health care 2012;**28**(2):138.
248. Gladstone DJ, Sharma M, Spence JD. Cryptogenic stroke and atrial fibrillation. The New England journal of medicine 2014;**371**(13):1260-1260.
249. Kirchhof P, Camm AJ, Goette A, Brandes A, Eckardt L, Elvan A*, et al.* Early rhythm-control therapy in patients with atrial fibrillation. New England Journal of Medicine 2020;**383**(14):1305-1316.
250. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C*, et al.* 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association 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. 2021;**42**(5):373-498.
251. Kemp Gudmundsdottir K, Fredriksson T, Svennberg E, Al-Khalili F, Friberg L, Frykman V*, et al.* Stepwise mass screening for atrial fibrillation using N-terminal B-type natriuretic peptide: the STROKESTOP II study. EP Europace 2020;**22**(1):24-32.
252. Himmelreich JC, Lucassen WA, Harskamp RE, Aussems C, van Weert HC, Nielen MMJOH. 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. 2021;**8**(1):e001459.
253. 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**(9):683-690.
254. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. British journal of clinical pharmacology 2010;**69**(1):4-14.
255. Chisholm J. The Read clinical classification. BMJ: British Medical Journal 1990;**300**(6732):1092.

256. Hill NR, Ayoubkhani D, McEwan P, Sugrue DM, Farooqui U, Lister S*, et al.* Predicting atrial fibrillation in primary care using machine learning. PloS one 2019;**14**(11):e0224582.
257. Arbel R, Hammerman A, Sergienko R, Friger M, Peretz A, Netzer D*, et al.* BNT162b2 vaccine booster and mortality due to Covid-19. New England Journal of Medicine 2021;**385**(26):2413-2420.
258. Arbel R, Sergienko R, Friger M, Peretz A, Beckenstein T, Yaron S*, et al.* Effectiveness of a second BNT162b2 booster vaccine against hospitalization and death from COVID-19 in adults aged over 60 years. Nature medicine 2022;**28**(7):1486-1490.
259. Arbel R, Wolff Sagy Y, Hoshen M, Battat E, Lavie G, Sergienko R*, et al.* Nirmatrelvir use and severe Covid-19 outcomes during the Omicron surge. New England Journal of Medicine 2022;**387**(9):790-798.
260. Hammerman A, Sergienko R, Friger M, Beckenstein T, Peretz A, Netzer D*, et al.* Effectiveness of the BNT162b2 vaccine after recovery from Covid-19. New England Journal of Medicine 2022;**386**(13):1221-1229.
261. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomstrom-Lundqvist C*, et al.* 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association of Cardio-Thoracic Surgery (EACTS). Eur Heart J 2020.
262. Riley RD, Snell KI, Ensor J, Burke DL, Harrell Jr FE, Moons KG*, et al.* Minimum sample size for developing a multivariable prediction model: PART II-binary and time-to-event outcomes. Statistics in medicine 2019;**38**(7):1276-1296.
263. Cowan JC, Wu J, Hall M, Orlowski A, West RM, Gale CP. A 10 year study of hospitalized atrial fibrillation-related stroke in England and its association with uptake of oral anticoagulation. European heart journal 2018;**39**(32):2975-2983.
264. Wu J, Alsaeed ES, Barrett J, Hall M, Cowan C, Gale CP. Prescription of oral anticoagulants and antiplatelets for stroke prophylaxis in atrial fibrillation: nationwide time series ecological analysis. EP Europace 2020;**22**(9):1311-1319.
265. Schnabel RB, Witt H, Walker J, Ludwig M, Geelhoed B, Kossack N*, et al.* Machine learning-based identification of risk-factor signatures for undiagnosed atrial fibrillation in primary prevention and post-stroke in clinical practice. European Heart Journal-Quality of Care and Clinical Outcomes 2023;**9**(1):16-23.
266. Chao TF, Chiang CE, Chen TJ, Liao JN, Tuan TC, Chen SA. Clinical Risk Score for the Prediction of Incident Atrial Fibrillation: Derivation in 7 220 654 Taiwan Patients With 438 930 Incident Atrial Fibrillations During a 16-Year Follow-Up. Journal of the American Heart Association 2021;**10**(17):e020194.
