National Institute for Health and Care Excellence

Final

Atrial fibrillation: diagnosis and management

Evidence reviews C and D: Tools to predict stroke in people with atrial fibrillation

NICE guideline NG196

Evidence reviews

April 2021

Final

Developed by the National Guideline Centre, Royal College of Physicians

Atrial fibrillation

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN: 978-1-4731-4043-1

Contents

1 Introduction

The risk of stroke caused by thromboembolism is up to 20% higher in patients with atrial fibrillation. The risk increases in the presence of additional risk factors, such as, hypertension, diabetes and high cholesterol.

Risk stratification tools help to predict the risk of embolic stroke in patients with atrial fibrillation and the presence of these other cardiovascular risks. The tools help to identify the risk of multiple risk factors, and based on this information, the clinician and patient can decide if the patient will benefit from anti-coagulation (e.g. DOAC or Vitamin K antagonists).

However Vitamin K antagonists and DOACs are not without risk. They increase the risk of bleeding, particularly in the elderly; hence the use of tools to predict the bleeding risk in patients exposed to these medications is also important. Knowing the predicted benefit of reducing the risk of stroke as well as the increased risk of bleeding helps the clinician and patient to make an informed decision about whether to use these anti-coagulants. The tools also help to discuss the recommendation with patients.

This chapter will outline the best tools available to assess the risk-benefit ratio of anti-coagulation in patients with atrial fibrillation. The high cost of the newer oral anti-coagulants in comparison to Vitamin K antagonists need to also be taken into account and a cost benefit analysis is presented. This is presented in two parts: a review of the clinical effectiveness of the tools, followed by a review of the accuracy of the tools.

2 Effectiveness of tools to predict stroke or thromboembolic events in people with atrial fibrillation

2.1 Review question: What is the most clinically and cost-effective risk stratification tool for predicting stroke or thromboembolic events in people with atrial fibrillation?

2.2 PICO table

For full details see the review protocol in appendix A.

Table 1: PICO characteristics of review question

Population	People aged over 18 with a diagnosis of AF.
Intervention(s)	Any stroke risk tool (for example, ABC stroke score, Q stroke, ATRIA, Troponin 2, CHADS2). Any version of CHADS2VASC with modifications [Note: **treat each test using a different threshold as a separate intervention**].
Comparison(s)	CHADS2VASC (the established method, as recommended by previous version of this guideline)
Outcomes	Critical • health-related quality of life • mortality • stroke or thromboembolic complications • major bleeding
Study design	Randomised controlled trials

2.3 Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual.[88] Methods specific to this review question are described in the review protocol in appendix A.

This review is not a 'prognostic accuracy' review, but is instead a review of trials that have compared later health outcomes in people randomised to different prediction tools. Tools with differing prognostic accuracies may differ in their influence on later health outcomes through stimulating a more or less appropriate treatment approach. Whilst accuracy is not measured directly in such randomised trials, the advantage of such studies is that they demonstrate clinical efficacy. In contrast a prognostic accuracy study can only demonstrate the intrinsic predictive accuracy of the tool and is unable to show that the accuracy affects health outcomes. However such randomised trials are not commonly undertaken, and may provide equivocal results, and so a prognostic accuracy review has also been undertaken (section 3).

Declarations of interest were recorded according to NICE's 2018 conflicts of interest policy.

2.4 Clinical evidence

2.4.1 Included studies

No relevant clinical studies comparing different stroke risk tools with CHADS2VASC were identified.

2.4.2 Excluded studies

See the excluded studies list in appendix I.

2.4.3 Summary of clinical studies included in the evidence review

No evidence identified.

2.4.4 Quality assessment of clinical studies included in the evidence review

No evidence identified.

2.5 Economic evidence

2.5.1 Included studies

No relevant health economic studies were identified.

2.5.2 Excluded studies

No health economic studies that were relevant to this question were excluded due to assessment of limited applicability or methodological limitations.

See also the health economic study selection flow chart in appendix G.

2.6 The committee's discussion of the evidence

No evidence was generated by this review. The committee discussed the predictive accuracy evidence (see section 3 below only, as this was felt to be sufficient to inform recommendations relevant to the most appropriate methods to predict stroke in people with AF, without the need for any consensus recommendations or research recommendations pertaining to this review.

3 Accuracy of tools to predict stroke or thromboembolic events

3.1 Review question: What is the most accurate risk stratification tool for predicting stroke or thromboembolic events in people with atrial fibrillation?

For full details see review protocol in Appendix A.

Table 2: 'PICO' characteristics of review question

Question	
Population	People aged >18 with a diagnosis of atrial fibrillation, who are not on anticoagulants
Risk tool	Any stroke risk tool (e.g., ABC stroke score, Q stroke, ATRIA, Troponin 2, CHADSVASC, CHADS2) Any other version of CHADSVASC with modifications
Reference standard	Later stroke and/or thromboembolic event at follow up used in study
Outcomes (in terms of predictive test accuracy, calibration)	Simple diagnostic (prognostic) accuracy outcomes, such as sensitivity and specificity C statistic (based on sensitivity and specificity but useful if >1 threshold used). Calibration outcomes Reclassification
Study types	cohort (external validation, internal validation)
Specific groups	Ethnic groups

3.2 Clinical evidence

The aim of this review was to evaluate the accuracy of stroke/thromboembolism (TE) prediction tools with reference to their discriminatory capabilities (sensitivity, specificity, C statistics, D statistics), calibration (R2 and Hosmer-Lemeshow statistics) and the Net Reclassification Index (NRI) in people with AF. The reference standard was the incidence (or not) of stroke and/or systemic thromboembolism (TE) at follow up.

We therefore searched for cohort studies evaluating stroke/TE prediction tools for people with AF. Only studies which analysed predictive accuracy in people who were not anticoagulated at baseline were included. If a study containing anticoagulated patients also contained a separately analysed sub-group who were not on anticoagulants then the study was also included, although only the data from the non-anticoagulated sub-group were included in the review. Non-anticoagulated cohorts were used because the purpose of stroke prediction tools is to evaluate who *requires* anticoagulation – that is, those people who are at risk of stroke if anticoagulants are *not* taken. For such a risk to be accurately estimated requires that the tool has been validated (with reference to later incidence of stroke/TE) in an analogous non-anticoagulated population. In contrast, use of an anticoagulated cohort would involve the stroke prediction tools identifying those people that have stroke/TE *despite* anticoagulation, which are not necessarily the people that require anticoagulation.

Nevertheless, non-anticoagulated cohorts present problems of their own. If a modern cohort is not anticoagulated this may mean that it is deemed very low risk or that it is 'special' in some way (perhaps by having contraindications to Warfarin or DOACS). This would make

such a cohort unrepresentative of the vast population of people with AF who have been recently diagnosed, and so the predictive capabilities of risk tools in such a cohort might differ from those in the target population. Hence during this review attention has been focussed upon the reasons why the cohort was not anti-coagulated, and whether the characteristics of the cohorts were noticeably different from the general population of people with AF. In general, the non-anticoagulated cohorts included in this review appear not to be low risk, nor do they seem 'special' in any way.

Thirty eight studies evaluating the accuracy of stroke/thromboembolism prediction tools for people with atrial fibrillation who were not anticoagulated were included in the review.[2, 3, 10, 21, 30, 33, 36, 38, 39, 41, 49, 60, 66, 68, 71, 74-76, 81, 83, 84, 95-97, 100, 110, 122-124, 127, 130, 131, 133, 136-140, 142] These studies are summarised in Table 3. The different stroke prediction tools are outlined in Table 4.

Quality of data was generally low or very low. This was partly due to serious or very serious risk of bias in all studies resulting from poor reporting of blinding of prediction tool and outcome data (and vice versa), and from a majority of studies having excessively short follow up periods (<5 years) and/or a relatively low number of events at follow up (<100). In addition, some pooled effects showed serious heterogeneity. This heterogeneity remained unexplained as we had not proposed any pre-hoc sub-grouping strategies.

Evidence from these studies is summarised in the GRADE clinical evidence profiles below (Table 5 to Table 10). See also the study selection flow chart in Appendix B:, study evidence tables in Appendix F:, forest plots in Appendix D:, and excluded studies list in Appendix I:. In summary, there did not appear to be clinically important differences in accuracy between different tools.

Summary of included studies

Table 3: Summary of studies included in the review

Study	Stroke prediction tool(s)	Population (not anti-coagulated)	Outcome event definition	Number of outcome events	Follow up
Abraham 2013[2]	CHADS2 CHADSVASC	5981 post-menopausal women with NVAF from USA. 64.9 hypertensive, 3.7% CHF, 9.2% DM, 2.6% prior stroke, 4.9% prior TIA, 10% prior CAD.	Stroke/TIA obtained from medical records and centrally adjudicated	457	Median 11.8 years
Abumaileq 2015a[3]	CHADSVASC R2CHADS2 ATRIA	154 consecutive patients with NVAF from Spain. Mean age was 74 years, mean SBP was 129, 30% were current smokers, 21% had DM, 6.5% had HF, 15% CHD. 85% CHADSVASC score of 2 points or more	TE event (Stroke/TIA, PE, Peripheral embolism) during follow-up. Stroke needed to last >24 hours and shown on CT/MRI with confirmation from a neurologist. A diagnosis of peripheral embolism was defined as non-central nervous system embolism with clinical or radiographic evidence of arterial occlusion.	9	11 months
Aspberg 2016[10]	ATRIA CHADS2 CHADSVASC	115,153 participants with AF from Sweden. 80.6% percent had score of 2 or more on CHADSVASC. Prior stroke 13%, 70.7% >65 years, 49.3% female, 15.8% DM, 28% HF, 6% Renal failure, 44% hypertension.	Acute ischaemic stroke (defined by ICD-10 code I63), excluding TIAs or other kinds of thromboembolism. The outcome diagnosis, ischaemic stroke, was retrieved from the National Patient Register.	11052	Up to 5 years
Chao 2016[21]	CHADSVASC Age-modified CHADSVASC	124, 271 patients with AF (diagnosed using ICD-9-CM code from the National health Insurance Research database in Taiwan). Age	Ischaemic stroke, with concomitant imaging studies of the brain (CT/MRI)	21,008	Up to 10 years

Study	Stroke prediction tool(s)	Population (not anti-coagulated)	Outcome event definition	Number of outcome events	Follow up
		72, 54% male, 56.8% hypertensive, 23% DM, 38% CHF, 28% previous stroke/TIA. Median CHADSVASC score 3.			
Fang 2008[30]	AFI 1994 SPAF CHADS2 Framingham ACCP 2004	5,588 patients with NVAF from USA. Sample data not given for this cohort. 81.3% at moderate or high risk of stroke	Hospital database searched for incident thromboembolic events, either ischemic stroke or other peripheral embolism. The validity of potential events was adjudicated by an outcomes committee of 3 physicians using a formal study protocol. If there was no consensus on the validity of an event, an expert neurologist adjudicated the event. Outcome events that occurred during hospitalization or as a complication from a diagnostic or interventional procedure were excluded	685	6 years
Fox, 2017[33]	GARFIELD CHADSVASC	2301 people with AF. Demographic data not available	Composite of IS, SE and TIA	51	3 years
Friberg 2012b[36]	CHADSVASC, CHADS2, revised CHADS2, SPAF 1999, AFI 1994, ACC/AHA/ESC, Framingham, NICE	90, 490 patients with AF defined by ICD-10 code 1489 with or without subscales A-F from Sweden. Demographic data not available.	First occurrence of Ischaemic stroke (defined by ICD-10 code 163). Events in first 14 days post inception excluded.	5359	1.4 years
Gage 2001[39]	CHADS2 AFI 1994 SPAF 1995	1733 patients from the US National Registry of AF cohort. Mean age 81, 58% women, 56% CHF, 56% hypertension, 23% DM, 25% history of cerebral ischaemia. 1204 were not prescribed any	Hospitalisation for ischaemic stroke as determined by Medicare claims. ICD-9-CM codes used to identify. 1.2 year FU	94	1.2 years

Study	Stroke prediction tool(s)	Population (not anti-coagulated)	Outcome event definition	Number of outcome events	Follow up
		antithrombotic therapy and 529 (31%) were prescribed aspirin. CHADS2 score of 2.1.			
Gage 2004 [38]	AFI 1994, SPAF 1995, ACCP 2001, CHADS2, Framingham	2580 patients with NVAF from 6 international RCTs. 37% women, mean age 72, 46% hypertension, 25% HF, 13% DM, 22% prior stroke or TIA, 18% prior MI/angina. 59% moderate or high risk.	Suspected stroke, confirmed by CT in 98% of incident neurological events. Strokes defined as neurological deficits that persisted > 24 hours and not associated with an intracranial haemorrhage.	207	1.9 years
Guo 2013[41]	CHADS2 CHADSVASC	885 patients with pre-existing diagnosis of permanent, persistent or paroxysmal AF at General Hospital in China between 2007 and 2010. Mean age 77, 27% female, 75% hypertensive, 39% DM, 23% HF, 63% CAD, 20.9% prior stroke, renal failure 9.6%. 81.2% high risk on CHADSVASC.	Major adverse events (stroke/TE). IS defined as focal neurological deficit of sudden onset lasting >24 hours diagnosed clinically by a neurologist. A TE was IS, PE or peripheral embolism.	85	1.9 years
Hippisley Cox 2013[49]	Q stroke CHADS2 CHADSVASC	7689 people with NVAF from UK. 71% percent high risk on CHADS2. People with prior stroke or TIA excluded. Demographic data not available for this cohort.	Stroke/TIA, excluding haemorrhagic stroke, as defined by ICD-10 codes: cerebral infarction (I63) and stroke not specified as haemorrhage or infarction (I64).	890	Up to 10 years
Kang 2017 [60]	CHADS2 CHADSVASC	10,846 patients with newly diagnosed NVAF from South Korea. Mean age	Ischaemic stroke. Stroke was defined according to ICD-10 codes (I63-64) for diagnoses made during hospitalization and	888	1.2 years

Study	Stroke prediction tool(s)	Population (not anti-coagulated)	Outcome event definition	Number of outcome events	Follow up
		63.7 years, 47% women, previous stroke 16.7%, CHF 25%, DM 21%, IHD 48%, CHADS more than or equal to 4: 16%, CHADSVASC more than or equal to 6: 10%.	according to brain imaging such as computed tomography and magnetic resonance imaging		
Kim 2017[66]	CHADS2, CHADSVASC, ATRIA	5855 NVAF patients from South Korea. Mean age 64, 48% women, CHADSVASC means core 3.28, 24.5% prior stroke, 13% MI, 32% HF, 76% hypertension, 20% DM.	The primary end point was incident ischemic stroke (including ischemic stroke–related death). Diagnosis made with concomitant brain imaging studies, including computed tomography or MRI.	819	4.2years
Larsen 2012[71]	CHADS2 CHADSVASC	1603 patients with incident AF (defined by ICD-08 [pre 1994] or ICD-10 codes) from a Danish cohort of 57,053 middle aged people. Age 67, 40% women, mean follow up 5.4 years, CHF 24.4%, 30% hypertension, 10% DM, 6% stroke history. 7% CHADS2 of 5 or above, 6% CHADSVASC score of 5 or above.	Stroke (not defined)	unclear	5.4 years
Lip 2006[74]	CHADS2, CHADSVASC (Birmingham 2009), CHADS2 with vWF, Birmingham with vWF	994 patients with NVAF, from USA. Mean age 69.3, 75% male, 53% hypertension, 14% diabetes, 19% recent HF, 13% previous TIA/stroke, 10% previous MI, 6% PVD, 9% LV systolic dysfunction,	Ischaemic stroke (not defined)	unclear	1.6 years

Study	Stroke prediction tool(s)	Population (not anti-coagulated)	Outcome event definition	Number of outcome events	Follow up
		8% current smokers. 43 IS events. 73.9% not low risk according to CHADS2.			
Lip 2010[75]	AFI 1994, SPAF 1999, CHADS2, revised CHADS2, Framingham, NICE, ACCA/AHA/ESC, ACCP 2008 and CHADSVASC (Birmingham)	1084 NVAF patients from USA. Age 66 years, 41% women, previous stroke 4.2%, TIA 4.3%, DM 17.3%, hypertension 67%, HF 23.5%, antiplatelets 74%, LVEF 53%. 17% classed as high risk and 61.9% as intermediate risk on CHADS2	Thromboembolic events: IS (focal neurological event lasting >24 hours diagnosed by neurologist), PE or peripheral embolism	25	1 year
Lip 2014[76]	SAMe-TT2R2	3,483 patients with AF (n=242 had valvular AF) who were not receiving OACs. Mean age 70, 43% female, 48% HF, 33% CAD, 17% previous MI, 5% previous CABG, 40% hypertensive, 7% previous stroke, 9% renal insufficiency. Mean CHADSVASC score 3.1.	Stroke/ TEs (not defined)	273	Up to 10 years
Maheshwari, 2019[81]	CHADSVASC P2-CHADSVASC	2229 participants from the ARIC study (Atherosclerosis Risk in Communities) and 700 participants from MESA (Multi-Ethnic Study of Atherosclerosis) with incident AF who were not on anticoagulants within 1 year of AF diagnosis;	Ischaemic stroke	47 (ARIC) 31 (MESA)	1 year (5 years for ARIC CHADSVASC)

Study	Stroke prediction tool(s)	Population (not anti-coagulated)	Outcome event definition	Number of outcome events	Follow up
		ARIC cohort: age 73; female 47%; DM 30%; hypertension 75%; previous MI 24%; HF 38%; PAD 9%; past stroke/TIA 15%; CHADSVASC 3.6; black 19%, white 81%; MESA cohort: age 76; female 45%; DM 18%; hypertension 68%; previous MI 6%; HF 8%; PAD 2%; past stroke/TIA 6%; CHADSVASC 3.0; black 20%, white 49%; Chines 13%; Hispanic 17%			
McAlister, 2017[83]	CHADS2, CHADSVASC, R2CHADS2 (71 point), ATRIA, CHADS2KDIGO, CHADS2Alb, CHADS2 eGFR	58,451 people from Alberta Canada with incident NVAF, and no anticoagulant use. eGFR < 60 24.4%; previous stroke 10.8%; previous bleed 11.2%; age >65 52.6%; female 47%; previous MI: 11.3%; HF: 21.8%; DM: 21.6%; PVD: 3.5%; hypertensive: 64.1%	Stroke/TE (not defined)	7340	2.5 years
McAlister, 2018[84]	CHADS2 CHADSVASC ATRIA	This was a sample of people (of an unknown size) with AF (defined as: ICD-9CM 427.3 or ICD-10CA I48) and who were not treated with OACs. No details are given of their characteristics. They were	First TE (first stroke, TIA or systemic arterial thromboembolism)	10,827	1 year

Study	Stroke prediction tool(s)	Population (not anti-coagulated)	Outcome event definition	Number of outcome events	Follow up
		drawn from a larger cohort of 147,952 adult Canadians with AF.			
Olesen 2011[96]	CHADS2, CHADSVASC	73,538 people with NVAF from Denmark. CHADSVASC of 2 or more was 80.5. Age >75 60%, female 51%, DM 9%, previous TE 18%, Vascular disease 18%, antiplatelets 35%.	Admission to hospital, or death, from TE (defined by codes I26,63,64 and 74).	unclear	1 year
Olesen 2012[95]	CHADS2with vascular disease added CHADSVASC	924 people aged <65 years with NVAF or atrial flutter. No demographic data for these provided.	IS/thromboembolism (not defined)	14	Up to 10 years
Olesen 2012b[97]	CHADS2 CHADSVASC	47,576 patients with atrial fibrillation (defined by ICD code I48 from Danish National Patient Registry), not on OACs. Mean age 69.4, CHF 2%, hypertension 17%, DM 2%, previous stroke 0%, vascular disease 12%, female 46.3%, aspirin 26%. 63% CHADSVASC score of 2 or more. All had CHADS2 scores of 0 or 1.	Hospitalisation or death from stroke/TE. ICD codes ICD-10: G458, G459, I63,I64,I74)	4599	12 years
Piccini, 2013[100]	CHADS2 R2CHADS2 score – CHADS2 with creatinine clearance	Sub-group from the ATRIA cohort that were NOT taking OACS (n=16,360). No information given on	Stroke – a composite of all stroke (both ischemic and haemorrhagic) and systemic embolism.	Unclear	3 years