267. Routen A, Akbari A, Banerjee A, Katikireddi SV, Mathur R, McKee M*, et al.* Strategies to record and use ethnicity information in routine health data. Nature Medicine 2022:1-4.
268. Groenwold RH. Informative missingness in electronic health record systems: the curse of knowing. Diagnostic and prognostic research 2020;**4**(1):1-6.
269. Breiman L. Random forests. Machine learning 2001;**45**(1):5-32.
270. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ*, et al.* An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. The Lancet 2019;**394**(10201):861-867.
271. Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. Dordrecht, The Netherlands: D. Reidel 1986;**81**(10.5555):26853.
272. Conrad N, Judge A, Canoy D, Tran J, Pinho-Gomes A-C, Millett ER*, et al.* Temporal trends and patterns in mortality after incident heart failure: a longitudinal analysis of 86 000 individuals. JAMA cardiology 2019;**4**(11):1102-1111.
273. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M*, et al.* 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special

contribution of the Heart Failure Association (HFA) of the ESC. European heart journal 2021;**42**(36):3599-3726.
274. Vahanian A, Beyersdorf F, Praz F, Milojevic M, Baldus S, Bauersachs J*, et al.* 2021 ESC/EACTS Guidelines for the management of valvular heart disease: developed by the Task Force for the management of valvular heart disease of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS). European heart journal 2022;**43**(7):561-632.
275. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S*, et al.* Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med 2013;**10**(2):e1001381.
276. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) The TRIPOD Statement. Circulation 2015;**131**(2):211-219.
277. Nicholls SG, Quach P, von Elm E, Guttmann A, Moher D, Petersen I*, et al.* The reporting of studies conducted using observational routinely-collected health data (RECORD) statement: methods for arriving at consensus and developing reporting guidelines. PLoS One 2015;**10**(5):e0125620.
278. Kotecha D, Asselbergs FW, Achenbach S, Anker SD, Atar D, Baigent C*, et al.* CODE-EHR best practice framework for the use of structured electronic healthcare records in clinical research. bmj 2022;**378**.
279. Kramer DB, Xu S, Kesselheim AS. Regulation of medical devices in the United States and European Union. In. *The Ethical Challenges of Emerging Medical Technologies*: Taylor and Francis; 2020, 41-49.
280. Ruigómez A, Johansson S, Wallander M-A, Rodríguez LAGa. Incidence of chronic atrial fibrillation in general practice and its treatment pattern. Journal of clinical epidemiology 2002;**55**(4):358-363.
281. Raileanu LE, Stoffel K. Theoretical comparison between the gini index and information gain criteria. Annals of Mathematics and Artificial Intelligence 2004;**41**(1):77-93.
282. Van Smeden M, Heinze G, Van Calster B, Asselbergs FW, Vardas PE, Bruining N*, et al.* Critical appraisal of artificial intelligence-based prediction models for cardiovascular disease. European Heart Journal 2022;**43**(31):2921-2930.
283. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. The Lancet Digital Health 2021;**3**(4):e260-e265.
284. Middeldorp ME, Pathak RK, Meredith M, Mehta AB, Elliott AD, Mahajan R*, et al.* PREVEntion and regReSsive Effect of weight-loss and risk factor modification on Atrial Fibrillation: the REVERSE-AF study. EP Europace 2018;**20**(12):1929-1935.
285. Uittenbogaart SB, Verbiest-van Gurp N, Lucassen WA, Winkens B, Nielen M, Erkens PM*, et al.* Opportunistic screening versus usual care for detection of atrial fibrillation in primary care: cluster randomised controlled trial. bmj 2020;**370**.
286. Nadarajah R, Wu J, Hogg D, Raveendra K, Nakao YM, Nakao K*, et al.* Prediction of short-term atrial fibrillation risk using primary care electronic health records. Heart 2023.
287. Padmanabhan S, Carty L, Cameron E, Ghosh RE, Williams R, Strongman H. Approach to record linkage of primary care data from Clinical Practice Research Datalink to other health-related patient data: overview and implications. European journal of epidemiology 2019;**34**(1):91-99.
288. Liyanage H, Liaw S-T, de Lusignan S. Reporting of Studies Conducted using Observational Routinely Collected Data (RECORD) Statement: call for contributions from the clinical informatics community. Informatics in primary care 2012;**20**(4):221-224.
289. Bonow RO, Greenland P. Population-wide trends in aortic stenosis incidence and outcomes. In: Am Heart Assoc; 2015, 969-971.
290. Hulme OL, Khurshid S, Weng LC, Anderson CD, Wang EY, Ashburner JM*, et al.* Development and Validation of a Prediction Model for Atrial Fibrillation Using Electronic Health Records. JACC. Clinical Electrophysiology 2019;**5**(11):1331-1341.

291. Robson J, Dostal I, Sheikh A, Eldridge S, Madurasinghe V, Griffiths C*, et al.* The NHS Health Check in England: an evaluation of the first 4 years. BMJ open 2016;**6**(1):e008840.
292. Dalton AR, Soljak M, Samarasundera E, Millett C, Majeed A. Prevalence of cardiovascular disease risk amongst the population eligible for the NHS Health Check Programme. European journal of preventive cardiology 2013;**20**(1):142-150.
293. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S*, et al.* Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. New England Journal of Medicine 2015;**373**(22):2117-2128.
294. EMPA-KIDNEY Collaborative Group. Empagliflozin in patients with chronic kidney disease. New England Journal of Medicine 2022.
295. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A*, et al.* Dapagliflozin and cardiovascular outcomes in type 2 diabetes. New England Journal of Medicine 2019;**380**(4):347-357.
296. Ledwidge M, Gallagher J, Conlon C, Tallon E, O'Connell E, Dawkins I*, et al.* Natriuretic peptide–based screening and collaborative care for heart failure: the STOP-HF randomized trial. Jama 2013;**310**(1):66-74.
297. Lyth J, Svennberg E, Bernfort L, Aronsson M, Frykman V, Al-Khalili F*, et al.* Cost-effectiveness of population screening for atrial fibrillation: the STROKESTOP study. European Heart Journal 2023;**44**(3):196-204.
298. Noseworthy PA, Attia ZI, Behnken EM, Giblon RE, Bews KA, Liu S*, et al.* Artificial intelligence-guided screening for atrial fibrillation using electrocardiogram during sinus rhythm: a prospective non-randomised interventional trial. The Lancet 2022;**400**(10359):1206-1212.
299. Hill NR, Groves L, Dickerson C, Ochs A, Pang D, Lawton S*, et al.* Identification of undiagnosed atrial fibrillation using a machine learning risk-prediction algorithm and diagnostic testing (PULsE-AI) in primary care: a multi-centre randomized controlled trial in England. European Heart Journal - Digital Health 2022;**3**(2):195-204.
300. Mant J, Burt J, Griffin S, Hobbs R, McManus R, Williams K*, et al.* An RCT to determine if screening for atrial fibrillation reduces stroke and mortality: SAFER trial - Screening for Atrial Fibrillation with ECG to Reduce stroke. In: *Royal College of General Practicioners*. *2020*.
301. Zenicor Medical Systems. *Zenicor-ECG*. https://zenicor.com/zenicor-ekg/.
302. National Institute of Health and Care Excellence. *Lead-I ECG devices for detecting symptomatic atrial fibrillation using single time point testing in primary care*. https://www.nice.org.uk/guidance/dg35.
303. Barrett PM, Komatireddy R, Haaser S, Topol S, Sheard J, Encinas J*, et al.* Comparison of 24-hour Holter monitoring with 14-day novel adhesive patch electrocardiographic monitoring. The American journal of medicine 2014;**127**(1):95. e11-95. e17.
304. Svennberg E, Stridh M, Engdahl J, Al-Khalili F, Friberg L, Frykman V*, et al.* Safe automatic 1-lead ECG screening for atrial fibrillation. In: *Congress of the European-Society-of-Cardiology (ESC) Conference Location Rome, ITALY*. *2016*, p. 689-689. OXFORD UNIV PRESS Location OXFORD.