Study	Stroke prediction tool(s)	Population (not anti-coagulated)	Outcome event definition	Number of outcome events	Follow up
	incorporated (2 points for CrCl <60mL/min) Sum of CrCl<60 ml and prior stroke/TIA	characteristics in Piccini, 2013.			
Reps, 2020[110]	ATRIA CHADS2 CHADSVASC FRAMINGHAM Q-Stroke	312,354 people from several health databases in USA and South Korea with AF and no prior anticoagulant use or stroke	Ischeamic or hemorrhagic stroke recorded with an inpatient or ER visit in datbase	7584	1 year
Schwartz, 2019[121]	Modified CHADSVASC (excluding pervious stroke/TIA)	Data from 11,443 patients with AF who were NOT on DOACs or VKAs were retrieved from the Northwestern Healthcare system's Enterprise Database Warehouse. The data allowed identification of stroke outcomes, and calculation of prior CHADSVASC scores. Mean age 67.6 for white patients and 63.1 for non-white patients. Mean CHADSVASC was 2.4 in whites and 2.2 in non-whites	Incident Stroke using ICD-9 codes and ICD-10 codes	205	971 days
Singer 2013[123]	ATRIA, CHADS2, CHADSVASC	25, 306 patients with NVAF from USA. TE rate of 1.9% per year (496 stroke or other TE events). No	IS, defined as sudden onset of a neurologic deficit lasting >24 hours and not attributable to other causes. Other TEs: sudden occlusion to an artery to a major organ documented by	496	1 year

Study	Stroke prediction tool(s)	Population (not anti-coagulated)	Outcome event definition	Number of outcome events	Follow up
		demographic data for this cohort.	imaging, surgery or pathology and not due to concomitant atherosclerosis or other causes.		
Siu 2014[124]	CHADS2 CHADSVASC	3881 patients with NVAF (not defined) who did not receive OACs. Mean age 77, 53.5% female, 47.5% hypertensive, 18% DM, 1.7% renal failure on dialysis, 19% HF, 8% CAD, 1.3% PAD, 17% prior stroke/TIA. Mean CHADSVASC 3.3.	Stroke (not defined)	847	3.2 years
Suzuki 2015[127]	CHADS2 CHADSVASC	3588 patients with AF. Taken from 3 Japanese databases. Age 68.1, 34% female, 50% hypertension, 15% DM, 8.5% previous stroke or TIA, 15% HF, 11% CAD, 42% antiplatelet use. No data on CHADSVASC scores at baseline	Ischaemic stroke (not defined)	69	1.4 years
Tomasdottir, 2019[130]	CHADSVASC	231 077 (48.1% women) non-selected patients with AF not receiving oral anticoagulation from 2006 to 2014. Data from cross-linked national Swedish registers. Age 75 (men), 82 (women); HF 28.5%; hypertension 48.4%; DM 17.2%; Stroke/TIA/SE 18.7%; Vascular disease 24.1%	Ischaemic stroke	17,540	2.5 years

Study	Stroke prediction tool(s)	Population (not anti-coagulated)	Outcome event definition	Number of outcome events	Follow up
Tomita 2015[131]	mCHADSVA mCHADSVASC CHADS2	294 women and 703 men with NVAF from Japan. Mean mCHADSVASC scores of 1.9 (male) and 3.3 (female). , Mean age 687% history of stroke/TIA, 58% antiplatelet use, 29% paroxysmal AF	Thromboembolic events (not defined)	30	2 years
Van dem Ham 2015[133]	ATRIA, CHADSVASC and CHADS2	60, 594 patients with NVAF from Netherlands. Mean age 74.4 years, female 48.7%, 50% past or present smokers; 12% DM, 17.5% CHF, 54.6% hypertension, 15% previous stroke/TIA, 31% vascular disease, 28% renal dysfunction (eGFR <60 ml/min/1.73m2).	Ischeamic stroke (defined by codes in CPRD, HES or both)	3751	2.1 years
Van Staa 2011[136]	AFI 1994 AFI 1998 ACCP 2001 ACCP2004 ACCP 2008 NICE 2006 ACC/AHA/ESC CHADSVASC CHADS2 Modified CHADS2 SPAF 1995 Hart 1999 Van Walraven 2002 Van Latum1995	79,884 patients with NVAF from Netherlands. Age 73.3, female 49.7%, 54.6% current or past smoker, CHADS score more than or equal to 3: 20%, CHF 29%, DM 17%, Hypertension 50%, previous stroke or TIA 18%.	Stroke as recorded in the GPRD, hospitalisation for stroke as recorded in the HES, and mortality resulting from stroke as recorded on death certificates.	1233	4 years

Study	Stroke prediction tool(s)	Population (not anti-coagulated)	Outcome event definition	Number of outcome events	Follow up
	Framingham 2003				
Wang 2003[137]	Framigham CHADS2 SPAF 1995 AFI 1994	705 participants with new onset AF from USA. Mean age 75, 48% women, 50% on hypertension therapy, 15% DM, 18% smoking, 34% prior CHF or MI.	Stroke – decided by a panel of 3 Framingham investigators, including a neurologist, based on a review of all medical records and clinical data, and an examination by the neurologist.	83	4 years
Wicke, 2019[138]	CHADSVASC	A broadly representative population with AF who were not on OACs from southern Germany (n=30,299). Claims data from a statutory health insurance (AOK Baden Wuerttemberg), the largest insurance fund in the German state of Baden-Wuerttemberg (population in 2014 was 10.7 million), were used. For the year 2014, the data contained information on 3.8 million individuals, which equals to about 35% of the state's population. Age 76.4; 46.6% male; CHADSVASC score 4.25; hypertension 85%; CHF 40.2%; stroke/TIA 7.96%; DM 10.1%;	Hospitalisation for Ischaemic stroke	961	2 years
Xing 2016[139]	CHADS2 CHADSVASC	413 patients with NVAF, from China. mean age 81, 71% male, median	Ischaemic stroke – new sudden focal neurological deficit resulting from a supposed CV cause that persisted >24 hours and not	59	2 years

Study	Stroke prediction tool(s)	Population (not anti-coagulated)	Outcome event definition	Number of outcome events	Follow up
		CHADSVASC score 4.77. Hypertension 77.5%, previous stroke/TIA 36.8%, DM 36.1%, antiplatelets 68%.	attributable to other causes. Brain imaging also used to differentiate from haemorrhage.		
Xing 2018[140]	CHADSVASC	389 consecutive patients with AF from China. Age 83.7, 77% female, 82% hypertension, 56% vascular disease, 36% DM, 36% previous IS, 25% HF, Cr 100 mg/dL, EF 62%, CHADSVASC 4.87.	Ischaemic stroke – new sudden focal neurological deficit resulting from a supposed CV cause that persisted >24 hours and not attributable to other causes. Brain imaging also used to differentiate from haemorrhage.	49	2.6 years
Yoshizawa 2017[142] and Komatzu, 2014[68]	R2CHADS CHADS2 CHADSVASC	332 people with NVAF from Japan. Age 65, male/female: 224:108, hypertension 43%, DM 13%, smoking 27%, underlying heart disease 20% (IHD 11.4%, non-ischaemic 8.6%), 18 month Hx of AF, 33% on aspirin, CHADSVASC score 2 points or more: 59%.	IS/STE. Cerebral TE confirmed based on clinical symptoms and the presence of a 3mm or larger infarct area on CT/MRI.	unclear	4.4 years

Table 4: Summary of stroke/TE prediction tools and their constituent variables and cut-offs (where available)

Risk tool	Variables and scoring
ACC/AHA/ESC guidelines 2006	No risk factors= low risk; age>75years, or hypertension, or heart failure, or LVEF <35%, or diabetes=intermediate risk; Previous stroke, TIA or embolism, or ≥2 moderate risk factors of (age ≥75y, hypertension, heart failure, LVEF ≤35%, diabetes)=high risk

Risk tool	Variables and scoring
ACCP (American College of Chest Physicians on Antithrombotic and Thrombolytic Therapy guidelines)2001	No risk factors=low risk; age 65-75, or diabetes or CAD=moderate risk; age >75 years or history of ischaemic stroke/TIA, or systemic embolism or hypertension or poor left ventricular systolic function or rheumatic valve disease or prosthetic valve disease=high risk
ACCP 2004	Age <65 years and no other risk factors=low risk; age 65-75 and no risk factors= moderate risk; age > 75 or history of stroke/TIA or systemic embolism or poor left ventricular function/HF or hypertension or diabetes=high risk
ACCP 2008	No risk factors=low risk; age >75 years, or hypertension, or moderately or severely impaired LVEF and/or heart failure, or diabetes=intermediate risk; previous stroke, TIA or embolism, or >2 moderate risk factors: age>75 years, hypertension, heart failure, LVEF <35%, diabetes=high risk
AFI (Atrial Fibrillation Investigators) 1994	Age<65 years and no other risk factors=low risk; Age >65 years and no other risk factors=intermediate risk; prior ischaemic stroke or TIA, history of hypertension, history of DM = high risk
AFI 1998	Risk factors: history of stroke/TIA, hypertension, diabetes. Age<65 years and no other risk factors=low risk; Age >65 years and no other risk factors=intermediate risk; moderate/severe left ventricular dysfunction (echocardiography) or age <65 years and >1 risk factor or age 65-75 years and >1 risk factors or age > 75 years = high risk
Age modified CHADSVASC[21]	As CHADSVASC but age category for intermediate risk extended from 65-74 to 50-74.
ATRIA	One point each for female sex, DM, CHF, hypertension, proteinuria, eGFR<45. Age >85 = 6 points (or 9 if prior stroke/TIA), age 75-84 = 5 points (or 7 if prior stroke/TIA), age 65-74 3 points (or 7 if prior stroke/TIA), age <65 0 points (or 8 with prior stroke/TIA).
CHADS2	One point each for CHF, hypertension, age 75 of older, and DM, and 2 points for prior stroke or TIA. Score 0=low risk; score 1-2=intermediate risk; score 3 to 6=high risk
CHADS2 Alb	As CHADS2 but with addition of the albuminuria measurements only. This additional albuminuria component was categorised as low (0 points), moderate (1), or high (3). These scores were added on to the conventional CHADS2 scores (with a maximum of 6) to create this new score with a maximum of 9 (6+3) points. On this scale high risk was deemed as >2 points.
CHADS2 eGFR	As CHADS2 but with addition of the eGFR measurements only. This additional eGFR component was categorised as >60 mL/min/1.73m2 (0 points), 45-59 mL/min/1.73m2 (4 points), 30-44 mL/min/1.73m2 (5 points), or <30 mL/min/1.73m2 (7 points). These scores were added on to the conventional CHADS2 scores (with a maximum of 6) to create this new score with a maximum of 13 (6+7) points. On this scale high risk was deemed as >2 points.
CHADS2 KDIGO	As CHADS2 but with addition of the KDIGO component. KDIGO score was based on both eGFR and albuminuria measurements, and was categorised as low (0 points), moderate (3), high (5) or very high (7). These scores were added on to the conventional CHADS2 scores (with a maximum of 6) to create this new score with a maximum of 13 (6+7) points. On this scale high risk was deemed as >3 points.

Risk tool	Variables and scoring
CHADS2 with vascular disease added[95]	Vascular disease added as a risk factor to CHADS2. No details given on relationship between scores and risk.
CHADS2 with vWF[74]	As CHADS2, with extra point for plasma von Willebrand Factor levels (vWf) > 158 IU/dL
CHADSVASC 2009 (Also known as BIRMINGHAM)	One point for female sex, history of CHF, history of hypertension, history of vascular disease or history of DM. 2 points for history of stroke/TE. Age <65=0 points, 65-74=1 point, >75=2 points. Maximum score 9 points. Low risk =0 points; 1 point=low/moderate; >2 points moderate/high
CHADSVASC with vWF[74]	As CHADSVASC, with extra point for plasma von Willebrand Factor levels (vWf) > 158 IU/dL
FRAMINGHAM	Age 0-10 points, female gender 6 points, systolic blood pressure 0-4 points, DM 5 points, prior stroke/TIA 6 points. Score 0-7=low risk; score 8 to 15 intermediate risk; score 16 to 31=high risk
GARFIELD AF Risk	Risk of ischemic stroke or systemic embolism =1-[0.991344397 exp(0.03048226*(**age**-60) + 0.952524717* history of **stroke** + 0.432357326* history of **bleed** + 0.319129628*history of **heart failure** +0.574919171*history of **chronic kidney disease** + 0.654249546*living in **Other Region (living in Aust, NZ or SA)** + 0.671380382***Black/ Mixed/ Other race** -0.582045773* **Oral Anticoagulant**)].
Hart 1998	No risk factors=low risk; hypertension+ age <75 years or diabetes=intermediate risk; history of stroke/TIA or women aged >75 years or men aged >75 years + hypertension or systolic >160=high risk
mCHADSVA – female gender removed[131]	As mCHADSVASC (above) but female category removed
mCHADSVASC – for the vascular disease criterion, only coronary artery disease is included as a risk factor [131]	As CHADSVASC but for the vascular disease component only coronary artery disease was included as a risk factor
Modified CHADS2[136]	Age 40-64 +1, age 65-69 +2, age 70-74 +3, age 75-79 +4, age 80-84 +5, age >85 +6, woman +1, DM +1, history of stroke/TIA +1. Score 0=low risk; score 1-5 moderate risk; score 6-14 high risk
Modified CHADSVASC (no previous stroke/TIA)[122]	As CHADSVASC but no previous stroke/TIA component included.
NICE	Age <65 with no moderate/high risk factors=low risk; age >65 with no high risk factors OR age <75years with hypertension, diabetes or vascular disease = intermediate risk; previous stroke/TIA or thromboembolic event OR age >75 years with hypertension, diabetes or vascular disease OR clinical evidence of valve disease or heart failure, or impaired left ventricular function=high risk.
P2-CHADSVASC[81]	As CHADSVASC with addition of *abnormal p-wave axis*, which was given a score of 2 if present.

Risk tool	Variables and scoring
Q STROKE	QStroke includes measurements of age, sex, deprivation, ethnicity, body mass index, systolic blood pressure, total cholesterol:HDL cholesterol ratio, smoking status (five levels), diabetes type, congestive cardiac failure, coronary heart disease, rheumatoid arthritis, chronic kidney disease, treated hypertension, valvular heart disease, and family history of premature coronary heart disease. A % score is derived that provides an absolute risk of stroke over a choice of durations, from 1 to 10 years.
R2 CHADS2[3, 100, 142]	R2CHADS2 was calculated by adding 2 points for renal dysfunction (i.e. estimated glomerular filtration rate [eGFR] <60 ml/min/1.73 m2); 2 points for prior stroke or TIA; and one point for each of the following factors: congestive heart failure, hypertension, age ≥ 75 and diabetes mellitus with a maximum score of 8 points.
R2 CHADS2 (71 points)[83]	This appears to be completely different to the R2 CHADS2 scheme outlined above. The score used by McAlister et al. (2017) was out of a total of 71 points, as follows: eGFR (0-29 points), previous stroke (18 points), age 65-75 (2 points), age >75 (3 points), female (5 points), previous MI (6 points), HF (-2 points), DM (4 points), Hypertension (5 points) and PVD (6 points). The authors stated that this score is normally given out of 100, but was reduced to 71 because there were no data on diastolic bp or HR, and patients had incident AF.
Revised CHADS2[36]	As CHADS2 risk factors but 0=low risk; 1=intermediate risk; 2to 6=high risk
SAMe-TT2R2	Calculated as the sum of points after addition of one point each for female sex, age<60 years, medical history of >2 co-morbidities (among hypertension, DM, CAD or MI, PAD, CHF, previous stroke/TIA, pulmonary disease or hepatic/renal disease), and two points each for smoking and non-white race. Scores of 0-1=low risk; 2=intermediate risk; >2=high risk
SPAF 1999	No risk factors=low risk; hypertension or DM = moderate risk; previous stroke/TIA or women aged >75 or men aged >75 with hypertension=high risk
SPAF (Stroke Prevention in Atrial Fibrillation) 1995	No risk factors=low risk; history of hypertension=intermediate risk; prior stroke, women older than 75 years, recent clinical heart failure or LV fractional shortening <25% on echocardiography, or systolic bp >160=high risk
Sum of CrCl <60 mL/min and prior stroke/TIA[100]	Unclear but probably 2 points for CrCl<60mL/min and 2 points for prior stroke/TIA
Van Latum	Risk factors: history of stroke/TIA, IHD, enlarged cardiothoracic ratio on chest roentgenogram, systolic bp>160, AF>1 year, visible ischaemic lesion on CT. No risk factors=low risk; 1-2 risk factors=moderate risk; >3 risk factors)=high risk
Van Walraven	No risk factors=low risk; history of stroke/TIA or treated hypertension or SBP >140 or previous MI/angina or DM=mod/high risk

3.2.1 Discrimination

Table 5: Clinical evidence profile: Discriminative capacity of stroke prediction tools featured in the studies (see table 3).

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
CHADS2	27	884,951 (one study n is unknown)	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**POOLED EFFECT: Random Effects: 0.66(0.64-0.69); I^2=98%**	VERY LOW
Modified CHADS2 (Van Staa, 2010)	1	79,884	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.69(0.67-0.71)	VERY LOW
Revised CHADS2 (Friberg 2012)	2	91,574	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	**POOLED EFFECT: Fixed Effects: 0.62(0.61-0.63); I^2=0%**	LOW
R2 CHADS2 (Abumaileq 2015, Yoshizawa, 2017, Piccini, 2013)	3	16846	Very serious risk of bias[a]	Very serious risk of inconsistency[b]	No serious indirectness	Serious imprecision	**POOLED EFFECT: Random Effects: 0.74(0.62-0.86); I^2=92%**	VERY LOW

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
R2CHADS2 (71 points) (McAlister, 2017)	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.66(0.64-0.67)	LOW
CHADS2 KDIGO (McAlister, 2017)	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.65(0.64-0.66)	LOW
CHADS2 Alb (McAlister, 2017)	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.65(0.64-0.67)	LOW
CHADS2 eGFR (McAlister, 2017)	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.67(0.65-0.68)	LOW
CHADS2 with vWF	1	994	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.69(0.60-0.77)	VERY LOW
CHADSVASC 2009	27	987,032 (in one study n unknown)	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**POOLED EFFECT: Random Effects: 0.67(0.65-0.69); I^2=99%**	VERY LOW
P2-CHADSVASC	2	2929	Very serious	No serious inconsistency	No serious indirectness	Serious imprecision	**POOLED EFFECT: Fixed effect 0.68 (0.62-0.75) I^2=0%**	VERY LOW

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
			risk of bias[a]					
Age modified CHADSVASC (Chao 2016)	1	124,271	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.71(0.70-0.71)	MODERATE
mCHADSVASC (modified in Tomita 2015)	1	997	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.60(0.51-0.68)	LOW
Modified CHADSVASC (no stroke/TIA)[108]	1	11433	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.65(0.57-0.72)(non-white) 0.68(0.64-0.72)(white)	VERY LOW
mCHADSVA – (Modified in Tomita 2015)	1	997	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.62(0.53-0.71)	VERY LOW
Q STROKE	2	320,043	Serious risk of bias[a]	Very serious inconsistency	No serious indirectness	No serious imprecision	**POOLED EFFECT: Random Effects: 0.61 (0.56-0.66); I^2=96%**	VERY LOW
ATRIA	7	572,012 (one study	Very serious	Very serious	No serious indirectness	serious imprecision	**POOLED EFFECT: Random Effects: 0.67 (0.64-0.70); I^2=98%**	VERY LOW

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
		unknown n)	risk of bias[a]	inconsistency				
AFI 1994	7	182,064	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**POOLED EFFECT: Random Effects: 0.62(0.57-0.66); I^2=92%**	VERY LOW
AFI 1998	1	79,884	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.61(0.60-0.62)	LOW
SPAF 1995	5	90,490	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	Serious imprecision	**POOLED EFFECT: Random Effects: 0.68(0.58-0.79); I^2=97%**	VERY LOW
SPAF 1999	2	91,574	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Very serious imprecision[c]	**POOLED EFFECT: Random Effects: 0.60(0.49-0.70); I^2=50%**	VERY LOW
FRAMINGHAM	7	492,685	Very serious risk of bias[a]	Very serious inconsistency	No serious indirectness	No serious imprecision	**POOLED EFFECT: Fixed Effects: 0.64(0.60-0.67); I^2=97%**	VERY LOW
ACCP 2001	2	82,464	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	**Range:0.58 to 0.62 Median: 0.60**	LOW
ACCP 2004	2	85,472	Very serious	No serious inconsistency	No serious indirectness	No serious imprecision	**Range: 0.60 to 0.61 Median: 0.605**	LOW

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
			risk of bias[a]					
ACCP 2008	2	80,968	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	**POOLED EFFECT: Fixed Effects: 0.64(0.62-0.66); I^2=0%**	LOW
ACC/AHA/ESC guidelines 2006	3	171,458	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	**POOLED EFFECT: Fixed Effects: 0.62(0.61-0.63); I^2=47%**	LOW
NICE	3	171,458	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	No serious imprecision	**POOLED EFFECT: Random Effects: 0.62(0.59-0.65); I^2=72%**	VERY LOW
Hart 1998	1	79,884	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.62(0.60-0.64)	LOW
Van Walraven	1	79,884	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.55(0.54-0.58)	LOW
Van Latum	1	79,884	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.57(0.55-0.59)	LOW
CHADSVASC with vWF	1	994	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.68(0.59-0.76)	VERY LOW

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
GARFIELD	1	2301	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.70(0.63-0.77)	VERY LOW
SAMe-TT2R2	1	3483	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.51(0.49-0.53)	LOW
Sum of CrCl <60 mL/min and prior stroke/TIA[88]	1	16,360	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.61 (0.58-0.64)	LOW

Pooling (meta-analysis) was carried out if there were at least two studies per risk tool with confidence intervals. RevMan was used to carry out the analyses. If pooling was not possible for risk tools with >1 data point then the range and median value of the study point estimates were recorded. If there were only one data point then only the result from the study was recorded.
a) Risk of bias was assessed using the PROBAST checklist (see Appendix F).Risk of bias was serious for some risk tools because none of the studies reported any blinding of assessors for risk tool data and outcome status, and most did not report loss to follow up, although follow up and number of events were appropriate. Risk of bias was very serious for the rest of the risk tools because many studies with the aforementioned limitations also had insufficient numbers of events (<100) and/or inappropriately short follow up times (<5 years) to be able to accurately predict risk.
b) Where data were pooled, an I^2 of 50-74% was deemed serious inconsistency and an I^2 of 75% or above was deemed very serious inconsistency. If no pooling were possible, inconsistency was assessed by inspection of the degree of overlap of confidence intervals between studies: if one of more Cis did not overlap then a rating of serious inconsistency was given. Reasons for heterogeneity between studies may include geographical/cultural/ethnic differences. Clinically the studies appeared reasonably homogeneous, with similar rates of hypertension, diabetes and former stroke.
c) The judgement of precision was based on the spread of confidence interval across two clinical thresholds: C statistics of 0.5 and 0.7. The threshold of 0.5 marked the boundary between no predictive value better than chance and a predictive value better than chance. The threshold of 0.7 marked the boundary above which the committee might consider recommendations. If the 95% Cis crossed one of these thresholds a rating of serious imprecision was given and if they crossed both of these thresholds a rating of very serious imprecision as given.

Table 6: Clinical evidence profile: sensitivity and specificity of stroke prediction tools featured in the studies (see table 3). For pooled data the 95% CIs of individual studies can be found in the Forest plots in the appendices. For individual or non-pooled data the 95% CIs are given below. The pooled sensitivity/specificity values have been calculated using Bayesian methodology and are expressed as medians (95% credible intervals).

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
CHADS2 at threshold of ≥1	6	172,747	**Pooled sensitivity: 0.874(0.676-0.960)**	**Pooled specificity: 0.228(0.131-0.501)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
					specificity				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
CHADS2 at threshold of ≥2	5	165,058	**Pooled sensitivity: 0.582(0.308-0.811)**	**Pooled specificity: 0.625(0.363-0.835)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					risk of bias[a]				
CHADS2 at threshold of ≥3	5	165,058	**Pooled sensitivity: 0.316(0.129-0.593)**	**Pooled specificity: 0.845(0.641-0.944)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	No serious imprecision	VERY LOW
Revised CHADS2 (Friberg 2012)	1	90,490	0.980 at standard threshold [no raw data in paper, and no 95% CIs reported]	0.150 at standard threshold [no raw data in paper, and no 95% CIs reported]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
R2CHADS2 (71 points) (McAlister, 2017)	1	7340	0.800 no specified threshold	0.511 no specified threshold	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
CHADS2 KDIGO (McAlister, 2017)	1	7340	0.726 no specified threshold	0.575 no specified threshold	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
CHADS2 Alb (McAlister, 2017)	1	7340	0.821 no specified threshold	0.488 no specified threshold	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
CHADS2 eGFR (McAlister, 2017)	1	7340	0.693 no specified threshold	0.640 no specified threshold	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
CHADSVASC 2009 at threshold of ≥1	9	440,691	**Pooled sensitivity: 0.977(0.947-0.992)**	**Pooled specificity: 0.092(0.051-0.156)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	No serious imprecision	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
CHADSVASC 2009 at threshold of ≥2	9	438983	**Pooled sensitivity: 0.923(0.850-0.964)**	**Pooled specificity: 0.223(0.144-0.328)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	No serious imprecision	VERY LOW
CHADSVASC 2009 at threshold of ≥3	8	569,938	**Pooled sensitivity: 0.809(0.631-0.913)**	**Pooled specificity: 0.431(0.287-0.582)**	**Sensitivity**				
					Serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
CHADSVASC 2009 at threshold of ≥4	8	438,829	**Pooled sensitivity: 0.524(0.347-0.695)**	**Pooled specificity: 0.646(0.477-0.781)**	**Sensitivity**				
					Serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
Modified CHADSVASC (no	1	11,433	0.821(0.759-0.872)	0.393(0.384-0.402)	**Sensitivity**				
					Very serious	NA	No serious indirectness	No serious imprecision	LOW

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
stroke/TIA)[122] at threshold for risk of ≥2					risk of bias[a]				
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
Modified CHADSVASC (no stroke/TIA)[122] at threshold for risk of ≥3	1	11,433	0.631(0.559-0.699)	0.612(0.603-0.621)	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	Serious imprecision	VERY LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
Modified CHADSVASC (no stroke/TIA)[122] at threshold for risk of ≥4	1	11,433	0.359(0.292-0.431)	0.798(0.791-0.805)	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
	1	7689			**Sensitivity**				

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Q STROKE with optimal cut-off at top 63%			0.825 (0.798-0.849) with optimal cut-off at top 63%	0.395(0.383-0.407) with optimal cut-off at top 63%	Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
					Specificity				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
Q STROKE with at top 90%	1	7689	0.992(0.984-0.997) with cut-off at top 90%	0.112(0.105-0.119) with cut-off at top 90%	**Sensitivity**				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
					Specificity				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
Q STROKE with at top 85%	1	7689	0.979(0.967-0.987) with cut-off at top 85%	0.167(0.158-0.176) with cut-off at top 85%	**Sensitivity**				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
					Specificity				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
Q STROKE with at top 80%	1	7689	0.958(0.943-0.971) with cut-off at top 80%	0.221(0.211-0.231) with cut-off at top 80%	**Sensitivity**				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
					Specificity				

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
Q STROKE with at top 70%	1	7689	0.890(0.868-0.909) with cut-off at top 70%	0.325(0.314-0.336) with cut-off at top 70%	**Sensitivity**				
					Serious risk of bias[a]	NA	No serious indirectness	Serious imprecision[c]	LOW
					Specificity				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
ATRIA at threshold for risk of ≥1	2	158004	**Median[d]: 0.985(0.983-0.987)**	**Median[d]: 0.091(0.089-0.168)[10]**	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	Serious imprecision[c]	VERY LOW
ATRIA at threshold for risk of ≥2	1	152149	0.967(0.964-0.970)	0.166(0.164-0.168)	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very serious	NA	No serious indirectness	No serious imprecision	LOW

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					risk of bias[a]				
ATRIA at threshold for risk of ≥3	1	152149	0.958(0.955-0.962)	0.192(0.189-0.194)	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
ATRIA at threshold for risk of ≥4	1	152149	0.936(0.931-0.940)	0.241(0.238-0.243)	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
ATRIA at threshold for risk of ≥5	1	152149	0.894(0.888-0.899)	0.309(0.307-0.312)	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very serious	NA	No serious indirectness	No serious imprecision	LOW

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					risk of bias[a]				
ATRIA at threshold for risk of ≥6	3	158158	**Median[d]: 0.444(0.137-0.788)**	**Median[d]: 0.510(0.426-0.594)**	**Sensitivity**				
					Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency	No serious indirectness	Serious imprecision[c]	VERY LOW
ATRIA at threshold for risk of ≥7	2	152303	**Median[d]: 0.444(0.137-0.788)**	**Median[d]: 0.607(0.522-0.687)**	**Sensitivity**				
					Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	LOW
AFI 1994	1	90,490	0.990 at standard threshold[no raw data in paper, and no 95% CIs reported]	0.090 at standard threshold[no raw data in paper, and no 95% CIs reported]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
SPAF 1999	1	90,490	0.890 at standard threshold[no raw data in paper, and no 95% CIs reported]	0.290 at standard threshold[no raw data in paper, and no 95% CIs reported]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
FRAMINGHAM	1	90,490	0.920 at standard threshold[no raw data in paper, and no 95% CIs reported]	0.260 at standard threshold[no raw data in paper, and no 95% CIs reported]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
ACC/AHA/ESC guidelines 2006	1	90,490	0.980 at standard threshold[no raw data in paper, and no 95% CIs reported]	0.150 at standard threshold[no raw data in paper, and no 95% CIs reported]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
NICE	1	90,490	1.000 at standard threshold[no raw data in paper, and no 95% CIs reported]	0.090 at standard threshold[no raw data in paper, and no 95% CIs reported]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW

Pooling (meta-analysis) was carried out if there were at least three studies per risk tool with confidence intervals. RevMan and WinBugs were used to carry out the analyses. If pooling was not possible for risk tools with >1 data point then the range and median value of the study point estimates were recorded. If there were only one data point then only the result from the study was recorded.
a) Risk of bias was assessed using the PROBAST checklist. Risk of bias was serious for some risk tools because none of the studies reported any blinding of assessors for risk tool data and outcome status, and most did not report loss to follow up, although follow up and number of events were appropriate. Risk of bias was very serious for the rest of the risk tools because many studies with the aforementioned limitations also had insufficient numbers of events (<100) and/or inappropriately short follow up times (<5 years) to be able to accurately predict risk.
b) Where data were pooled, inconsistency was assessed by visual inspection of the sensitivity/specificity plots, or data (if 2 studies). The evidence was downgraded by 1 increment if there was no overlap of 95% confidence intervals. For single studies no evaluation was made and 'not applicable' was recorded.
c) Imprecision was assessed based on inspection of the confidence region in the meta-analysis or, where meta-analysis has not been conducted, assessed according to the range of confidence intervals in the individual studies. The evidence was downgraded by 1 increment when the confidence interval around the point estimate crossed one of the clinical thresholds (0.90 or 0.60 for sensitivity and 0.5 and 0.1 for specificity), and downgraded by 2 increments when the confidence interval around the point estimate crossed both of the clinical thresholds. The upper clinical threshold marked the point above which recommendations would be possible, and the lower clinical threshold marked the point below which the tool would be regarded as of little clinical use.
d)For unpooled data the median value was given (of data with 95% CIs). If there were an even number of data points in the unpooled data, the data point chosen in the central pair was the one with lower sensitivity, with its paired specificity.

Additional discrimination measures – D statistic

Table 7: Clinical evidence profile: D statistics of prediction tools featured in the studies (see table 3)

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	D statistic (95%CI)	Quality
Q Stroke [female]	1	3180	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.820(0.660-0.990) [Female]	MODERATE
Q Stroke [male]	1	4509	serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[b]	1.150(1.000 to 1.300) [Male]	LOW
CHADS2 [female]	1	3180	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.640(0.490-0.810) [Female]	MODERATE
CHADS2 [male]	1	4509	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.810(0.660 to 0.960) [Male]	MODERATE
CHADSVASC [female]	1	3180	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.670(0.510-0.830) [Female]	MODERATE
CHADSVASC [male]	1	4509	serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[b]	0.970(0.820 to 1.120) [Male]	LOW

a) Risk of bias was assessed using the PROBAST checklist (see Appendix F).Risk of bias was serious for most risk tools because none of the studies reported any blinding of assessors for risk tool data and outcome status, and most did not report loss to follow up, although follow up and number of events were appropriate. Risk of bias was very serious for the Framingham risk tool because the study concerned also had insufficient numbers of events (<100) and/or inappropriately short follow up times (<5 years) to be able to accurately predict risk.
b) The judgement of precision was based on the spread of confidence intervals around the clinically important point at 1.1. If the CIs crossed 1.1 then they were graded as seriously imprecise.

3.2.2 Calibration

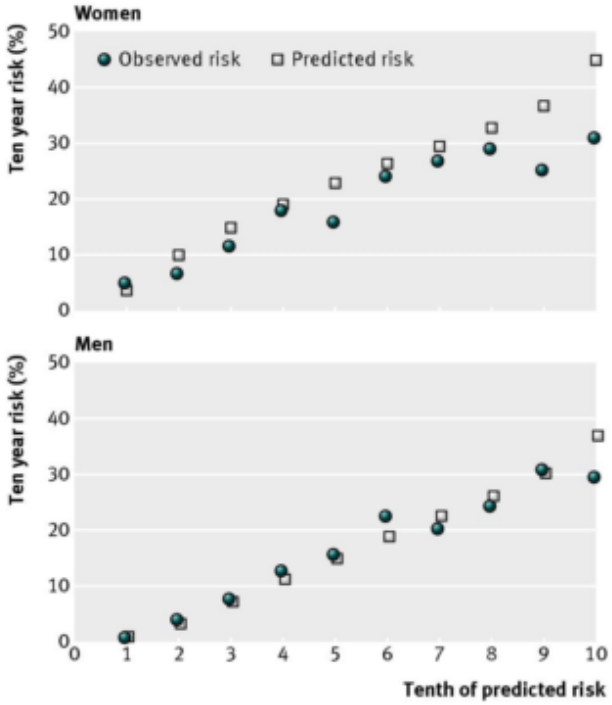
Table 8: Clinical evidence profile: calibration statistics of prediction tools featured in the studies (see table 3)

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	R^2 (95%CI)	Hosmer-Lemeshow statistics	Quality
Q Stroke [female]	1	3180	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision[b]	0.140(0.092-0.187)[Female]	-	MODERATE
Q Stroke [male]	1	4509	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision[b]	0.241(0.193-0.289)[Male]	-	MODERATE
CHADS2 [female]	1	3180	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision[b]	0.091(0.049-0.132)[Female]	-	MODERATE
CHADS2 [male]	1	4509	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision[b]	0.135(0.091-0.179) [Male]	-	MODERATE
CHADSVASC [female]	1	3180	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision[b]	0.096(0.055-0.138)[Female]	-	MODERATE
CHADSVASC [male]	1	4509	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision[b]	0.183(0.137-0.228) [Male]	-	MODERATE

Framingham	1	705	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	-	7.6 (values <20 indicate good calibration. No CIs or p value provided in study.	LOW	

a) Risk of bias was assessed using the PROBAST checklist (see Appendix F).Risk of bias was serious for most risk tools because none of the studies reported any blinding of assessors for risk tool data and outcome status, and most did not report loss to follow up, although follow up and number of events were appropriate. Risk of bias was very serious for the Framingham risk tool because the study concerned also had insufficient numbers of events (<100) and/or inappropriately short follow up times (<5 years) to be able to accurately predict risk.
b) The judgement of precision was based on the spread of confidence intervals around the clinically important point at 0.5. If the CIs crossed 0.5 then they were graded as seriously imprecise.

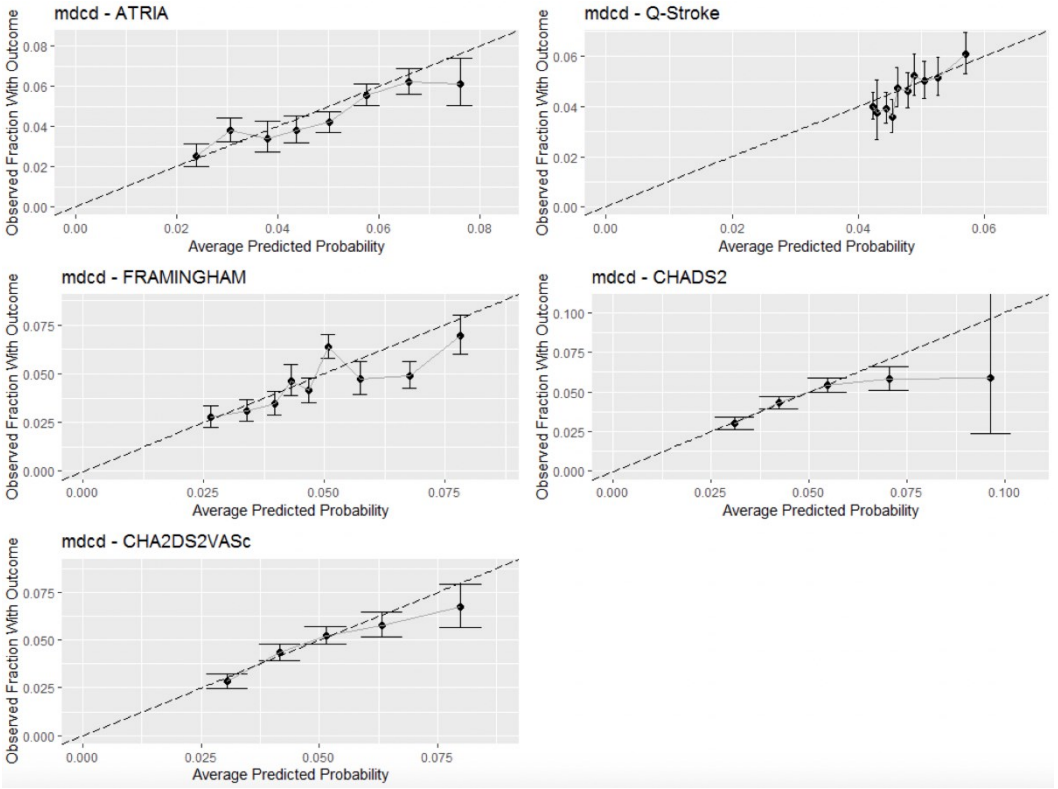
The figure below shows there is good calibration of Q stroke with observed risk[49], with close agreement between predicted and observed risk of stroke across all 10ths of risk.