305. Wijesurendra R, Pessoa-Amorim G, Bulbulia R, Jones N, A'court C, Taylor K*, et al.* Active monitoring for atrial fibrillation (AMALFI): a streamlined randomised controlled trial of remote screening for subclinical atrial fibrillation. Europace 2023;**25**(Supplement_1):euad122. 052.
306. Nadarajah R, Wu J, Frangi AF, Hogg D, Cowan C, Gale CP. What is next for screening for undiagnosed atrial fibrillation? Artificial intelligence may hold the key. European Heart Journal-Quality of Care and Clinical Outcomes 2021.
307. Wu J, Nadarajah R, Nakao YM, Nakao K, Wilkinson C, Mamas M*, et al.* Temporal trends and patterns in atrial fibrillation incidence: a population-based study of 3.4 million individuals. Lancet Regional Health - Europe 2022.
308. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. Journal of the American statistical association 1979;**74**(368):829-836.

309. Chamberlain 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**(1):85-91.
310. Lau Y, Yiu K, Siu C, Tse H. Hypertension and atrial fibrillation: epidemiology, pathophysiology and therapeutic implications. Journal of human hypertension 2012;**26**(10):563-569.
311. Staerk L, Wang B, Preis SR, Larson MG, Lubitz SA, Ellinor PT*, et al.* Lifetime risk of atrial fibrillation according to optimal, borderline, or elevated levels of risk factors: cohort study based on longitudinal data from the Framingham Heart Study. Bmj 2018;**361**.
312. Li YG, Bisson A, Bodin A, Herbert J, Grammatico-Guillon L, Joung B*, et al.* C2 HEST score and prediction of incident atrial fibrillation in poststroke patients: a French nationwide study. Journal of the American Heart Association 2019;**8**(13):e012546.
313. Segan L, Canovas R, Nanayakkara S, Chieng D, Prabhu S, Voskoboinik A*, et al.* New-onset atrial fibrillation prediction: the HARMS2-AF risk score. European Heart Journal 2023:ehad375.
314. Hammond-Haley M, Providencia R, Lambiase PD. Temporal pattern/episode duration-based classification of atrial fibrillation as paroxysmal vs. persistent: is it time to develop a more integrated prognostic score to optimize management? EP Europace 2018;**20**(FI_3):f288-f298.
315. Ooka T, Johno H, Nakamoto K, Yoda Y, Yokomichi H, Yamagata Z. Random forest approach for determining risk prediction and predictive factors of type 2 diabetes: large-scale health check-up data in Japan. BMJ Nutrition, Prevention & Health 2021:bmjnph-2020-000200.
316. Williams K, Modi RN, Dymond A, Hoare S, Powell A, Burt J*, et al.* Cluster randomised controlled trial of screening for atrial fibrillation in people aged 70 years and over to reduce stroke: protocol for the pilot study for the SAFER trial. BMJ open 2022;**12**(9):e065066.
317. Engler D, Hanson CL, Desteghe L, Boriani G, Diederichsen SZ, Freedman B*, et al.* Feasible approaches and implementation challenges to atrial fibrillation screening: a qualitative study of stakeholder views in 11 European countries. BMJ open 2022;**12**(6):e059156.
318. Engler D, Heidbuchel H, Schnabel RB, Investigators A-E. Digital, risk-based screening for atrial fibrillation in the European community—the AFFECT-EU project funded by the European Union. European heart journal 2021.
319. Brik T, Lucassen WA, Harskamp RE, Karregat EP, Himmelreich JC, Busschers WB*, et al.* Personalized approach using wearable technology for early detection of atrial fibrillation in high-risk primary care patients (PATCH-AF): study protocol for a cluster randomized controlled trial. American Heart Journal 2022;**254**:172-182.
320. Marston L, Carpenter JR, Walters KR, Morris RW, Nazareth I, Petersen I. Issues in multiple imputation of missing data for large general practice clinical databases. Pharmacoepidemiology and drug safety 2010;**19**(6):618-626.
321. Verberne LD, Nielen MM, Leemrijse CJ, Verheij RA, Friele RD. Recording of weight in electronic health records: an observational study in general practice. BMC Family Practice 2018;**19**:1-8.