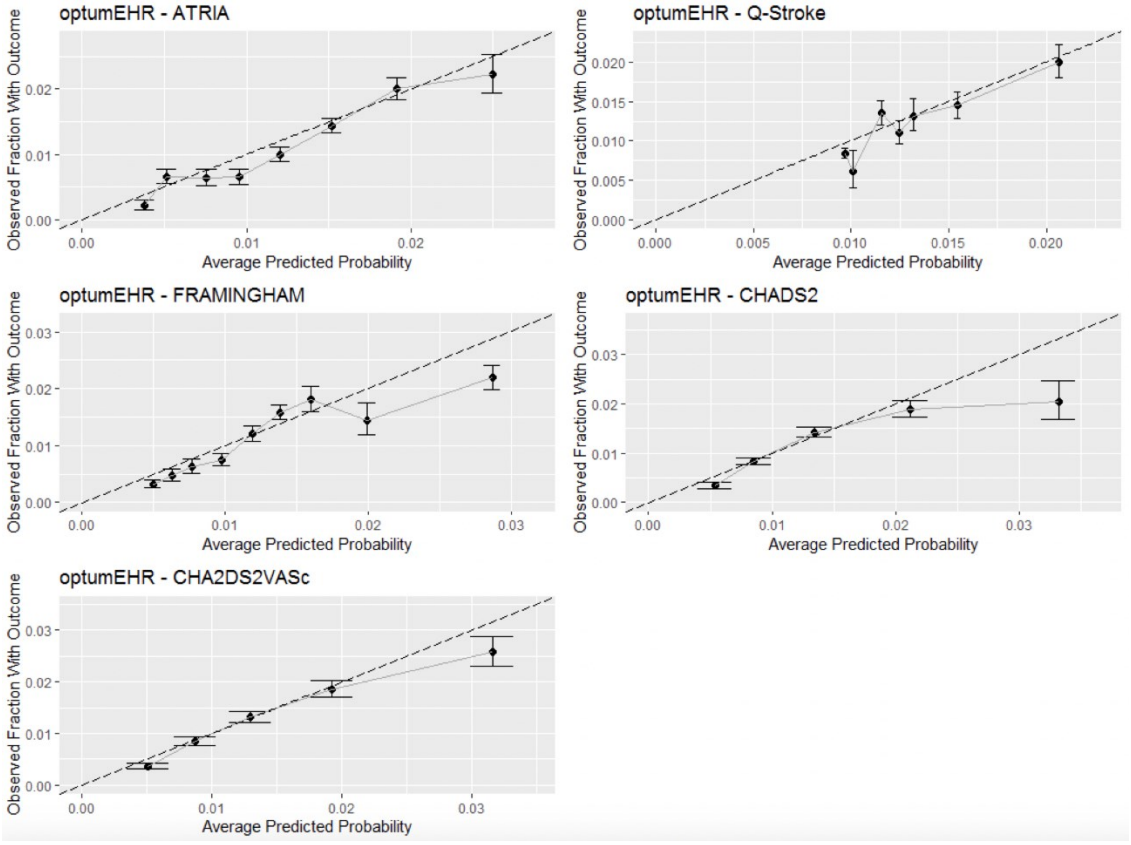


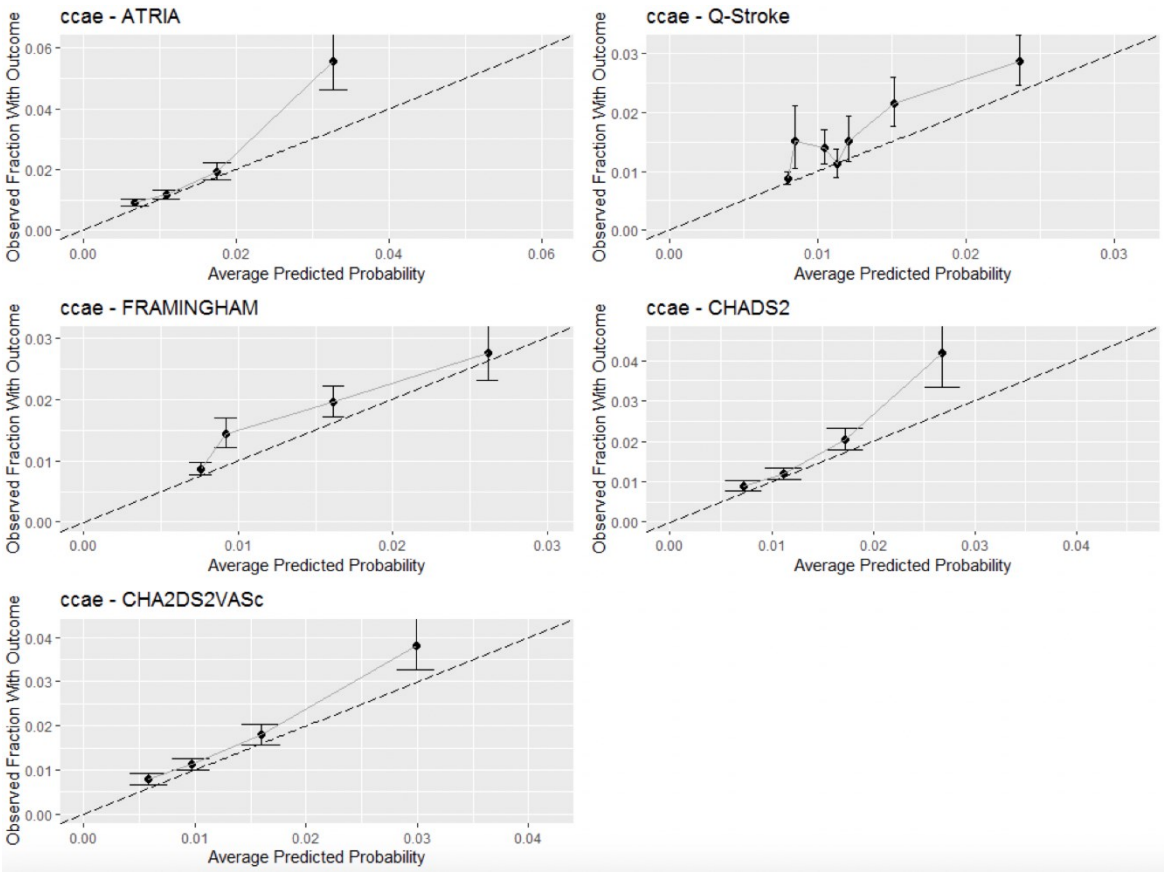
Data from QResearch database version 34, all patients free of free of stroke or transient ischaemic attack

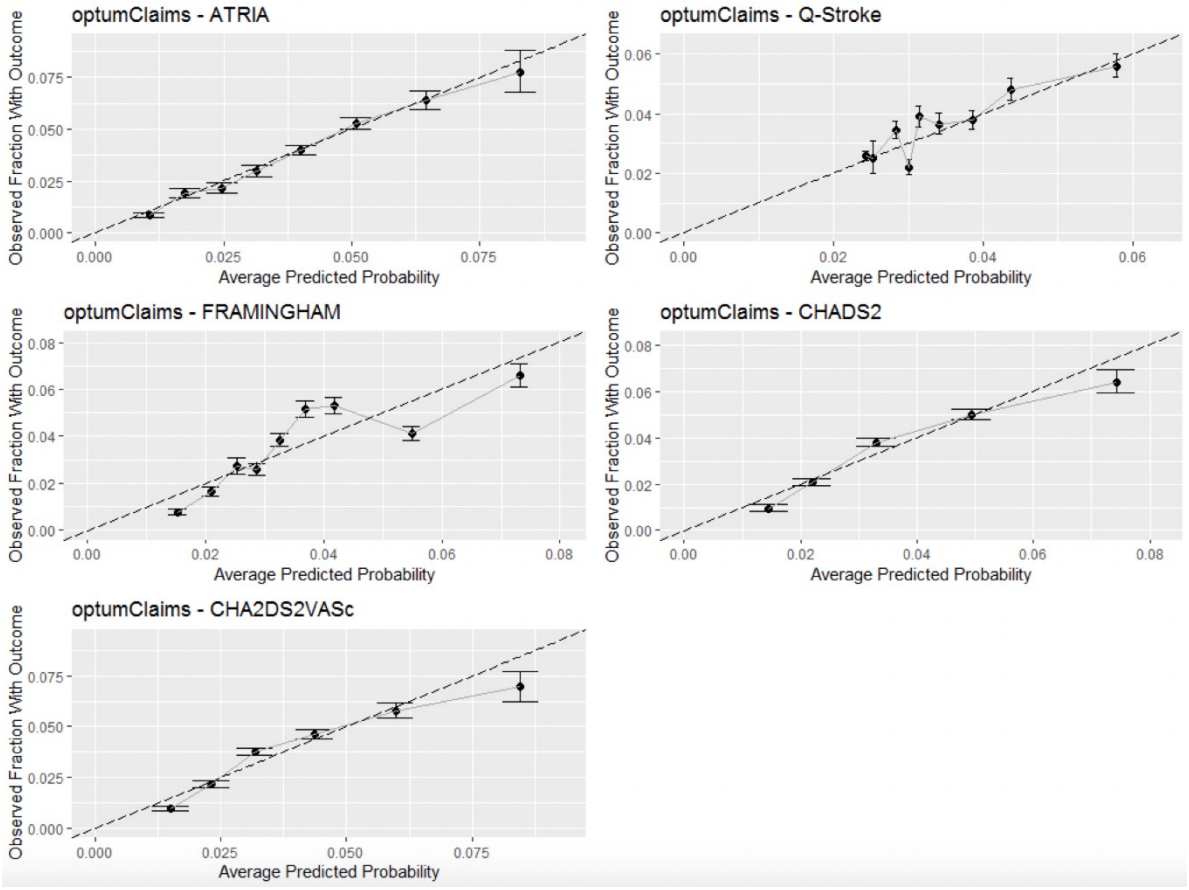
Fig 2 Mean predicted risks and observed risk of stroke or transient ischaemic attack at 10 years by tenth of pred applying the QStroke risk prediction scores to the subset of patients with atrial fibrillation.

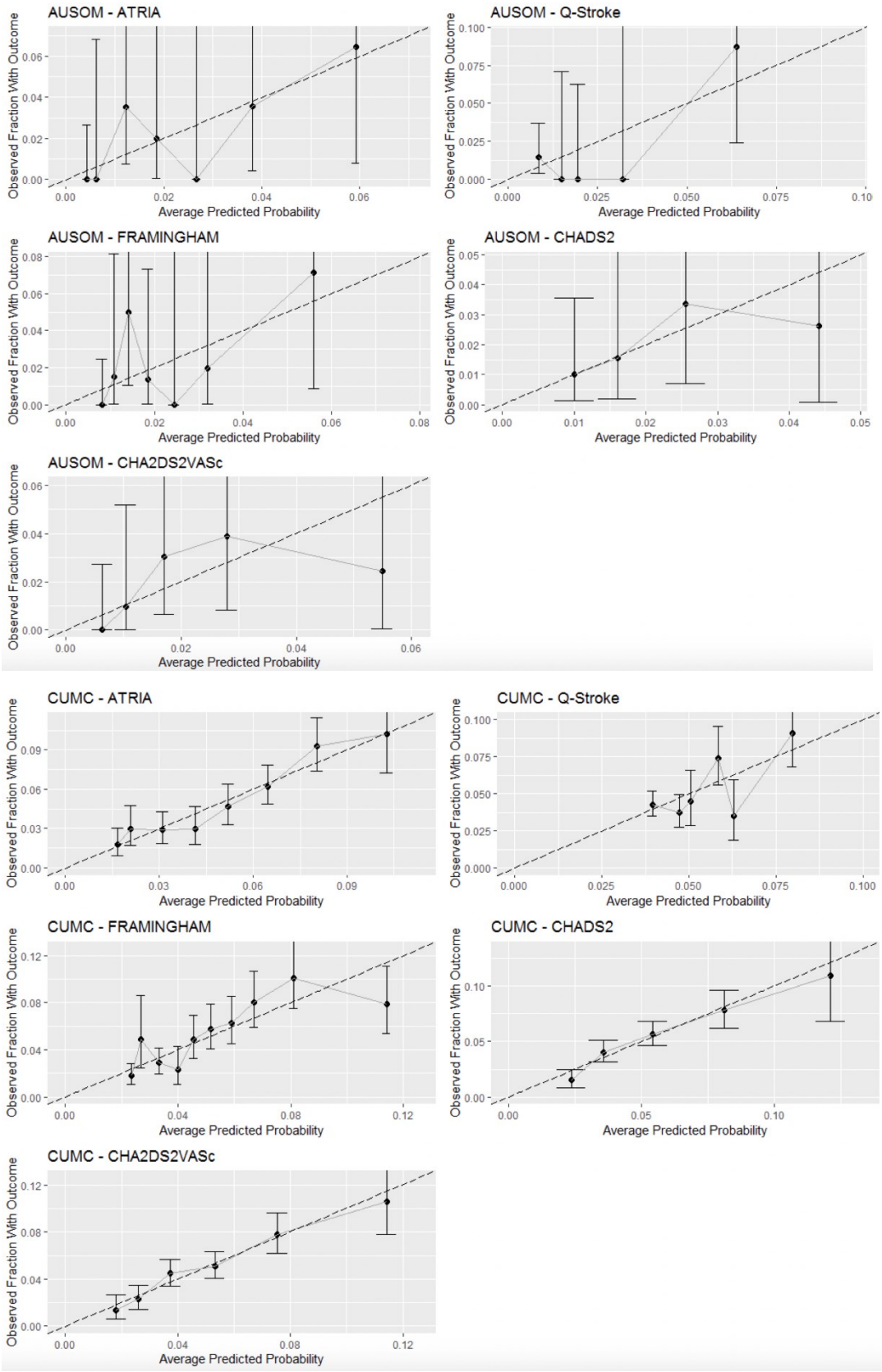
In the study by Reps, 2020[110], calibration plots for the ATRIA, CHADS2, CHADSVASC, FRAMINGHAM and Q STROKE were provided for each of several databases. It was reported that recalibrating the total scores using a linear model worked for ATRIA, Q-stroke, CHADS2 and CHA2DS2VASc but not the Framingham model, which under-estimated risk in the middle risk groups. The plots for each of the different databases are reported here:

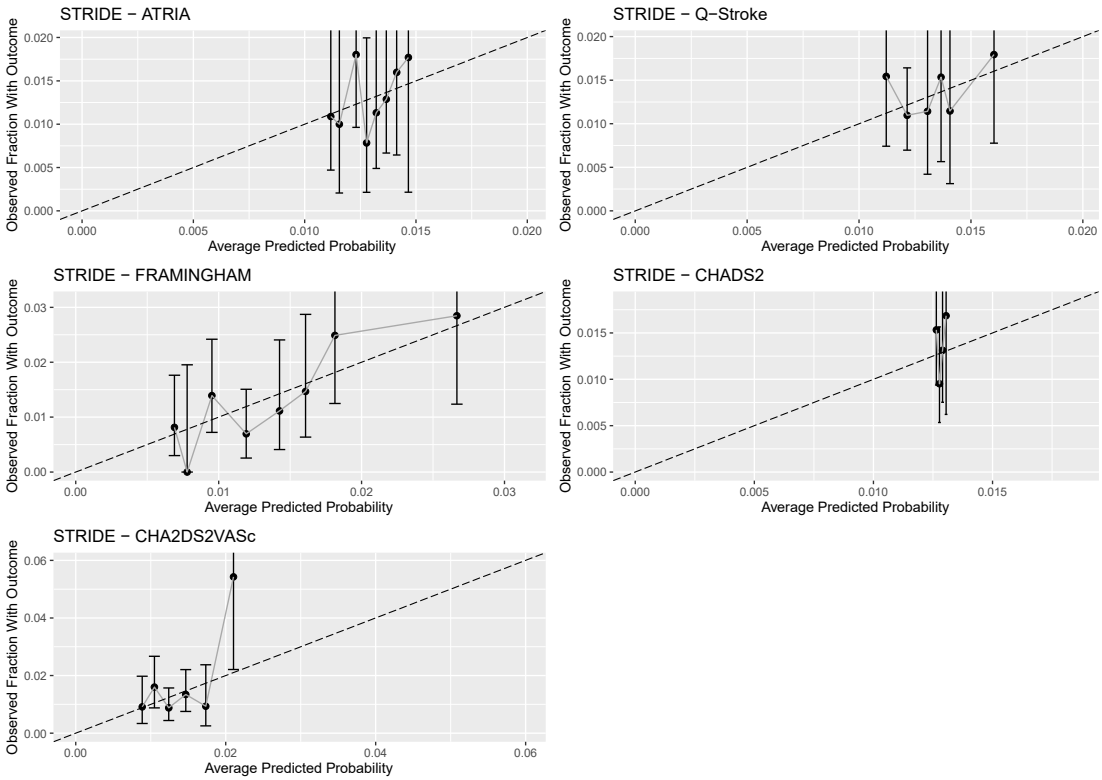

mdcd - ATRIA
mdcd - Q-Stroke
mdcd - FRAMINGHAM
mdcd - CHADS2
mdcd - CHA2DS2VASc
Observed Fraction With Outcome
Average Predicted Probability


optumEHR - ATRIA
optumEHR - Q-Stroke
optumEHR - FRAMINGHAM
optumEHR - CHADS2
optumEHR - CHA2DS2VASc
Observed Fraction With Outcome
Average Predicted Probability


ccae - ATRIA
ccae - Q-Stroke
ccae - FRAMINGHAM
ccae - CHADS2
ccae - CHA2DS2VASc
Observed Fraction With Outcome
Average Predicted Probability


optumClaims - ATRIA
optumClaims - Q-Stroke
optumClaims - FRAMINGHAM
optumClaims - CHADS2
optumClaims - CHA2DS2VASc
Observed Fraction With Outcome
Average Predicted Probability


AUSOM - ATRIA
AUSOM - Q-Stroke
AUSOM - FRAMINGHAM
AUSOM - CHADS2
AUSOM - CHA2DS2VASc
CUMC - ATRIA
CUMC - Q-Stroke
CUMC - FRAMINGHAM
CUMC - CHADS2
CUMC - CHA2DS2VASc
Observed Fraction With Outcome
Average Predicted Probability



P2-CHADSVASC was well-calibrated in the ARIC cohort but less so in the MESA cohort in the study by Maheswari, 2019 [81]

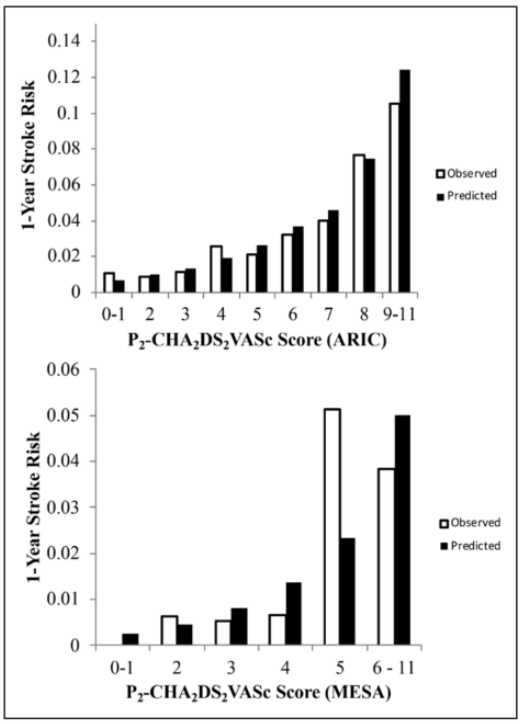


Figure 2. Calibration of the P_2-CHA_2DS_2VASc score in ARIC and MESA. Observed (white bars) and predicted (black bars) 1-year stroke risk for P_2-CHA_2DS_2-VASc score categories in the ARIC study (Atherosclerosis Risk in Communities) and the MESA (Multi-Ethnic Study of Atherosclerosis).

3.2.3 Net Reclassification improvement

Several studies reported the Net Reclassification Improvement (NRI). This is expressed in terms of one (index) risk tool to another (comparator) risk tool and gives a score between -2 and +2 (with +2 representing the best possible performance of the index tool relative to the comparator, and -2 the worst). The score represents the net improvement of the index test relative to the comparator in terms of the proportion of true cases (judged by later development of stroke/TE) that are correctly up-classified by the tool (relative to any false negative classifications yielded by the comparator), and the proportion of false cases (judged by the lack of later stroke/TE) that are correctly down-classified by the tool (relative to any false positive classifications yielded by the comparator). Meanwhile, incorrect up-classification or incorrect down-classification of the index relative to the comparator convey negative scores to the NRI, and so if a score is negative overall this indicates the index is less accurate than the comparator.

NRI data are given below for each risk tool comparison. The data have been divided into two tables, by comparator.

Table 9: Clinical evidence profile: NRI of prediction tools featured in the studies (see table 3) with CHADS2 as the comparator

Prediction tool comparison	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	[NRI(95% CI)](95%CI)	Quality
ATRIA versus CHADS2	4	259,504	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	No serious imprecision	-0.0 **POOLED EFFECT: Random effects NRI +0.130 (+0.050 to +0.220); I^2=98%**	VERY LOW
R2CHADS2 (71 point) versus CHADS2	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	-0.015 (-0.036 to 0.006)	VERY LOW
R2CHADS2 versus CHADS2	1	16,360	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.226(0.125 to 0.307)	LOW

CHADS2 KDIGO versus CHADS2	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.026(-0.049 to -0.002)	LOW
CHADS2 Alb versus CHADS2	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	-0.018 (-0.026 to 0.028)	VERY LOW
CHADS2 eGFR versus CHADS2	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.006 (-0.017 to 0.030)	VERY LOW
CHASDS2 with vascular disease versus CHADS2	1	2002	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.400 (0.000 to +0.800)	MODERATE

Pooling (meta-analysis) was carried out if there were at least two studies per risk tool with confidence intervals. RevMan was used to carry out the analyses. If pooling was not possible for risk tools with >1 data point then the range and median value of the study point estimates were recorded. If there were only one data point then only the result from the study was recorded.

a) Risk of bias was assessed using the PROBAST checklist (see Appendix F).Risk of bias was serious for some risk tools because none of the studies reported any blinding of assessors for risk tool data and outcome status, and most did not report loss to follow up, although follow up and number of events were appropriate. Risk of bias was very serious for the rest of the risk tools because many studies with the aforementioned limitations also had insufficient numbers of events (<100) and/or inappropriately short follow up times (<5 years) to be able to accurately predict risk.

b) Where data were pooled, an I^2 of 50-74% was deemed serious inconsistency and an I^2 of 75% or above was deemed very serious inconsistency. If no pooling were possible, inconsistency was assessed by inspection of the degree of overlap of confidence intervals between studies: if one of more Cis did not overlap then a rating of serious inconsistency was given. Reasons for heterogeneity between studies may include geographical/cultural/ethnic differences. Clinically the studies appeared reasonably homogeneous, with similar rates of hypertension, diabetes and former stroke.

c) The judgement of precision was based on the spread of confidence intervals. If the lower 95% CI passed across 0 then this was graded as seriously imprecise

Table 10: Clinical evidence profile: NRI of prediction tools featured in the studies (see table 3) with CHADSVASC (or CHADSVASC derivatives) as the comparator

Prediction tool comparison	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	[NRI(95% CI)](95%CI)	Quality
ATRIA versus CHADSVASC	3	210,053	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	No serious imprecision	to 0.300)[123] **POOLED EFFECT: Random effects NRI +0.230 (+0.200 to +0.250); I^2=79%**	VERY LOW
Age-modified CHADSVASC versus CHADSVASC	1	124,271	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.039 (0.0216 to 0.0459)	MODERATE
CHADS2 versus CHADSVASC	8	210,854	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	)[139] **POOLED EFFECT: Random effects NRI -0.020 (-0.060 to +0.020); I^2=84%**	VERY LOW
Revised CHADS2 versus CHADSVASC	1	90,490	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Not applicable	+0.070	LOW
Framingham versus CHADSVASC	1	90,490	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Not applicable	+0.120	LOW
SPAF 1999 versus CHADSVASC	1	90,490	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Not applicable	+0.120	LOW
ACC/AHA/ESC versus CHADSVASC	1	90,490	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Not applicable	+0.070	LOW

NICE versus CHADSVASC	1	90,490	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Not applicable	+0.000	LOW
AFI 1994 versus CHADSVASC	1	90,490	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Not applicable	+0.000	LOW
CHADS2 versus mCHADSVASC	1	997	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	-0.100(-0.280 to 0.080)	VERY LOW
CHADS2 versus mCHADSVA	1	997	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	-0.030 (-0.210 to 0.160)	VERY LOW
mCHADSVASC versus mCHADSVA	1	997	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.110(0.010 to 0.200)	LOW
P2-CHADSVASC versus CHADSVASC	2	2929	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	No serious imprecision	**POOLED EFFECT: Random effects NRI +0.330 (+0.100 to +0.570); I^2=53%**	VERY LOW

Pooling (meta-analysis) was carried out if there were at least two studies per risk tool with confidence intervals. RevMan was used to carry out the analyses. If pooling was not possible for risk tools with >1 data point then the range and median value of the study point estimates were recorded. If there were only one data point then only the result from the study was recorded.

a) Risk of bias was assessed using the PROBAST checklist (see Appendix F).Risk of bias was serious for some risk tools because none of the studies reported any blinding of assessors for risk tool data and outcome status, and most did not report loss to follow up, although follow up and number of events were appropriate. Risk of bias was very serious for the rest of the risk tools because many studies with the aforementioned limitations also had insufficient numbers of events (<100) and/or inappropriately short follow up times (<5 years) to be able to accurately predict risk.

b) Where data were pooled, an I^2 of 50-74% was deemed serious inconsistency and an I^2 of 75% or above was deemed very serious inconsistency. If no pooling were possible, inconsistency was assessed by inspection of the degree of overlap of confidence intervals between studies: if one of more Cis did not overlap then a rating of serious inconsistency was given. Reasons for heterogeneity between studies may include geographical/cultural/ethnic differences. Clinically the studies appeared reasonably homogeneous, with similar rates of hypertension, diabetes and former stroke.

c) The judgement of precision was based on the spread of confidence intervals. If the lower 95% CI passed across 0 then this was graded as seriously imprecise

Q Stroke versus CHADSVASC

Data relevant to classification were given in one study[49], but there was insufficient information on true events and non-events to allow calculation of the NRI

Q Stroke versus CHADS2

Data relevant to classification were given in one study[49], but there was insufficient information on true events and non-events to allow calculation of the NRI

Sum of CrCL <60mL/min and prior stroke/TIA versus **R2CHADS 2**

+0.024 (-0.077 to + 0.029)[100]

3.3 Economic evidence

3.3.1 Included studies

No relevant health economic studies were identified.

3.3.2 Excluded studies

No health economic studies that were relevant to this question were excluded due to assessment of limited applicability or methodological limitations.

See also the health economic study selection flow chart in appendix H.

4 The committee's discussion of the evidence

4.1 Interpreting the evidence

4.1.1 The outcomes that matter most

The committee agreed that the most critical accuracy data for decision-making were sensitivity/specificity, net reclassification improvement (NRI), and calibration data. Sensitivity and specificity measures for specific thresholds were deemed useful outcomes, because they allow for the differing importance placed on sensitivity or specificity, and are specific to clinically relevant thresholds of risk. Reclassification measures were also favoured because they are sensitive to small changes in a tool, such as an additional parameter contributing to the score. Calibration measures were deemed the most useful outcome, however, because they give the most realistic impression of how well a tool predicts the actual risk of the event at a particular test threshold. Unfortunately, because these measures were unavailable for many tools they played a smaller part than anticipated in decision-making.

C statistics data were deemed important, but less critical than the other outcomes, because they do not take the relative importance of sensitivity and specificity into account. For example, tool A may have a higher C statistic than tool B, but this superior C statistic may be because tool A tends towards very high overall specificity, even though its overall sensitivity may be inferior to that of tool B. If sensitivity is deemed the more important aspect of predictive accuracy, then the C statistic may be a misleading measure in this context. In addition, C statistics effectively measure the overall accuracy at *all* risk thresholds defined by a tool (quantified by the area under the curve described by sensitivity and 1-specificity co-ordinates at each possible risk threshold). In practice, however, a test will be used at a specific threshold, relating to the point where risk is deemed to change from an acceptable to an unacceptable risk, and so the overall accuracy at all thresholds, including clinically non-relevant ones, may be misleading. Finally, C-statistics are insensitive to small changes in the risk model (when new prognostic factors are added to an existing model). Nevertheless, the committee included the C statistic as an outcome as it gives a general measure of a tool's ability to differentiate between high and low risk patients, and is commonly reported in these studies.

The committee confirmed that the recommendations on anticoagulation applied to all patients with AF irrespective of whether they were symptomatic, to all categories of AF (paroxysmal, persistent and permanent), to patients following cardioversion considered at continuing risk of arrhythmia recurrence, and to patients with atrial flutter.

4.1.2 The quality of the evidence

Evidence was generally deemed low or very low quality. Risk of bias was serious or very serious due to unclear methodology in terms of blinding of risk tool and outcome data, and in many studies the follow up time was short (<5 years) or involved few events (<100). The quality was also affected by serious or very serious heterogeneity.

4.1.3 Benefits and harms

Sub-optimal predictive accuracy can lead to two harms, in the context of predicting stroke in people with AF. Sub-optimal accuracy caused by low sensitivity will lead to more people having strokes or thromboembolic events because they are incorrectly deemed to be at too low a level of risk to be prescribed anticoagulants. Sub-optimal accuracy caused by low

specificity will lead to more people having unnecessary bleeding episodes or other side-effects of anticoagulants because they have been prescribed anticoagulants when their risk of stroke is actually low.

The judgement of which is the most important harm depends on the severity of these harms and also their probability of occurring. Scoring systems generally have a trade-off between sensitivity and specificity. The committee agreed that the greater emphasis should be on avoiding strokes because bleeding events were both less probable than strokes and also less likely to have such serious consequences as strokes if they occurred. This was judged to be particularly so given the new generation of anticoagulants: non-vitamin K antagonist oral anticoagulants (DOACs). Thus tools favouring sensitivity were preferred. However it was also recognised that it is easy to design a perfectly sensitive test if specificity is not considered at all (for example, simply giving anticoagulants to all people with AF is equivalent to the use of a perfectly sensitive but completely non-specific test). It was recognised that the ideal tool would have high sensitivity but also have enough specificity to allow the people with lowest risk to avoid unnecessary anticoagulation, with the excess risk that would entail.

The CHADS2 was similar to the CHADSVASC in terms of the C statistic, but it was felt too insensitive at even the lowest thresholds to be able to rival the CHADSVASC. However, there were two new tools that were regarded as potential rivals to the CHADSVASC in terms of predictive accuracy: the Q stroke and ATRIA.

The Q stroke was viewed as highly promising, as it had excellent sensitivity and reasonable specificity at the 85th percentile of scores. The D statistic point estimates of the Q stroke were numerically superior to those in the CHADSVASC, with the Q stroke values in men suggesting a clinically important degree of discrimination. However, there was some overlap of 95% confidence intervals between Q stroke and CHADSVASC suggesting that these differences could be explained by sampling error. There was also good calibration of the Q stroke at lower risks, particularly in men, and the R^2 data were again numerically superior to the CHADSVASC, although again the overlap of 95% confidence intervals suggested that sampling error could be a factor. The above data were based on one derivation/validation study from the UK. This study contained a separate sample for the validation analysis but the committee noted that despite the obvious potential of this tool, a single study was insufficient to inform recommendations, and that further work in other AF samples was required before this tool could be recommended over the CHADSVASC. A further paper was subsequently found during the final literature search reruns in September 2020. This was a large-scale external validation study, using data from women with AF who were not on anticoagulants, drawn from several separate databases from the USA and South Korea. This provided precise estimates for C statistics and calibration plots, but the data from this paper did not support the evidence from the UK study. C statistics were inferior to the CHADSVASC and ATRIA, whilst the calibration plots did not suggest any advantage over CHADSVASC.

The ATRIA was also regarded as an excellent tool, with a C statistic that was comparable to the CHADSVASC, and with similar calibration performance. It also had a significantly better NRI compared to the CHADSVASC. However this was largely due to down-classification of non-events. Accordingly, this was accompanied by better specificity but lower sensitivity (around 0.80+) than the CHADSVASC (around 0.90+) at standard thresholds (threshold of $\geq$2 for CHADSVASC and $\geq$6 for ATRIA). At lower ATRIA thresholds the sensitivity/specificity profile of ATRIA was very similar to CHADSVASC (at CHADSVASC thresholds of $\geq$1 or $\geq$2) but did not become any better than it. The decision of the committee was therefore that CHADSVASC was slightly more useful because of its better ability to ensure that people truly at risk of stroke were anticoagulated.

In addition to ATRIA having potentially more harms than CHADSVASC in terms of ATRIA leading to more people at risk of stroke not being anticoagulated, the ATRIA was also believed to be more difficult to use. The committee discussed the time delays in getting a dip-stick assessment of proteinuria done and retesting eGFR for the ATRIA, although it was pointed out that ATRIA might, on occasions, be able to utilise data already in the patients' notes rather than requiring the acquisition of new data.

Thus CHADSVASC was regarded as the best available tool. The ideal threshold for the CHADSVASC in terms of anticoagulation was agreed to be $\geq$2, as this gave an excellent compromise between high sensitivity and reasonable specificity. This fitted with current practice. The reviewed data did not allow the committee to decide if men and women should have different thresholds.

4.1.4 Cost effectiveness and resource use

No relevant health economic analyses were identified for this review. The committee discussed the different resource use for the different tests, in particular ATRIA compared to CHADSVASC. It was noted that testing proteinuria for ATRIA required a urine dipstick test. This can be particularly challenging with older frail patients who may require assistance to provide a urine sample and therefore may incur additional cost over CHADVASC. In addition, further blood tests would be needed for ATRIA as the eGFR would need repeating. The committee noted that these additional tests would create delays and disruption to clinics.

The committee also discussed the potential harm associated with ATRIA compared to CHADSVASC, in terms of ATRIA leading to more people at risk of stroke not being anticoagulated (as a result of the lower sensitivity). The committee noted that this harm would likely make ATRIA less cost-effective than CHADSVASC due to the high cost of a stroke to the NHS and detrimental impact on QALYs. This would likely outweigh the increased anticoagulation as a result of the lower specificity of CHADSVASC. Health economic modelling of ATRIA compared to CHADSVASC was not prioritised by the guideline committee as other areas of the guideline were considered to have a greater potential resource impact (ablation and anticoagulation).

The committee agreed that there was not sufficient clinical evidence of superiority for ATRIA to warrant a change in practice and the potential harms and costs associated with this new tool.

4.1.5 Other factors the committee took into account

Patient views are central when considering the trade-off between the benefits and harms. The committee agreed that it is important to ensure that information and education are provided to ensure the benefits and harms are fully understood (see the NICE patient experience guideline).

References

1. Aakre CA, McLeod CJ, Cha SS, Tsang TS, Lip GY, Gersh BJ. Comparison of clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation. Stroke. 2014; 45(2):426-431

2. Abraham JM, Larson J, Chung MK, Curtis AB, Lakshminarayan K, Newman JD et al. Does CHA2DS2-VASc improve stroke risk stratification in postmenopausal women with atrial fibrillation? American Journal of Medicine. 2013; 126(12):1143.e1141-1143.e1148

3. Abumuaileq RR, Abu-Assi E, Lopez-Lopez A, Raposeiras-Roubin S, Rodriguez-Manero M, Martinez-Sande L et al. Comparison between CHA2DS2-VASc and the new R2CHADS2 and ATRIA scores at predicting thromboembolic event in non-anticoagulated and anticoagulated patients with non-valvular atrial fibrillation. BMC Cardiovascular Disorders. 2015; 15:156

4. Abumuaileq RR, Abu-Assi E, Raposeiras-Roubin S, Lopez-Lopez A, Redondo-Dieguez A, Alvarez-Iglesias D et al. Evaluation of SAMe-TT2R2 risk score for predicting the quality of anticoagulation control in a real-world cohort of patients with non-valvular atrial fibrillation on vitamin-K antagonists. Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology. 2015; 17(5):711-717

5. Al-Radeef MY. Status of CHADS2 and CHA2DS2-VASc scores in predicting risk of stroke and its prevention in Iraqi patients with atrial fibrillation. Iraqi Journal of Pharmaceutical Sciences. 2019; 28(1):101-105

6. Al-Turaiki AM, Al-Ammari MA, Al-Harbi SA, Khalidi NS, Alkatheri AM, Aldebasi TM et al. Assessment and comparison of CHADS2, CHA2DS2-VASc, and HAS-BLED scores in patients with atrial fibrillation in Saudi Arabia. Annals of Thoracic Medicine. 2016; 11(2):146-150

7. Alraies MC, Norby FL, Chen LY. The CHA2DS2-VASc score and the risk of ischemic stroke in community-dwelling individuals with and without atrial fibrillation: The Atherosclerosis Risk In Communities (ARIC) study. Journal of the Saudi Heart Association. 2017; 29(4):316-317

8. Andersson T, Magnuson A, Bryngelsson IL, Frobert O, Henriksson KM, Edvardsson N et al. Patients with atrial fibrillation and outcomes of cerebral infarction in those with treatment of warfarin versus no warfarin with references to CHA2DS2-VASc score, age and sex - A Swedish nationwide observational study with 48 433 patients. PloS One. 2017; 12(5):e0176846

9. Asberg S, Henriksson KM, Farahmand B, Asplund K, Norrving B, Appelros P et al. Ischemic stroke and secondary prevention in clinical practice: a cohort study of 14,529 patients in the Swedish Stroke Register. Stroke. 2010; 41(7):1338-1342

10. Aspberg S, Chang Y, Atterman A, Bottai M, Go AS, Singer DE. Comparison of the ATRIA, CHADS2, and CHA2DS2-VASc stroke risk scores in predicting ischaemic stroke in a large Swedish cohort of patients with atrial fibrillation. European Heart Journal. 2016; 37(42):3203-3210

11. Atzema CL, Dorian P, Fang J, Tu JV, Lee DS, Chong AS et al. A clinical decision instrument for 30-day death after an emergency department visit for atrial fibrillation:

the atrial fibrillation in the emergency room (AFTER) study. Annals of Emergency Medicine. 2015; 66(6):658-668.e656

12. Banerjee A, Fauchier L, Bernard-Brunet A, Clementy N, Lip GY. Composite risk scores and composite endpoints in the risk prediction of outcomes in anticoagulated patients with atrial fibrillation. The Loire Valley Atrial Fibrillation Project. Thrombosis and Haemostasis. 2014; 111(3):549-556

13. Banerjee A, Fauchier L, V'Ourch P, Andres CR, Taillandier S, Halimi JM et al. Renal impairment and stroke risk assessment in patients with atrial fibrillation: The loire valley atrial fibrillation project. European Journal of Preventive Cardiology. 2013; 1:S61

14. Banerjee A, Fauchier L, Vourc'h P, Andres CR, Taillandier S, Halimi JM et al. Renal impairment and ischemic stroke risk assessment in patients with atrial fibrillation: the Loire Valley Atrial Fibrillation Project. Journal of the American College of Cardiology. 2013; 61(20):2079-2087

15. Baruch L, Gage BF, Horrow J, Juul-Moller S, Labovitz A, Persson M et al. Can patients at elevated risk of stroke treated with anticoagulants be further risk stratified? Stroke. 2007; 38(9):2459-2463

16. Basili S, Loffredo L, Pastori D, Proietti M, Farcomeni A, Vestri AR et al. Carotid plaque detection improves the predictive value of CHA2DS2-VASc score in patients with non-valvular atrial fibrillation: The ARAPACIS Study. International Journal of Cardiology. 2017; 231:143-149

17. Berg DD, Ruff CT, Jarolim P, Giugliano RP, Nordio F, Lanz HJ et al. Performance of the ABC scores for assessing the risk of stroke or systemic embolism and bleeding in patients with atrial fibrillation in ENGAGE AF-TIMI 48. Circulation. 2019; 139(6):760-771

18. Borre ED, Goode A, Raitz G, Shah B, Lowenstern A, Chatterjee R et al. Predicting thromboembolic and bleeding event risk in patients with non-valvular atrial fibrillation: a systematic review. Thrombosis and Haemostasis. 2018; 118(12):2171-2187

19. Camelo-Castillo A, Rivera-Caravaca JM, Marin F, Vicente V, Lip GYH, Roldan V. Predicting Adverse Events beyond Stroke and Bleeding with the ABC-Stroke and ABC-Bleeding Scores in Patients with Atrial Fibrillation: The Murcia AF Project. Thrombosis and Haemostasis. 2020; 120(8):1200-1207

20. Chan PH, Lau CP, Tse HF, Chiang CE, Siu CW. CHA2DS2-VASc recalibration with an additional age category (50-64 years) enhances stroke risk stratification in chinese patients with atrial fibrillation. Canadian Journal of Cardiology. 2016; 32(12):1381-1387

21. Chao TF, Lip GY, Liu CJ, Tuan TC, Chen SJ, Wang KL et al. Validation of a modified CHA2DS2-VASc score for stroke risk stratification in Asian patients with atrial fibrillation: a nationwide cohort study. Stroke. 2016; 47(10):2462-2469

22. Chao TF, Liu CJ, Chen SJ, Wang KL, Lin YJ, Chang SL et al. Atrial fibrillation and the risk of ischemic stroke: does it still matter in patients with a CHA2DS2-VASc score of 0 or 1? Stroke. 2012; 43(10):2551-2555

23. Chao TF, Liu CJ, Wang KL, Lin YJ, Chang SL, Lo LW et al. Should atrial fibrillation patients with 1 additional risk factor of the CHA2DS2-VASc score (beyond sex) receive oral anticoagulation? Journal of the American College of Cardiology. 2015; 65(7):635-642

24. Chao TF, Wang KL, Liu CJ, Lin YJ, Chang SL, Lo LW et al. Age threshold for increased stroke risk among patients with atrial fibrillation: a nationwide cohort study from Taiwan. Journal of the American College of Cardiology. 2015; 66(12):1339-1347

25. Charfeddine S, Ben Halima M, Krichene S, Hassine M, Battikh K, Larbi N et al. SAMe-TT2R2 in Tunisian population: results from the NATURE AF registry. Archives of Cardiovascular Diseases Supplements. 2020; 12(1):124