322. Firman N, Robson J, Ahmed Z, Boomla K, Dezateux C. Completeness and representativeness of body mass index in children's electronic general practice records: Linked cross-sectional study in an ethnically-diverse urban population in the United Kingdom. Pediatric Obesity 2021;**16**(8):e12772.
323. Petersen I, Welch CA, Nazareth I, Walters K, Marston L, Morris RW*, et al.* Health indicator recording in UK primary care electronic health records: key implications for handling missing data. Clinical epidemiology 2019:157-167.
324. Collins GS, Dhiman P, Navarro CLA, Ma J, Hooft L, Reitsma JB*, et al.* Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ open 2021;**11**(7):e048008.

325. Barker J, Li X, Khavandi S, Koeckerling D, Mavilakandy A, Pepper C*, et al.* Machine learning in sudden cardiac death risk prediction: a systematic review. Europace 2022;**24**(11):1777-1787.
326. Hulme OL, Khurshid S, Weng L-C, Anderson CD, Wang EY, Ashburner JM*, et al.* Development and validation of a prediction model for atrial fibrillation using electronic health records. JACC Clinical Electrophysiology 2019;**5**(11):1331-1341.
327. Guthrie B, Payne K, Alderson P, McMurdo ME, Mercer SW. Adapting clinical guidelines to take account of multimorbidity. Bmj 2012;**345**:e6341.
328. Beles M, Masuy I, Verstreken S, Bartunek J, Dierckx R, Heggermont W*, et al.* Cardio–renal–metabolic syndrome: clinical features and dapagliflozin eligibility in a real-world heart failure cohort. ESC Heart Failure 2023.
329. Connell AW, Sowers JR. The CardioRenal metabolic syndrome. Journal of the American Society of Hypertension: JASH 2014;**8**(8):604.
330. Wu J, Nadarajah R, Nakao YM, Nakao K, Wilkinson C, Cowan JC*, et al.* Temporal trends and patterns in cause-specific mortality and hospitalisations after incident atrial fibrillation: a longitudinal analysis of 72 000 individuals European Heart Journal 2023.
331. Lee KK, Doudesis D, Anwar M, Astengo F, Chenevier-Gobeaux C, Claessens Y-E*, et al.* Development and validation of a decision support tool for the diagnosis of acute heart failure: systematic review, meta-analysis, and modelling study. bmj 2022;**377**.
332. Faizal ASM, Thevarajah TM, Khor SM, Chang S-W. A review of risk prediction models in cardiovascular disease: conventional approach vs. artificial intelligent approach. Computer methods and programs in biomedicine 2021;**207**:106190.
333. Li Y, Rao S, Solares JRA, Hassaine A, Ramakrishnan R, Canoy D*, et al.* BEHRT: transformer for electronic health records. Scientific reports 2020;**10**(1):1-12.
334. Solares JRA, Raimondi FED, Zhu Y, Rahimian F, Canoy D, Tran J*, et al.* Deep learning for electronic health records: A comparative review of multiple deep neural architectures. Journal of biomedical informatics 2020;**101**:103337.
335. Rasmy L, Wu Y, Wang N, Geng X, Zheng WJ, Wang F*, et al.* A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set. Journal of biomedical informatics 2018;**84**:11-16.
336. Buhrmester V, Münch D, Arens M. Analysis of explainers of black box deep neural networks for computer vision: A survey. arXiv preprint arXiv:1911.12116 2019.
337. Hoogland J, van Barreveld M, Debray TP, Reitsma JB, Verstraelen TE, Dijkgraaf MG*, et al.* Handling missing predictor values when validating and applying a prediction model to new patients. Statistics in medicine 2020;**39**(25):3591-3607.
338. Tsvetanova A, Sperrin M, Peek N, Buchan I, Hyland S, Martin GP. Missing data was handled inconsistently in UK prediction models: a review of method used. Journal of Clinical Epidemiology 2021;**140**:149-158.
339. MDCalc. *Framingham coronary heart disease risk score*. https://www.mdcalc.com/framingham-coronary-heart-disease-risk-score.
340. University of Washington. *Seattle heart failure model*. https://depts.washington.edu/shfm/index.php?width=1920&height=1080.
341. QRISK3. https://qrisk.org/three/.
342. Saar-Tsechansky M, Provost F. Handling missing values when applying classification models. 2007.
343. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Statistics in medicine 2004;**23**(7):1111-1130.
344. Berkelmans GF, Read SH, Gudbjörnsdottir S, Wild SH, Franzen S, Van Der Graaf Y*, et al.* Population median imputation was noninferior to complex approaches for imputing missing values in cardiovascular prediction models in clinical practice. Journal of Clinical Epidemiology 2022;**145**:70-80.
345. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clinical Kidney Journal 2021;**14**(1):49-58.

346. Bleeker S, Moll H, Steyerberg Ea, Donders A, Derksen-Lubsen G, Grobbee D*, et al.* External validation is necessary in prediction research:: A clinical example. Journal of clinical epidemiology 2003;**56**(9):826-832.
347. de Jong Y, Ramspek CL, van der Endt VH, Rookmaaker MB, Blankestijn PJ, Vernooij RW*, et al.* A systematic review and external validation of stroke prediction models demonstrates poor performance in dialysis patients. Journal of clinical epidemiology 2020;**123**:69-79.
348. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. Journal of clinical epidemiology 2016;**69**:245-247.
349. Riley RD, Ensor J, Snell KI, Debray TP, Altman DG, Moons KG*, et al.* External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. bmj 2016;**353**.
350. Mokgokong R, Schnabel R, Witt H, Miller R, Lee TC. Performance of an electronic health record-based predictive model to identify patients with atrial fibrillation across countries. Plos one 2022;**17**(7):e0269867.
351. Collerton J, Davies K, Jagger C, Kingston A, Bond J, Eccles MP*, et al.* Health and disease in 85 year olds: baseline findings from the Newcastle 85+ cohort study. Bmj 2009;**339**.
352. Goldstein BA, Bhavsar NA, Phelan M, Pencina MJ. Controlling for informed presence bias due to the number of health encounters in an electronic health record. American journal of epidemiology 2016;**184**(11):847-855.
353. Ahmad FS, Chan C, Rosenman MB, Post WS, Fort DG, Greenland P*, et al.* Validity of cardiovascular data from electronic sources: the multi-ethnic study of atherosclerosis and HealthLNK. Circulation 2017;**136**(13):1207-1216.
354. Yahya T, Jilani MH, Khan SU, Mszar R, Hassan SZ, Blaha MJ*, et al.* Stroke in young adults: Current trends, opportunities for prevention and pathways forward. American journal of preventive cardiology 2020;**3**:100085.
355. Harrington J, Piccini JP, Alexander JH, Granger CB, Patel MR. Clinical evaluation of factor XIa inhibitor drugs: JACC review topic of the week. Journal of the American College of Cardiology 2023;**81**(8):771-779.
356. Lane DA, Skjøth F, Lip GY, Larsen TB, Kotecha D. Temporal trends in incidence, prevalence, and mortality of atrial fibrillation in primary care. Journal of the American Heart Association 2017;**6**(5):e005155.
357. International Medical Device Regulators Forum. *"Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations*. https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf.
358. The Secretary of State. *The Medical Devices Regulations*. https://www.legislation.gov.uk/uksi/2002/618/contents/made.
359. NHS England - Transformation Directorate. *Digital Technology Assessment Criteria (DTAC)*.
360. NHS Digital. *DCB0129: Clinical Risk Management: its Application in the Manufacture of Health IT Systems*. https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/dcb0129-clinical-risk-management-its-application-in-the-manufacture-of-health-it-systems.
361. Sharma V, Ali I, van der Veer S, Martin G, Ainsworth J, Augustine T. Adoption of clinical risk prediction tools is limited by a lack of integration with electronic health records. BMJ Health & Care Informatics 2021;**28**(1).
362. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A*, et al.* Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. Bmj 2008;**336**(7659):1475-1482.