26. Chen YY, Lin YJ, Chien KL, Chen SA. Model-based point scoring system for predicting risk of stroke in the era of catheter of atrial fibrillation. Journal of Arrhythmia. 2019; 35 (Suppl 1):238

27. Dalgaard F, Pieper K, Verheugt F, Camm AJ, Fox KA, Kakkar AK et al. GARFIELD-AF model for prediction of stroke and major bleeding in atrial fibrillation: a Danish nationwide validation study. BMJ Open. 2019; 9(11):e033283

28. Di Toro D, Hadid C, Gallino S, Labadet C. Application and comparison of the CHADS2 and CHA2DS2-VASC risk scores in a population with atrial fibrillation. Argentine Journal of Cardiology. 2013; 81(6):463-467

29. Dzeshka MS, Lane DA, Lip GYH. Stroke and bleeding risk in atrial fibrillation: navigating the alphabet soup of risk-score acronyms (CHADS2, CHA2DS2-VASc, R2CHADS2, HAS-BLED, ATRIA, and more). Clinical Cardiology. 2014; 37(10):634-644

30. Fang MC, Go AS, Chang Y, Borowsky L, Pomernacki NK, Singer DE. Comparison of risk stratification schemes to predict thromboembolism in people with nonvalvular atrial fibrillation. Journal of the American College of Cardiology. 2008; 51(8):810-815

31. Fauchier L, Clementy N, Bisson A, Ivanes F, Angoulvant D, Babuty D et al. Should atrial fibrillation patients with only 1 nongender-related CHA2DS2-VASc risk factor be anticoagulated? Stroke. 2016; 47(7):1831-1836

32. Forslund T, Wettermark B, Wandell P, von Euler M, Hasselstrom J, Hjemdahl P. Risks for stroke and bleeding with warfarin or aspirin treatment in patients with atrial fibrillation at different CHA(2)DS(2)VASc scores: experience from the Stockholm region. European Journal of Clinical Pharmacology. 2014; 70(12):1477-1485

33. Fox KAA, Lucas JE, Pieper KS, Bassand JP, Camm AJ, Fitzmaurice DA et al. Improved risk stratification of patients with atrial fibrillation: an integrated GARFIELD-AF tool for the prediction of mortality, stroke and bleed in patients with and without anticoagulation. BMJ Open. 2017; 7(12):e017157

34. Friberg L, Benson L, Rosenqvist M, Lip GY. Assessment of female sex as a risk factor in atrial fibrillation in Sweden: nationwide retrospective cohort study. BMJ. 2012; 344:e3522

35. Friberg L, Lund LH. Heart failure: a weak link in CHA2 DS2 -VASc. ESC heart failure. 2018; 5(3):231-239

36. Friberg L, Rosenqvist M, Lip GY. Evaluation of risk stratification schemes for ischaemic stroke and bleeding in 182 678 patients with atrial fibrillation: the Swedish Atrial Fibrillation cohort study. European Heart Journal. 2012; 33(12):1500-1510

37. Friberg L, Skeppholm M, Terent A. Benefit of anticoagulation unlikely in patients with atrial fibrillation and a CHA2DS2-VASc score of 1. Journal of the American College of Cardiology. 2015; 65(3):225-232

38. Gage BF, van Walraven C, Pearce L, Hart RG, Koudstaal PJ, Boode BS et al. Selecting patients with atrial fibrillation for anticoagulation: stroke risk stratification in patients taking aspirin. Circulation. 2004; 110(16):2287-2292

39. 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. Journal of the American Medical Association. 2001; 285(22):2864-2870

40. Gazova A, Leddy JJ, Rexova M, Hlivak P, Hatala R, Kyselovic J. Predictive value of CHA2DS2-VASc scores regarding the risk of stroke and all-cause mortality in patients with atrial fibrillation (CONSORT compliant). Medicine. 2019; 98(31):e16560

41. Guo Y, Apostolakis S, Blann AD, Wang H, Zhao X, Zhang Y et al. Validation of contemporary stroke and bleeding risk stratification scores in non-anticoagulated Chinese patients with atrial fibrillation. International Journal of Cardiology. 2013; 168(2):904-909

42. Guo Y, Chen Y, Lane DA, Liu L, Wang Y, Lip GYH. Mobile health technology for atrial fibrillation management integrating decision support, education, and patient involvement: mAF App trial. American Journal of Medicine. 2017; 130(12):1388-1396.e1386

43. Gupta N, Haft JI, Bajaj S, Samuel A, Parikh R, Pandya D et al. Role of the combined CHADS2 score and echocardiographic abnormalities in predicting stroke in patients with paroxysmal atrial fibrillation. Journal of Clinical Neuroscience. 2012; 19(9):1242-1245

44. Hijazi Z, Lindahl B, Oldgren J, Andersson U, Lindback J, Granger CB et al. Repeated measurements of cardiac biomarkers in atrial fibrillation and validation of the ABC stroke score over time. Journal of the American Heart Association. 2017; 6(6):23

45. Hijazi Z, Lindback J, Alexander JH, Hanna M, Held C, Hylek EM et al. The ABC (age, biomarkers, clinical history) stroke risk score: a biomarker-based risk score for predicting stroke in atrial fibrillation. European Heart Journal. 2016; 37(20):1582-1590

46. Hijazi Z, Lindback J, Ostlund O, Siegbahn A, Alexander JH, Granger CB et al. External validation of the biomarker-based ABC-stroke risk score for atrial fibrillation. European Heart Journal. 2015; 1:710-711

47. Hijazi Z, Oldgren J, Lindback J, Alexander JH, Connolly SJ, Eikelboom JW et al. The novel biomarker-based ABC (age, biomarkers, clinical history)-bleeding risk score for patients with atrial fibrillation: a derivation and validation study. Lancet. 2016; 387(10035):2302-2311

48. Hijazi Z, Oldgren J, Lindback J, Alexander JH, Hanna M, Hylek EM et al. Performance of the novel biomarker-based abcstroke risk score over time in patients with atrial fibrillation. Circulation. 2016; 134(Suppl. 1):A14726

49. Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. BMJ. 2013; 346:f2573

50. Hippisley-Cox J, Coupland C, Brindle P. The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study. BMJ Open. 2014; 4(8):e005809

51. Holt TA, Kirkpatrick S, Hislop J, Kearley K, Mollison J, Yu LM et al. Barriers to a software reminder system for risk assessment of stroke in atrial fibrillation: a process evaluation of a cluster randomised trial in general practice. British Journal of General Practice. 2018; 68(677):E844-E851

52. Horne BD, Jacobs V, May HT, Graves KG, Bunch TJ. Augmented intelligence decision tool for stroke prediction combines factors from CHA2DS2-VASc and the intermountain risk score for patients with atrial fibrillation. Journal of Cardiovascular Electrophysiology. 2019; 30(9):1452-1461

53. Hu WS, Lin CL. CHA2DS2-VASc score for ischaemic stroke risk stratification in patients with chronic obstructive pulmonary disease with and without atrial fibrillation: a nationwide cohort study. Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology. 2018; 20(4):575-581

54. Huang J, Wu SL, Xue YM, Fei HW, Lin QW, Ren SQ et al. Association of CHADS2 and CHA2DS2-VASc scores with left atrial thrombus with nonvalvular atrial fibrillation: a single center based retrospective study in a cohort of 2695 Chinese subjects. BioMed Research International. 2017; 2017:6839589

55. Inohara T, Kimura T, Ueda I, Ikemura N, Tanimoto K, Nishiyama N et al. Effect of compliance to updated AHA/ACC performance and quality measures among patients with atrial fibrillation on outcome (from Japanese multicenter registry). American Journal of Cardiology. 2017; 120(4):595-600

56. Inoue H, Nozawa T, Hirai T, Iwasa A, Okumura K, Lee JD et al. Accumulation of risk factors increases risk of thromboembolic events in patients with nonvalvular atrial fibrillation. Circulation Journal. 2006; 70(6):651-656

57. Jaakkola S, Kiviniemi TO, Nuotio I, Hartikainen J, Mustonen P, Palomaki A et al. Usefulness of the CHADS2-VASc and HAS-BLED Scores in predicting the risk of stroke versus intracranial bleeding in patients with atrial fibrillation (from the FibStroke study). American Journal of Cardiology. 2018; 121(10):1182-1186

58. Joundi RA, Cipriano LE, Sposato LA, Saposnik G. Ischemic stroke risk in patients with atrial fibrillation and CHA2DS2-VASc score of 1: Systematic review and meta-analysis. Stroke. 2016; 47(5):1364-1367

59. Kabra R, Girotra S, Vaughan Sarrazin M. Refining stroke prediction in atrial fibrillation patients by addition of african-american ethnicity to CHA2DS2-VASc score. Journal of the American College of Cardiology. 2016; 68(5):461-470

60. Kang SH, Choi EK, Han KD, Lee SR, Lim WH, Cha MJ et al. Risk of ischemic stroke in patients with non-valvular atrial fibrillation not receiving oral anticoagulants- korean nationwide population-based study. Circulation Journal. 2017; 81(8):1158-1164

61. Kang SH, Kim J, Park JJ, Oh IY, Yoon CH, Kim HJ et al. Risk of stroke in congestive heart failure with and without atrial fibrillation. International Journal of Cardiology. 2017; 248:182-187

62. Karlsson LO, Nilsson S, Bang M, Nilsson L, Charitakis E, Janzon M. A clinical decision support tool for improving adherence to guidelines on anticoagulant therapy in patients with atrial fibrillation at risk of stroke: A cluster-randomized trial in a Swedish primary care setting (the CDS-AF study). PLoS Medicine. 2018; 15(3):e1002528

63. Karlsson LO, Nilsson S, Charitakis E, Bang M, Johansson G, Nilsson L et al. Clinical decision support for stroke prevention in atrial fibrillation (CDS-AF): rationale and

design of a cluster randomized trial in the primary care setting. American Heart Journal. 2017; 187:45-52

64. Kearon C. In AF, ABC scores predicted stroke or major bleeding better than CHA2DS2-VASc and HAS-BLED scores, respectively. Annals of Internal Medicine. 2019; 170(12):JC71

65. Kim MN, Kim SA, Choi JI, Park SM, Park SW, Kim YH et al. Improvement of predictive value for thromboembolic risk by incorporating left atrial functional parameters in the CHADS2 and CHA2DS2-VASc Scores. International Heart Journal. 2015; 56(3):286-292

66. Kim TH, Yang PS, Kim D, Yu HT, Uhm JS, Kim JY et al. CHA2DS2-VASc score for identifying truly low-risk atrial fibrillation for stroke: a Korean nationwide cohort study. Stroke. 2017; 48(11):2984-2990

67. Kim TH, Yang PS, Uhm JS, Kim JY, Pak HN, Lee MH et al. CHA2DS2-VASc Score (congestive heart failure, hypertension, age >=75 [doubled], diabetes mellitus, prior stroke or transient ischemic attack [doubled], vascular disease, age 65-74, female) for stroke in asian patients with atrial fibrillation: a Korean nationwide sample cohort study. Stroke. 2017; 48(6):1524-1530

68. Komatsu T, Sato Y, Ozawa M, Kunugita F, Yoshizawa R, Morino Y et al. Comparison between CHADS2 and CHA2DS2-VASc score for risk stratification of ischemic stroke in Japanese patients with non-valvular paroxysmal atrial fibrillation not receiving anticoagulant therapy. International Heart Journal. 2014; 55(2):119-125

69. Komatsu T, Tachibana H, Satoh Y, Ozawa M, Kunugita F, Ueda H et al. Relationship between CHA(2)DS(2)-VASc scores and ischemic stroke/cardiovascular events in Japanese patients with paroxysmal atrial fibrillation not receiving anticoagulant therapy. Journal of Cardiology. 2012; 59(3):321-328

70. Laguna P, Martin A, Del Arco C, Millan I, Gargantilla P, Spanish Atrial Fibrillation in Emergency Medicine Study Group. Differences among clinical classification schemes for predicting stroke in atrial fibrillation: implications for therapy in daily practice. Academic Emergency Medicine. 2005; 12(9):828-834

71. Larsen TB, Lip GY, Skjoth F, Due KM, Overvad K, Hvilsted Rasmussen L. Added predictive ability of the CHA2DS2VASc risk score for stroke and death in patients with atrial fibrillation: the prospective Danish Diet, Cancer, and Health cohort study. Circulation: Cardiovascular Quality and Outcomes. 2012; 5(3):335-342

72. Larsen TB, Lip GYH, Due KM, Skjoeth F, Tjoenneland A, Overvad K et al. Comparison of the CHADS2 and CHA2DS2VASc prediction rules for stroke in patients with atrial fibrillation. the Danish diet, cancer and health cohort study. European Heart Journal. 2011; 32(Suppl 1):131

73. Lin YS, Chen YL, Chen TH, Lin MS, Liu CH, Yang TY et al. Comparison of Clinical Outcomes Among Patients With Atrial Fibrillation or Atrial Flutter Stratified by CHA2DS2-VASc Score. JAMA Network Open. 2018; 1(4):e180941

74. Lip GY, Lane D, Van Walraven C, Hart RG. Additive role of plasma von Willebrand factor levels to clinical factors for risk stratification of patients with atrial fibrillation. Stroke. 2006; 37(9):2294-2300

75. 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

76. Lip GYH, Haguenoer K, Saint-Etienne C, Fauchier L. Relationship of the SAMe-TT2R2 score to poor-quality anticoagulation, stroke, clinically relevant bleeding, and mortality in patients with atrial fibrillation. Chest. 2014; 146(3):719-726

77. Lip GYH, Jensen M, Melgaard L, Skjoth F, Nielsen PB, Larsen TB. Stroke and bleeding risk scores in patients with atrial fibrillation and valvular heart disease: evaluating 'valvular heart disease' in a nationwide cohort study. Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology. 2019; 21(1):33-40

78. Lip GYH, Lane DA, Buller H, Apostolakis S. Development of a novel composite stroke and bleeding risk score in patients with atrial fibrillation: the AMADEUS Study. Chest. 2013; 144(6):1839-1847

79. Lowres N, Olivier J, Chao TF, Chen SA, Chen Y, Diederichsen A et al. Estimated stroke risk, yield, and number needed to screen for atrial fibrillation detected through single time screening: a multicountry patient-level meta-analysis of 141,220 screened individuals. PLoS Medicine. 2019; 16(9):e1002903

80. Maeda T, Nishi T, Funakoshi S, Tada K, Tsuji M, Satoh A et al. Risks of bleeding and stroke based on CHA2DS2-VASc scores in Japanese patients with atrial fibrillation: a large-scale observational study using real-world data. Journal of the American Heart Association. 2020; 9(5):e014574

81. Maheshwari A, Norby FL, Roetker NS, Soliman EZ, Koene RJ, Rooney MR et al. Refining prediction of atrial fibrillation-related stroke using the P2-CHA2DS2-VASc score: ARIC and MESA. Circulation. 2019; 139(2):180-191

82. Masaki N, Suzuki M, Iwatsuka R, Mizukami A, Kumasaka L, Nagahori W et al. Effectiveness of risk stratification according to CHADS2 score in Japanese patients with nonvalvular atrial fibrillation. International Heart Journal. 2009; 50(3):323-329

83. McAlister FA, Wiebe N, Jun M, Sandhu R, James MT, McMurtry MS et al. Are existing risk scores for nonvalvular atrial fibrillation useful for prediction or risk adjustment in patients with chronic kidney disease? Canadian Journal of Cardiology. 2017; 33(2):243-252

84. McAlister FA, Wiebe N, Ronksley PE, Healey JS. Although non-stroke outcomes are more common, stroke risk scores can be used for prediction in patients with atrial fibrillation. International Journal of Cardiology. 2018; 269:145-151

85. Naccarelli GV, Panaccio MP, Cummins G, Tu N. CHADS2 and CHA2DS2-VASc risk factors to predict first cardiovascular hospitalization among atrial fibrillation/atrial flutter patients. American Journal of Cardiology. 2012; 109(10):1526-1533

86. Nagahara D, Kamiyama N, Fujito T, Mochizuki A, Shimoshige S, Miura T. A novel scoring system for stroke risk stratification in Japanese patients with low CHADS2 scores: Study using a transesophageal-echocardiogram endpoint. Journal of Arrhythmia. 2020; 36(4):624-631

87. Nakagawa K, Hirai T, Takashima S, Fukuda N, Ohara K, Sasahara E et al. Chronic kidney disease and CHADS(2) score independently predict cardiovascular events and mortality in patients with nonvalvular atrial fibrillation. American Journal of Cardiology. 2011; 107(6):912-916

88. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual [Updated October 2018]. London. National Institute for Health and Care Excellence, 2014. Available from: https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview

89. Nielsen PB, Overvad TF. Female sex as a risk modifier for stroke risk in atrial fibrillation: using CHA2DS2-VASc versus CHA2DS2-VA for stroke risk stratification in atrial fibrillation: a note of caution. Thrombosis and Haemostasis. 2020; 120(6):894-898

90. Nielsen PB, Overvad TF, Andersen SD, Larsen TB, Skjoth F, Sogaard M et al. Risk stratification for ischemic cerebrovascular events and mortality among intracerebral hemorrhage patients with and without atrial fibrillation: a nationwide cohort study. Cerebrovascular Diseases. 2019; 48(3-6):236-243

91. Ntaios G, Georgiopoulos G, Koroboki E, Vemmos K. External validation of the PREMISE score in the Athens Stroke Registry. Journal of Stroke and Cerebrovascular Diseases. 2019; 28(7):1806-1809

92. O'Brien EC, Simon DN, Thomas L, Fonarow GC, Kowey PR, Mahaffey KW et al. Comparative performance of the R2CHADS2, CHADS2, and CHA2DS2-VASC scores in atrial fibrillation. Circulation. 2015; 132(Suppl. 3):A17202

93. Oldgren J, Hijazi Z, Lindback J, Alexander J, Connolly S, Eikelboom J et al. External validation of the biomarker-based ABC-stroke risk score for atrial fibrillation. Journal of the American College of Cardiology. 2016; 67(13 Suppl):879

94. Oldgren J, Hijazi Z, Lindback J, Alexander JH, Connolly SJ, Eikelboom JW et al. Performance and validation of a novel biomarker-based stroke risk score for atrial fibrillation. Circulation. 2016; 134(22):1697-1707

95. Olesen JB, Fauchier L, Lane DA, Taillandier S, Lip GYH. Risk factors for stroke and thromboembolism in relation to age among patients with atrial fibrillation: the Loire Valley Atrial Fibrillation Project. Chest. 2012; 141(1):147-153

96. Olesen JB, Lip GY, Hansen ML, Hansen PR, Tolstrup JS, Lindhardsen J et al. Validation of risk stratification schemes for predicting stroke and thromboembolism in patients with atrial fibrillation: nationwide cohort study. BMJ. 2011; 342:D124

97. Olesen JB, Torp-Pedersen C, Hansen ML, Lip GY. The value of the CHA2DS2-VASc score for refining stroke risk stratification in patients with atrial fibrillation with a CHADS2 score 0-1: a nationwide cohort study. Thrombosis and Haemostasis. 2012; 107(6):1172-1179

98. Pandya E, Masood N, Wang Y, Krass I, Bajorek B. Impact of a computerized antithrombotic risk assessment tool on the prescription of thromboprophylaxis in atrial fibrillation: hospital setting. Clinical and Applied Thrombosis/Hemostasis. 2018; 24(1):85-92

99. Parsons C, Cha S, Shen WK, Chamberlain AM, Luis SA, Keddis M et al. Usefulness of the addition of renal function to the CHA2DS2-vasc score as a predictor of thromboembolism and mortality in patients without atrial fibrillation. American Journal of Cardiology. 2018; 122(4):597-603

100. Piccini JP, Stevens SR, Chang Y, Singer DE, Lokhnygina Y, Go AS et al. Renal dysfunction as a predictor of stroke and systemic embolism in patients with nonvalvular atrial fibrillation: validation of the R(2)CHADS(2) index in the ROCKET AF (Rivaroxaban Once-daily, oral, direct factor Xa inhibition compared with vitamin K

antagonism for prevention of stroke and Embolism Trial in Atrial Fibrillation) and ATRIA (AnTicoagulation and Risk factors In Atrial fibrillation) study cohorts. Circulation. 2013; 127(2):224-232

101. Piyaskulkaew C, Singh T, Szpunar S, Saravolatz L, 2nd, Rosman H. CHA(2)DS(2)-VASc versus CHADS(2) for stroke risk assessment in low-risk patients with atrial fibrillation: a pilot study from a single center of the NCDR-PINNACLE registry. Journal of Thrombosis and Thrombolysis. 2014; 37(4):400-403