363. Sharma V, Davies A, Ainsworth J. Clinical risk prediction models: the canary in the coalmine for artificial intelligence in healthcare? BMJ Health & Care Informatics 2021;**28**(1).

364. Leeming G, Ainsworth J, Clifton DA. Blockchain in health care: hype, trust, and digital health. The Lancet 2019;**393**(10190):2476-2477.
365. Cunningham J, Ainsworth J. Enabling patient control of personal electronic health records through distributed ledger technology. Stud Health Technol Inform 2018;**245**:45-48.
366. Kwan JL, Lo L, Ferguson J, Goldberg H, Diaz-Martinez JP, Tomlinson G*, et al.* Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. Bmj 2020;**370**.
367. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. BMJ Quality & Safety 2010;**19**(Suppl 3):i68-i74.
368. Vermeeren AP, Law EL-C, Roto V, Obrist M, Hoonhout J, Väänänen-Vainio-Mattila K. User experience evaluation methods: current state and development needs. In: *Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries*. *2010*, p. 521-530.
369. Takamine L, Forman J, Damschroder LJ, Youles B, Sussman J. Understanding providers' attitudes and key concerns toward incorporating CVD risk prediction into clinical practice: a qualitative study. BMC Health Services Research 2021;**21**:1-11.
370. Gov.uk Sevice Manual. *User research*. https://www.gov.uk/service-manual/user-research.
371. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Annals of internal medicine 2006;**144**(3):201-209.
372. National Institute for Health and Care Excellence. *Evidence standards framework (ESF) for digital health technologies*. https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies#ai-update.
373. King S, Fitzgerald A, Bartlett C, Mahon J, Arber M, Carr E*, et al. Evidence Summary for Screening for Atrial Fibrillation in Adults*. https://view-health-screening-recommendations.service.gov.uk/atrial-fibrillation/.
374. Ahmad T, Pencina MJ, Schulte PJ, O'Brien E, Whellan DJ, Piña IL*, et al.* Clinical implications of chronic heart failure phenotypes defined by cluster analysis. Journal of the American College of Cardiology 2014;**64**(17):1765-1774.
375. Inohara T, Piccini JP, Mahaffey KW, Kimura T, Katsumata Y, Tanimoto K*, et al.* A cluster analysis of the Japanese Multicenter Outpatient Registry of patients with atrial fibrillation. The American Journal of Cardiology 2019;**124**(6):871-878.
376. Ogawa H, An Y, Nishi H, Fukuda S, Ishigami K, Ikeda S*, et al.* Characteristics and clinical outcomes in atrial fibrillation patients classified using cluster analysis: the Fushimi AF Registry. EP Europace 2021;**23**(9):1369-1379.
377. Watanabe E, Inoue H, Atarashi H, Okumura K, Yamashita T, Kodani E*, et al.* Clinical phenotypes of patients with non-valvular atrial fibrillation as defined by a cluster analysis: A report from the J-RHYTHM registry. IJC Heart & Vasculature 2021;**37**:100885.
378. Pastori D, Antonucci E, Milanese A, Menichelli D, Palareti G, Farcomeni A*, et al.* Clinical Phenotypes of Atrial Fibrillation and Mortality Risk—A Cluster Analysis from the Nationwide Italian START Registry. Journal of Personalized Medicine 2022;**12**(5):785.
379. Inohara T, Shrader P, Pieper K, Blanco RG, Thomas L, Singer DE*, et al.* Association of of atrial fibrillation clinical phenotypes with treatment patterns and outcomes: a multicenter registry study. JAMA cardiology 2018;**3**(1):54-63.
380. Vitolo M, Proietti M, Shantsila A, Boriani G, Lip GY. Clinical phenotype classification of atrial fibrillation patients using cluster analysis and associations with trial-adjudicated outcomes. Biomedicines 2021;**9**(7):843.
381. Saito Y, Omae Y, Nagashima K, Miyauchi K, Nishizaki Y, Miyazaki S*, et al.* Phenotyping of atrial fibrillation with cluster analysis and external validation. Heart 2023.