102. Poli D, Antonucci E, Grifoni E, Abbate R, Gensini GF, Prisco D. Stroke risk in atrial fibrillation patients on warfarin. Predictive ability of risk stratification schemes for primary and secondary prevention. Thrombosis and Haemostasis. 2009; 101(2):367-372

103. Poli D, Antonucci E, Pengo V, Testa S, Palareti G. Comparison of HAS-BLED and HAS-BED versus CHADS2 and CHA2DS2VASC stroke and bleeding scores in patients with atrial fibrillation. American Journal of Cardiology. 2017; 119(7):1012-1016

104. Poli D, Antonucci E, Testa S, Lip GY. A prospective validation of the SAME-TT2R 2 score: how to identify atrial fibrillation patients who will have good anticoagulation control on warfarin. Internal and Emergency Medicine. 2014; 9(4):443-447

105. Poli D, Lip GY, Antonucci E, Grifoni E, Lane D. Stroke risk stratification in a "real-world" elderly anticoagulated atrial fibrillation population. Journal of Cardiovascular Electrophysiology. 2011; 22(1):25-30

106. Potpara TS, Polovina MM, Licina MM, Marinkovic JM, Prostran MS, Lip GY. Reliable identification of "truly low" thromboembolic risk in patients initially diagnosed with "lone" atrial fibrillation: the Belgrade atrial fibrillation study. Circulation: Arrhythmia and Electrophysiology. 2012; 5(2):319-326

107. Potpara TS, Stankovic GR, Beleslin BD, Polovina MM, Marinkovic JM, Ostojic MC et al. A 12-year follow-up study of patients with newly diagnosed lone atrial fibrillation: implications of arrhythmia progression on prognosis: the Belgrade Atrial Fibrillation study. Chest. 2012; 141(2):339-347

108. Proietti M, Farcomeni A, Romiti GF, Di Rocco A, Placentino F, Diemberger I et al. Association between clinical risk scores and mortality in atrial fibrillation: systematic review and network meta-regression of 669,000 patients. European Journal of Preventive Cardiology. 2018; 27(6): 633-644

109. Puurunen MK, Kiviniemi T, Schlitt A, Rubboli A, Dietrich B, Karjalainen P et al. CHADS2, CHA2DS2-VASc and HAS-BLED as predictors of outcome in patients with atrial fibrillation undergoing percutaneous coronary intervention. Thrombosis Research. 2014; 133(4):560-566

110. Reps JM, Williams RD, You SC, Falconer T, Minty E, Callahan A et al. Feasibility and evaluation of a large-scale external validation approach for patient-level prediction in an international data network: validation of models predicting stroke in female patients newly diagnosed with atrial fibrillation. BMC Medical Research Methodology. 2020; 20(1):102

111. Rietbrock S, Heeley E, Plumb J, van Staa T. Chronic atrial fibrillation: incidence, prevalence, and prediction of stroke using the Congestive Heart Failure, Hypertension, Age >75, Diabetes Mellitus, and Prior Stroke or Transient Ischemic

Attack (CHADS2) risk stratification scheme. American Heart Journal. 2008; 156(1):57-64

112. Rivera-Caravaca JM, Marin F, Esteve-Pastor MA, Rana-Miguez P, Anguita M, Muniz J et al. Usefulness of the 2MACE score to predicts adverse cardiovascular events in patients with atrial fibrillation. American Journal of Cardiology. 2017; 120(12):2176-2181

113. Rivera-Caravaca JM, Roldan V, Esteve-Pastor MA, Valdes M, Vicente V, Lip GYH et al. Importance of time in therapeutic range on bleeding risk prediction using clinical risk scores in patients with atrial fibrillation. Scientific Reports. 2017; 7(1):12066

114. Rivera-Caravaca JM, Roldan V, Esteve-Pastor MA, Valdes M, Vicente V, Lip GYH et al. Long-term stroke risk prediction in 'real world' atrial fibrillation patients: a comparison of the ABC-stroke and CHA2DS2-VASc scores. Research and Practice in Thrombosis and Haemostasis. 2017; 1 (Suppl 1):335-336

115. Rivera-Caravaca JM, Roldan V, Esteve-Pastor MA, Valdes M, Vicente V, Marin F et al. Prediction of long-term net clinical outcomes using the TIMI-AF score: comparison with CHA2DS2-VASc and HAS-BLED. American Heart Journal. 2018; 197:27-34

116. Roldan V, Rivera-Caravaca JM, Shantsila A, Garcia-Fernandez A, Esteve-Pastor MA, Vilchez JA et al. Enhancing the 'real world' prediction of cardiovascular events and major bleeding with the CHA2DS2-VASc and HAS-BLED scores using multiple biomarkers. Annals of Medicine. 2018; 50(1):26-34

117. Ruff CT, Giugliano RP, Braunwald E, Murphy SA, Brown K, Jarolim P et al. Cardiovascular biomarker score and clinical outcomes in patients with atrial fibrillation: a subanalysis of the ENGAGE AF-TIMI 48 randomized clinical trial. JAMA Cardiology. 2016; 1(9):999-1006

118. Ruiz Ortiz M, Romo E, Mesa D, Delgado M, Anguita M, Castillo JC et al. Oral anticoagulation in nonvalvular atrial fibrillation in clinical practice: impact of CHADS(2) score on outcome. Cardiology. 2010; 115(3):200-204

119. Saito Y, Okumura Y, Nagashima K, Fukamachi D, Yokoyama K, Matsumoto N et al. Impact of the Fibrosis-4 Index on risk stratification of cardiovascular events and mortality in patients with atrial fibrillation: findings from a Japanese multicenter registry. Journal of Clinical Medicine. 2020; 9(2):21

120. Sander Van Doorn S, Debray TPA, Kaasenbrood F, Hoes AW, Rutten FH, Moons KGM et al. Predictive performance of the CHA2DS2-VASc rule in atrial fibrillation: a systematic review and meta-analysis. European Journal of Preventive Cardiology. 2018; 2(2 Suppl 1):S72

121. Schwartz S, Tedla Y, Greenland P, Yadlapati A, Passman R. Evaluation of Cha2ds2-Vasc and Has-Bled Discriminative Ability in Whites and Non-Whites. Journal of the American College of Cardiology. 2019; 73(9 Suppl 1):449

122. Schwartz SM, Tedla YG, Greenland P, Yadlapati A, Passman RS. Discriminative ability of CHA2DS2-VASc and HAS-BLED score in whites and nonwhites. American Journal of Cardiology. 2019; 123(12):1949-1954

123. Singer DE, Chang Y, Borowsky LH, Fang MC, Pomernacki NK, Udaltsova N et al. A new risk scheme to predict ischemic stroke and other thromboembolism in atrial fibrillation: The ATRIA study stroke risk score. Journal of the American Heart Association. 2013; 2(3):e000250

124. Siu CW, Lip GY, Lam KF, Tse HF. Risk of stroke and intracranial hemorrhage in 9727 Chinese with atrial fibrillation in Hong Kong. Heart Rhythm. 2014; 11(8):1401-1408

125. Somme D, Corvol A, Lazarovici C, Lahjibi-Paulet H, Gisselbrecht M, Saint-Jean O. Clinical usefulness in geriatric patients of combining CHADS2 and HEMORR2HAGES scores to guide antithrombotic prophylaxis in atrial fibrillation. Aging-Clinical and Experimental Research. 2010; 22(4):289-294

126. Sun Y, Zhu J, Ma C, Liu S, Yang Y, Hu D. Stroke risk status, anticoagulation treatment, and quality-of-life in chinese patients with atrial fibrillation: China Registry of Atrial Fibrillation (CRAF). Cardiovascular Therapeutics. 2019; 2019:Article ID 7372129

127. Suzuki S, Yamashita T, Okumura K, Atarashi H, Akao M, Ogawa H et al. Incidence of ischemic stroke in Japanese patients with atrial fibrillation not receiving anticoagulation therapy--pooled analysis of the Shinken Database, J-RHYTHM Registry, and Fushimi AF Registry. Circulation Journal. 2015; 79(2):432-438

128. Tanaka K, Yamada T, Torii T, Furuta K, Matsumoto S, Yoshimura T et al. Pre-admission CHADS2, CHA2DS2-VASc, and R2CHADS2 scores on severity and functional outcome in acute ischemic stroke with atrial fibrillation. Journal of Stroke and Cerebrovascular Diseases. 2015; 24(7):1629-1635

129. Tanaka S, Hirai T, Inao K, Fukuda N, Nakagawa K, Inoue H et al. High cardiac troponin I Is associated with transesophageal echocardiographic risk of thromboembolism and ischemic stroke events in non-valvular atrial fibrillation patients. Circulation Journal. 2018; 82(6):1699-1704

130. Tomasdottir M, Friberg L, Hijazi Z, Lindback J, Oldgren J. Risk of ischemic stroke and utility of CHA2 DS2 -VASc score in women and men with atrial fibrillation. Clinical Cardiology. 2019; 06:06

131. Tomita H, Okumura K, Inoue H, Atarashi H, Yamashita T, Origasa H et al. Validation of risk scoring system excluding female sex from CHA2DS2-VASc in Japanese patients with nonvalvular atrial fibrillation:Subanalysis of the J-RHYTHM registry. Circulation Journal. 2015; 79(8):1719-1726

132. Tsai CT, Chang SH, Chang SN, Hwang JJ, Wu CK, Wang YC et al. Additive effect of the metabolic syndrome score to the conventional CHADS2 score for the thromboembolic risk stratification of patients with atrial fibrillation. Heart Rhythm. 2014; 11(3):352-357

133. Van Den Ham HA, Klungel OH, Singer DE, Leufkens HG, van Staa TP. Comparative performance of ATRIA, CHADS2, and CHA2DS2-VASc risk scores predicting stroke in patients with atrial fibrillation: results from a national primary care database. Journal of the American College of Cardiology. 2015; 66(17):1851-1859

134. Van Den Ham HA, Klungel OH, Singer DE, Leufkens HGM, Van Staa TP. Comparison of atria and CHA2DS2-vasc risk stratification schemes for the prediction of stroke in the individual patient with atrial fibrillation and the impact on treatment decisions. Pharmacoepidemiology and Drug Safety. 2014; 1:359-360

135. Van Mieghem W, Lancellotti P. CHADS2 risk score and rate of stroke or systemic embolism and major bleeding in patients with non-valvular atrial fibrillation receiving non-vitamin K antagonist oral anticoagulants. Acta Cardiologica. 2017; 72(4):390-396

136. Van Staa TP, Setakis E, Di Tanna GL, Lane DA, Lip GYH. A comparison of risk stratification schemes for stroke in 79 884 atrial fibrillation patients in general practice. Journal of Thrombosis and Haemostasis. 2011; 9(1):39-48

137. Wang TJ, Massaro JM, Levy D, Vasan RS, Wolf PA, D'Agostino RB et al. A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: the Framingham Heart Study. JAMA. 2003; 290(8):1049-1056

138. Wicke FS, Schaller MA, Karymova K, Beyer M, Muller BS. Ischemic stroke risk estimation in patients without oral anticoagulation: an observational cohort study based on secondary data from Germany. BMC Cardiovascular Disorders. 2019; 19(1):94

139. Xing Y, Ma Q, Ma X, Wang C, Zhang D, Sun Y. CHADS2 score has a better predictive value than CHA2DS2-VASc score in elderly patients with atrial fibrillation. Clinical Interventions in Aging. 2016; 11:941-946

140. Xing Y, Sun Y, Li H, Tang M, Huang W, Zhang K et al. CHADS2-VASc score as a predictor of long-term cardiac outcomes in elderly patients with or without atrial fibrillation. Clinical Interventions in Aging. 2018; 13:497-504

141. Yang LT, Tsai WC, Su HM. Echocardiographic parameters versus CHA2DS2-VASc score in prediction of overall cardiac events, heart failure, and stroke in non-valvular atrial fibrillation. Cardiology Journal. 2018; 25(1):60-71

142. Yoshizawa R, Komatsu T, Kunugita F, Ozawa M, Ohwada S, Satoh Y et al. Comparison of the CHADS2, CHA2DS2-VASC and R2CHADS2 scores in Japanese patients with non-valvular paroxysmal atrial fibrillation not receiving anticoagulation therapy. Internal Medicine. 2017; 56(21):2827-2836

143. Zhu W, Fu L, Ding Y, Huang L, Xu Z, Hu J et al. Meta-analysis of ATRIA versus CHA2DS2-VASc for predicting stroke and thromboembolism in patients with atrial fibrillation. International Journal of Cardiology. 2017; 227:436-442

144. Zhu W, Wu Y, Zhou Y, Liang W, Xue R, Wu Z et al. CHA2DS2-VASc and ATRIA scores and clinical outcomes in patients with heart failure with preserved ejection fraction. Cardiovascular Drugs and Therapy. 2020; 34:763–772

Appendices

Appendix A: Review protocols

Table 11: Review protocol: What is the most clinically and cost-effective risk stratification tool for predicting stroke or thromboembolic events in people with atrial fibrillation?

ID	Field	Content
0.	PROSPERO registration number	[Complete this section with the PROSPERO registration number once allocated]
1.	Review title	Clinical and cost-effectiveness of tools for assessing stroke risk in people with atrial fibrillation
2.	Review question	What is the most clinically and cost-effective risk stratification tool for predicting stroke or thromboembolic events in people with atrial fibrillation?
3.	Objective	To identify the most clinically and cost effective tool to measure the risk of stroke and thromboembolic complications in this population
4.	Searches	The following databases will be searched: Cochrane Central Register of Controlled Trials (CENTRAL) Cochrane Database of Systematic Reviews (CDSR) Embase MEDLINE Searches will be restricted by: English language Human studies Letters and comments are excluded. Other searches: Inclusion lists of relevant systematic reviews will be checked by the reviewer. The searches may be re-run 6 weeks before final submission of the review and further studies retrieved for inclusion if relevant. The full search strategies for MEDLINE database will be published in the final review.
5.	Condition or domain being studied	Atrial Fibrillation
6.	Population	Inclusion: People aged over 18 with AF. Exclusion: People with AF due to severe valvular disease
7.	Intervention/Exposure/Test	Any stroke risk tool (for example, ABC stroke score, Q stroke, ATRIA, Troponin 2, CHADS2). Any version of CHADS2VASC with modifications

ID	Field	Content
		[Note: treat each test using a different threshold as a separate intervention; for example, Q stroke using the threshold of X for 'need for anticoagulation' is treated as a separate intervention to Q stroke using the threshold of Y for 'need for anticoagulation'].
8.	Comparator/Reference standard/Confounding factors	CHADS2VASC (the established method, as recommended by previous version of this guideline)
9.	Types of study to be included	Systematic reviews RCTs (including those with a cross-over design). Non-randomised studies will be excluded.
10.	Other exclusion criteria	Non-English language studies. Abstracts will be excluded as it is expected there will be sufficient full text published studies available.
11.	Context	N/A
12.	Primary outcomes (critical outcomes)	health-related quality of life mortality stroke or thromboembolic complications major bleeding Longest follow up point always used
13.	Secondary outcomes (important outcomes)	None
14.	Data extraction (selection and coding)	EndNote will be used for reference management, sifting, citations and bibliographies. Titles and/or abstracts of studies retrieved using the search strategy and those from additional sources will be screened for inclusion. The full text of potentially eligible studies will be retrieved and will be assessed for eligibility in line with the criteria outlined above. 10% of the abstracts will be reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. An in-house developed database; EviBase, will be used for data extraction. A standardised form is followed to extract data from studies (see Developing NICE guidelines: the manual section 6.4) and for undertaking assessment of study quality. Summary evidence tables will be produced including information on: study setting; study population and participant demographics and baseline characteristics; details of the intervention and control interventions; study methodology' recruitment and missing data rates; outcomes and times of measurement; critical appraisal ratings. A second reviewer will quality assure the extracted data. Discrepancies will be identified and resolved through discussion (with a third reviewer where necessary).

ID	Field	Content
15.	Risk of bias (quality) assessment	Risk of bias will be assessed using the appropriate checklist as described in Developing NICE guidelines: the manual. For Intervention reviews the following checklist will be used according to study design being assessed: Systematic reviews: Risk of Bias in Systematic Reviews (ROBIS) Randomised Controlled Trial: Cochrane RoB (2.0) Disagreements between the review authors over the risk of bias in particular studies will be resolved by discussion, with involvement of a third review author where necessary.
16.	Strategy for data synthesis	Where possible, data will be meta-analysed. Pairwise meta-analyses will be performed using Cochrane Review Manager (RevMan5) to combine the data given in all studies for each of the outcomes stated above. A fixed effect meta-analysis, with weighted mean differences for continuous outcomes and risk ratios for binary outcomes will be used, and 95% confidence intervals will be calculated for each outcome. Heterogeneity between the studies in effect measures will be assessed using the I^2 statistic and visually inspected. We will consider an I^2 value greater than 50% indicative of substantial heterogeneity. Sensitivity analyses will be conducted based on pre-specified subgroups using stratified meta-analysis to explore the heterogeneity in effect estimates. If this does not explain the heterogeneity, the results will be presented using random-effects. GRADE pro will be used to assess the quality of each outcome, taking into account individual study quality and the meta-analysis results. The 4 main quality elements (risk of bias, indirectness, inconsistency and imprecision) will be appraised for each outcome. Publication bias is tested for when there are more than 5 studies for an outcome. Other bias will only be taken into consideration in the quality assessment if it is apparent. Where meta-analysis is not possible, data will be presented and quality assessed individually per outcome. If sufficient data is available to make a network of treatments, WinBUGS will be used for network meta-analysis.
17.	Analysis of sub-groups	Stratification None Sub-grouping If serious or very serious heterogeneity (I2>50%) is present within any stratum, sub-grouping will occur according to the following strategies: None
18.	Type and method of review	☐ Intervention ☐ Diagnostic ☐ Prognostic ☐ Qualitative ☐ Epidemiologic

ID	Field	Content		
		☐	Service Delivery	
		☒	Other (please specify): RCT review of prediction tools	
19.	Language	English		
20.	Country	England		
21.	Anticipated or actual start date			
22.	Anticipated completion date			
23.	Stage of review at time of this submission	Review stage	Started	Completed
		Preliminary searches	☐	☑
		Piloting of the study selection process	☐	☑
		Formal screening of search results against eligibility criteria	☐	☑
		Data extraction	☐	☑
		Risk of bias (quality) assessment	☐	☑
		Data analysis	☐	☑
24.	Named contact	5a. Named contact National Guideline Centre 5b Named contact e-mail 5e Organisational affiliation of the review National Institute for Health and Care Excellence (NICE) and the National Guideline Centre		
25.	Review team members	From the National Guideline Centre: Sharon Swain Mark Perry Nicole Downes Sophia Kemmis Betty Elizabeth Pearton		
26.	Funding sources/sponsor	This systematic review is being completed by the National Guideline Centre which receives funding from NICE.		

ID	Field	Content	
27.	Conflicts of interest	All guideline committee members and anyone who has direct input into NICE guidelines (including the evidence review team and expert witnesses) must declare any potential conflicts of interest in line with NICE's code of practice for declaring and dealing with conflicts of interest. Any relevant interests, or changes to interests, will also be declared publicly at the start of each guideline committee meeting. Before each meeting, any potential conflicts of interest will be considered by the guideline committee Chair and a senior member of the development team. Any decisions to exclude a person from all or part of a meeting will be documented. Any changes to a member's declaration of interests will be recorded in the minutes of the meeting. Declarations of interests will be published with the final guideline.	
28.	Collaborators	Development of this systematic review will be overseen by an advisory committee who will use the review to inform the development of evidence-based recommendations in line with section 3 of Developing NICE guidelines: the manual. Members of the guideline committee are available on the NICE website: [NICE guideline webpage].	
29.	Other registration details		
30.	Reference/URL for published protocol		
31.	Dissemination plans	NICE may use a range of different methods to raise awareness of the guideline. These include standard approaches such as: notifying registered stakeholders of publication publicising the guideline through NICE's newsletter and alerts issuing a press release or briefing as appropriate, posting news articles on the NICE website, using social media channels, and publicising the guideline within NICE.	
32.	Keywords	Atrial Fibrillation, stroke prediction tools	
33.	Details of existing review of same topic by same authors	N/A	
34.	Current review status	☒	Ongoing
		☐	Completed but not published
		☐	Completed and published
		☐	Completed, published and being updated
		☐	Discontinued
35..	Additional information	N/A	
36.	Details of final publication	www.nice.org.uk	

Table 12: Review protocol: What is the most accurate risk stratification tool for predicting stroke or thromboembolic events in people with atrial fibrillation?