382. Kottkamp H. Fibrotic atrial cardiomyopathy: a specific disease/syndrome supplying substrates for atrial fibrillation, atrial tachycardia, sinus node disease, AV node disease, and thromboembolic complications. Journal of cardiovascular electrophysiology 2012;**23**(7):797-799.
383. Kottkamp H. Human atrial fibrillation substrate: towards a specific fibrotic atrial cardiomyopathy. European heart journal 2013;**34**(35):2731-2738.
384. Spach MS, Boineau JP. Microfibrosis produces electrical load variations due to loss of side-to-side cell connections; a major mechanism of structural heart disease arrhythmias. Pacing and clinical electrophysiology 1997;**20**(2):397-413.
385. Oakes RS, Badger TJ, Kholmovski EG, Akoum N, Burgon NS, Fish EN*, et al.* Detection and quantification of left atrial structural remodeling with delayed-enhancement magnetic resonance imaging in patients with atrial fibrillation. Circulation 2009;**119**(13):1758-1767.
386. Platonov PG, Mitrofanova LB, Orshanskaya V, Ho SY. Structural abnormalities in atrial walls are associated with presence and persistency of atrial fibrillation but not with age. Journal of the American College of Cardiology 2011;**58**(21):2225-2232.
387. Jørgensen HS, Nakayama H, Reith J, Raaschou HO, Olsen TS. Acute stroke with atrial fibrillation: the Copenhagen Stroke Study. Stroke 1996;**27**(10):1765-1769.
388. Daccarett M, Badger TJ, Akoum N, Burgon NS, Mahnkopf C, Vergara G*, et al.* Association of left atrial fibrosis detected by delayed-enhancement magnetic resonance imaging and the risk of stroke in patients with atrial fibrillation. Journal of the American College of Cardiology 2011;**57**(7):831-838.
389. Chung MK, Refaat M, Shen W-K, Kutyifa V, Cha Y-M, Di Biase L*, et al.* Atrial fibrillation: JACC council perspectives. Journal of the American College of Cardiology 2020;**75**(14):1689-1713.
390. Xu Y, Zhao L, Zhang L, Han Y, Wang P, Yu S. Left atrial enlargement and the risk of stroke: a meta-analysis of prospective cohort studies. Frontiers in Neurology 2020;**11**:26.
391. John B, Stiles MK, Kuklik P, Brooks AG, Chandy ST, Kalman JM*, et al.* Reverse remodeling of the atria after treatment of chronic stretch in humans: implications for the atrial fibrillation substrate. Journal of the American College of Cardiology 2010;**55**(12):1217-1226.
392. Nadarajah R, Wu J, Hurdus B, Asma S, Bhatt DL, Biondi-Zoccai G*, et al.* The collateral damage of COVID-19 to cardiovascular services: a meta-analysis. European Heart Journal 2022;**43**(33):3164-3178.
393. Savelieva I, Kakouros N, Kourliouros A, Camm AJ. Upstream therapies for management of atrial fibrillation: review of clinical evidence and implications for European Society of Cardiology guidelines. Part I: primary prevention. Europace 2011;**13**(3):308-328.
394. January CT, Wann LS, Alpert JS, Calkins H, Cigarroa JE, Cleveland JC*, et al.* 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 Practice Guidelines and the Heart Rhythm Society. Journal of the American College of Cardiology 2014;**64**(21):e1-e76.
395. John B, Stiles MK, Kuklik P, Chandy ST, Young GD, Mackenzie L*, et al.* Electrical remodelling of the left and right atria due to rheumatic mitral stenosis. European heart journal 2008;**29**(18):2234-2243.
396. Herrett E, Tomlin K, Lin L-Y, Tomlinson LA, Jit M, Briggs A*, et al.* Protocol for an OpenSAFELY cohort study collecting patient-reported outcome measures using the TPP Airmid smartphone application and linked big data to quantify the health and economic costs of long COVID (OpenPROMPT). BMJ open 2023;**13**(2):e071261.
397. Khan MS, Usman MS, Talha KM, Van Spall HG, Greene SJ, Vaduganathan M*, et al.* Leveraging electronic health records to streamline the conduct of cardiovascular clinical trials. European Heart Journal 2023;**44**(21):1890-1909.