ID	Field	Content
0.	PROSPERO registration number	Not registered

ID	Field	Content
1.	Review title	Accuracy of risk stratification tools for predicting stroke or thromboembolic events in people with atrial fibrillation.
2.	Review question	What is the most accurate risk stratification tool for predicting stroke or thromboembolic events in people with atrial fibrillation?
3.	Objective	To identify the most accurate tool to measure the risk of stroke or any thromboembolic event in this population.
4.	Searches	The following databases will be searched: Cochrane Central Register of Controlled Trials (CENTRAL) Cochrane Database of Systematic Reviews (CDSR) Embase MEDLINE Searches will be restricted by: English language Other searches: None The searches may be re-run 6 weeks before final submission of the review and further studies retrieved for inclusion if relevant. The full search strategies for MEDLINE database will be published in the final review.
5.	Condition or domain being studied	Atrial Fibrillation
6.	Population	People aged over 18 with a diagnosis of AF who are not being anticoagulated. Exclusion: People who are already being anticoagulated
7.	Index Test	Any stroke risk tool (e.g ABC stroke score, Q stroke, ATRIA, Troponin 2, CHADSVASC) Any other version of CHADSVASC with modifications
8.	Comparator/Reference standard/Confounding factors	Later stroke or thromboembolic event
9.	Types of study to be included	Prognostic prediction tool evaluation studies.
10.	Other exclusion criteria	Non-English language studies.
11.	Context	N/A
12.	Primary outcomes (critical outcomes)	Simple diagnostic (prognostic) accuracy outcomes, such as sensitivity and specificity C statistic (based on sensitivity and specificity but useful if >1 threshold used). Calibration outcomes Reclassification – scored from -2 (worst) to +2 (best), and based on the degree of correct (+1 for each) and incorrect (-1 for each) up-

ID	Field	Content		
		classifications and down-classifications of one test relative to another test, using the outcome of stroke or thromboembolic events as reference.		
13.	Secondary outcomes (important outcomes)	None		
14.	Data extraction (selection and coding)	EndNote will be used for reference management, sifting, citations and bibliographies. All references identified by the searches and from other sources will be screened for inclusion. 10% of the abstracts will be reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. The full text of these potentially eligible studies will be retrieved and assessed in line with the criteria outlined above. A standardised form will be used to extract data from the included studies (see Developing NICE guidelines: the manual section 6.4). Data extraction will be independently quality assured by a second reviewer, discrepancies will be identified and resolved through discussion (with a third party where necessary).		
15.	Risk of bias (quality) assessment	Risk of bias quality assessment will be assessed using PROBAST. Assessment will be independently quality assured by a second reviewer. Disagreements between the reviewers will be resolved by discussion, with involvement of a third party where necessary.		
16.	Strategy for data synthesis	Where possible C statistic and NRI data will be meta-analysed where appropriate (if at least 3 studies reporting data at the same diagnostic threshold) in RevMan. Summary diagnostic outcomes will be reported from the meta-analyses with their 95% confidence intervals in adapted GRADE tables. Heterogeneity will be assessed using I2 thresholds. If meta-analysis is not possible, data will be presented as individual values in adapted GRADE profile tables.		
17.	Analysis of sub-groups	If heterogeneity is identified, where data is available, subgroup analysis will be carried out for the following subgroups: None		
18.	Type and method of review	☐	Intervention	
		☐	Diagnostic	
		☒	Prognostic	
		☐	Qualitative	
		☐	Epidemiologic	
		☐	Service Delivery	
		☐	Other (please specify)	
19.	Language	English		
20.	Country	England		
21.	Anticipated or actual start date			
22.	Anticipated completion date			
23.		Review stage	Started	Completed

ID	Field	Content		
	Stage of review at time of this submission	Preliminary searches	☐	☑
		Piloting of the study selection process	☐	☑
		Formal screening of search results against eligibility criteria	☐	☑
		Data extraction	☐	☑
		Risk of bias (quality) assessment	☐	☑
		Data analysis	☐	☑
24.	Named contact	5a. Named contact National Guideline Centre 5b Named contact e-mail 5e Organisational affiliation of the review National Institute for Health and Care Excellence (NICE) and the National Guideline Centre		
25.	Review team members	From the National Guideline Centre: Sharon Swain Mark Perry Nicole Downes Sophia Kemmis Betty Elizabeth Pearton		
26.	Funding sources/sponsor	This systematic review is being completed by the National Guideline Centre which receives funding from NICE.		
27.	Conflicts of interest	All guideline committee members and anyone who has direct input into NICE guidelines (including the evidence review team and expert witnesses) must declare any potential conflicts of interest in line with NICE's code of practice for declaring and dealing with conflicts of interest. Any relevant interests, or changes to interests, will also be declared publicly at the start of each guideline committee meeting. Before each meeting, any potential conflicts of interest will be considered by the guideline committee Chair and a senior member of the development team. Any decisions to exclude a person from all or part of a meeting will be documented. Any changes to a member's declaration of interests will be recorded in the minutes of the meeting. Declarations of interests will be published with the final guideline.		

ID	Field	Content	
28.	Collaborators	Development of this systematic review will be overseen by an advisory committee who will use the review to inform the development of evidence-based recommendations in line with section 3 of Developing NICE guidelines: the manual. Members of the guideline committee are available on the NICE website: [NICE guideline webpage].	
29.	Other registration details	N/A	
30.	Reference/URL for published protocol		
31.	Dissemination plans	NICE may use a range of different methods to raise awareness of the guideline. These include standard approaches such as: notifying registered stakeholders of publication publicising the guideline through NICE's newsletter and alerts issuing a press release or briefing as appropriate, posting news articles on the NICE website, using social media channels, and publicising the guideline within NICE. [Add in any additional agree dissemination plans.]	
32.	Keywords	Diagnosis, Atrial Fibrillation	
33.	Details of existing review of same topic by same authors	N/A	
34.	Current review status	☐	Ongoing
		☒	Completed but not published
		☐	Completed and published
		☐	Completed, published and being updated
		☐	Discontinued
35..	Additional information	N/A	
36.	Details of final publication	www.nice.org.uk	

Table 13: Health economic review protocol

Review question	All questions – health economic evidence
Objectives	To identify health economic studies relevant to any of the review questions.
Search criteria	• Populations, interventions and comparators must be as specified in the clinical review protocol above. • Studies must be of a relevant health economic study design (cost–utility analysis, cost-effectiveness analysis, cost–benefit analysis, cost–consequences analysis, comparative cost analysis). • Studies must not be a letter, editorial or commentary, or a review of health economic evaluations. (Recent reviews will be ordered although not reviewed. The bibliographies will be checked for relevant studies, which will then be ordered.) • Unpublished reports will not be considered unless submitted as part of a call for evidence. • Studies must be in English.
Search strategy	A health economic study search will be undertaken using population-specific terms and a health economic study filter – see appendix B below. For questions being updated from NICE guideline CG180, the search will be run from October 2013, which was the cut-off date for the searches. For questions being updated from the NICE guideline CG36 and for new questions, the search will be run from 2003.
Review strategy	Studies not meeting any of the search criteria above will be excluded. Studies published before 2003, abstract-only studies and studies from non-OECD countries or the USA will also be excluded. Studies published after 2003 that were included in the previous guideline(s) will be reassessed for inclusion and may be included or selectively excluded based on their relevance to the questions covered in this update and whether more applicable evidence is also identified. Each remaining study will be assessed for applicability and methodological limitations using the NICE economic evaluation checklist which can be found in appendix H of Developing NICE guidelines: the manual.[88] **Inclusion and exclusion criteria** • If a study is rated as both 'Directly applicable' and with 'Minor limitations' then it will be included in the guideline. A health economic evidence table will be completed and it will be included in the health economic evidence profile. • If a study is rated as either 'Not applicable' or with 'Very serious limitations' then it will usually be excluded from the guideline. If it is excluded then a health economic evidence table will not be completed and it will not be included in the health economic evidence profile. • If a study is rated as 'Partially applicable', with 'Potentially serious limitations' or both then there is discretion over whether it should be included. **Where there is discretion** The health economist will make a decision based on the relative applicability and quality of the available evidence for that question, in discussion with the guideline committee if required. The ultimate aim is to include health economic studies that are helpful for decision-making in the context of the guideline and the current NHS setting. If several studies are considered of sufficiently high applicability and methodological quality that they could all be included, then the health economist, in discussion with the committee if required, may decide to include only the most applicable studies and to selectively exclude the remaining studies. All studies excluded on the basis of applicability or methodological limitations will be listed with explanation in the excluded health economic studies appendix below.

The health economist will be guided by the following hierarchies.

Setting:

- UK NHS (most applicable).
- OECD countries with predominantly public health insurance systems (for example, France, Germany, Sweden).
- OECD countries with predominantly private health insurance systems (for example, Switzerland).
- Studies set in non-OECD countries or in the USA will be excluded before being assessed for applicability and methodological limitations.

Health economic study type:

- Cost–utility analysis (most applicable).
- Other type of full economic evaluation (cost–benefit analysis, cost-effectiveness analysis, cost–consequences analysis).
- Comparative cost analysis.
- Non-comparative cost analyses including cost-of-illness studies will be excluded before being assessed for applicability and methodological limitations.

Year of analysis:

- The more recent the study, the more applicable it will be.
- Studies published in 2003 or later (including any such studies included in the previous guideline(s)) but that depend on unit costs and resource data entirely or predominantly from before 2003 will be rated as 'Not applicable'.
- Studies published before 2003 (including any such studies included in the previous guideline(s))will be excluded before being assessed for applicability and methodological limitations.

Quality and relevance of effectiveness data used in the health economic analysis:

- The more closely the clinical effectiveness data used in the health economic analysis match with the outcomes of the studies included in the clinical review the more useful the analysis will be for decision-making in the guideline.

Appendix B: Literature search strategies

This literature search strategy was used for the following reviews:

- **What is the most clinically and cost-effective risk stratification tool for predicting stroke or thromboembolic events in people with atrial fibrillation?**

- **What is the most accurate risk stratification tool for predicting stroke or thromboembolic events in people with atrial fibrillation?**

The literature searches for this review are detailed below and complied with the methodology outlined in Developing NICE guidelines: the manual.[88]

For more information, please see the Methods Report published as part of the accompanying documents for this guideline.

B.1 Clinical search literature search strategy

Searches were constructed using a PICO framework where population (P) terms were combined with Intervention (I) and in some cases Comparison (C) terms. Outcomes (O) are rarely used in search strategies for interventions as these concepts may not be well described in title, abstract or indexes and therefore difficult to retrieve. Searches were constructed using the following approaches:

- Population AND Prognostic/risk factor terms AND Study filter(s)

Table 14: Database date parameters and filters used

Database	Dates searched	Search filter used
Medline (OVID)	1946 – 10 September 2020	Exclusions Randomised controlled trials Systematic review studies Observational studies Diagnostic tests studies Risk/Prognostic studies
Embase (OVID)	1974 – 10 September 2020	Exclusions Randomised controlled trials Systematic review studies Observational studies Diagnostic tests studies Risk/Prognostic studies
The Cochrane Library (Wiley)	Cochrane Reviews to 2020 Issue 9 of 12 CENTRAL to 2020 Issue 9 of 12	None

Medline (Ovid) search terms

1.	exp atrial fibrillation/
2.	((atrial or atria or atrium or auricular) adj3 fibrillat*).ti,ab.
3.	AF.ti,ab.

4.	1 or 2 or 3
5.	letter/
6.	editorial/
7.	news/
8.	exp historical article/
9.	Anecdotes as Topic/
10.	comment/
11.	case report/
12.	(letter or comment*).ti.
13.	or/5-12
14.	randomized controlled trial/ or random*.ti,ab.
15.	13 not 14
16.	animals/ not humans/
17.	exp Animals, Laboratory/
18.	exp Animal Experimentation/
19.	exp Models, Animal/
20.	exp Rodentia/
21.	(rat or rats or mouse or mice).ti.
22.	or/15-21
23.	4 not 22
24.	limit 23 to English language
25.	(stroke or strokes).ti,ab.
26.	((cerebro* or cerebral*) adj2 (accident* or apoplexy)).ti,ab.
27.	(CVA or poststroke or poststrokes).ti,ab.
28.	exp Intracranial Hemorrhages/
29.	(brain adj2 (attack* or hemorrhag* or haemorrhag* or bleed* or infarct*)).ti,ab.
30.	((intracerebral or intracranial or cerebral* or cerebro* or cerebrum or cerebellum or subarachnoid* or choroidal or basal ganglia or subdural) adj3 (hemorrhag* or haemorrhag* or bleed*)).ti,ab.
31.	exp Brain infarction/
32.	*Thromboembolism/
33.	exp "Intracranial Embolism and Thrombosis"/
34.	exp Carotid Artery Thrombosis/
35.	((brain or brainstem or cerebr* or cerebell* or vertebrobasil* or verte brobasil* or hemisphere* or intracran* or intracerebral or infratentorial or supratentorial or mca* or anterior circulation or carotid or transient or lacunar) adj3 (infarct* or thrombo* or emboli* or occlus* or hypoxi*)).ti,ab.
36.	((thrombo* or emboli*) adj2 event*).ti,ab.
37.	troponin*.ti,ab.
38.	or/25-37
39.	38 not 22
40.	limit 39 to English language
41.	exp risk/
42.	(risk adj3 (assess* or scheme* or rating* or tool* or rule* or index* or indices or score* or scoring or scale* or model* or system* or algorithm* or stratif* or criteria or calculat*)).ti,ab.

43.	Decision Support Systems, Clinical/ or Decision Support Techniques/
44.	((decision or assess* or screen*) adj3 (tool* or rule* or instrument* or index* or test* or technique* or analy* or system* or model*)).ti,ab.
45.	(logistic* adj model*).mp.
46.	exp Prognosis/
47.	exp "Predictive Value of Tests"/
48.	(prognos* or predict*).ti,ab.
49.	or/41-48
50.	40 and 49
51.	chads*.ti,ab.
52.	cha2ds2*.ti,ab.
53.	"cha(2)ds(2)-vasc".ti,ab.
54.	("Anticoagulation and Risk Factors in Atrial Fibrillation" or ATRIA).ti,ab.
55.	Q stroke.ti,ab.
56.	ABC Stroke.ti,ab.
57.	or/50-56
58.	24 and 57
59.	randomized controlled trial.pt.
60.	controlled clinical trial.pt.
61.	randomi#ed.ab.
62.	placebo.ab.
63.	randomly.ab.
64.	clinical trials as topic.sh.
65.	trial.ti.
66.	or/59-65
67.	Meta-Analysis/
68.	Meta-Analysis as Topic/
69.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
70.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
71.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
72.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
73.	(search* adj4 literature).ab.
74.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
75.	cochrane.jw.
76.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
77.	or/67-76
78.	exp "sensitivity and specificity"/
79.	(sensitivity or specificity).ti,ab.
80.	((pre test or pretest or post test) adj probability).ti,ab.
81.	(predictive value* or PPV or NPV).ti,ab.
82.	likelihood ratio*.ti,ab.
83.	likelihood function/

84.	((area under adj4 curve) or AUC).ti,ab.
85.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
86.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness)).ti,ab.
87.	gold standard.ab.
88.	or/78-87
89.	Epidemiologic studies/
90.	Observational study/
91.	exp Cohort studies/
92.	(cohort adj (study or studies or analys* or data)).ti,ab.
93.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
94.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
95.	Controlled Before-After Studies/
96.	Historically Controlled Study/
97.	Interrupted Time Series Analysis/
98.	(before adj2 after adj2 (study or studies or data)).ti,ab.
99.	exp case control study/
100.	case control*.ti,ab.
101.	Cross-sectional studies/
102.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
103.	or/89-102
104.	58 and (66 or 77 or 88 or 103)

Embase (Ovid) search terms

1.	exp Atrial Fibrillation/
2.	((atrial or atria or atrium or auricular) adj3 fibrillat*).ti,ab.
3.	AF.ti,ab.
4.	1 or 2 or 3
5.	letter.pt. or letter/
6.	note.pt.
7.	editorial.pt.
8.	case report/ or case study/
9.	(letter or comment*).ti.
10.	or/5-9
11.	randomized controlled trial/ or random*.ti,ab.
12.	10 not 11
13.	animal/ not human/
14.	nonhuman/
15.	exp Animal Experiment/
16.	exp Experimental Animal/
17.	animal model/
18.	exp Rodent/
19.	(rat or rats or mouse or mice).ti.

20.	or/12-19
21.	4 not 20
22.	limit 21 to English language
23.	(stroke or strokes).ti,ab.
24.	((cerebro* or cerebral*) adj2 (accident* or apoplexy)).ti,ab.
25.	(CVA or poststroke or poststrokes).ti,ab.
26.	*brain hemorrhage/ or *brain ventricle hemorrhage/ or *cerebellum hemorrhage/ or *subarachnoid hemorrhage/
27.	(brain adj2 (attack* or hemorrhag* or haemorrhag* or bleed* or infarct*)).ti,ab.
28.	((intracerebral or intracranial or cerebral* or cerebro* or cerebrum or cerebellum or subarachnoid* or choroidal or basal ganglia or subdural) adj3 (hemorrhag* or haemorrhag* or bleed*)).ti,ab.
29.	*brain infarction/ or *brain infarction size/ or *brain stem infarction/ or *cerebellum infarction/
30.	*thromboembolism/
31.	*brain embolism/
32.	*Carotid Artery Thrombosis/
33.	((brain or brainstem or cerebr* or cerebell* or vertebrobasil* or verte brobasil* or hemisphere* or intracran* or intracerebral or infratentorial or supratentorial or mca* or anterior circulation or carotid or transient or lacunar) adj3 (infarct* or thrombo* or emboli* or occlus* or hypoxi*)).ti,ab.
34.	((thrombo* or emboli*) adj2 event*).ti,ab.
35.	troponin*.ti,ab.
36.	or/23-35
37.	36 not 20
38.	limit 37 to English language
39.	risk/ or risk factor/ or risk assessment/
40.	(risk adj3 (assess* or scheme* or rating* or tool* or rule* or index* or indices or score* or scoring or scale* or model* or system* or algorithm* or stratif* or criteria or calculat*)).ti,ab.
41.	decision support system/
42.	((decision or assess* or screen*) adj3 (tool* or rule* or instrument* or index* or test* or technique* or analy* or system* or model*)).ti,ab.
43.	(logistic* adj model*).mp.
44.	prognosis/
45.	predictive value/
46.	(prognos* or predict*).ti,ab.
47.	or/39-46
48.	38 and 47
49.	chads*.ti,ab.
50.	cha2ds2*.ti,ab.
51.	"cha(2)ds(2)-vasc".ti,ab.
52.	("Anticoagulation and Risk Factors in Atrial Fibrillation" or ATRIA).ti,ab.
53.	Q stroke.ti,ab.
54.	ABC Stroke.ti,ab.
55.	or/48-54
56.	22 and 55

57.	random*.ti,ab.
58.	factorial*.ti,ab.
59.	(crossover* or cross over*).ti,ab.
60.	((doubl* or singl*) adj blind*).ti,ab.
61.	(assign* or allocat* or volunteer* or placebo*).ti,ab.
62.	crossover procedure/
63.	single blind procedure/
64.	randomized controlled trial/
65.	double blind procedure/
66.	or/57-65
67.	systematic review/
68.	Meta-Analysis/
69.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
70.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
71.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
72.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
73.	(search* adj4 literature).ab.
74.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
75.	cochrane.jw.
76.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
77.	or/67-76
78.	Epidemiologic studies/
79.	Observational study/
80.	exp Cohort studies/
81.	(cohort adj (study or studies or analys* or data)).ti,ab.
82.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
83.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
84.	Controlled Before-After Studies/
85.	Historically Controlled Study/
86.	Interrupted Time Series Analysis/
87.	(before adj2 after adj2 (study or studies or data)).ti,ab.
88.	exp case control study/
89.	case control*.ti,ab.
90.	Cross-sectional studies/
91.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
92.	or/78-91
93.	exp "sensitivity and specificity"/
94.	(sensitivity or specificity).ti,ab.
95.	((pre test or pretest or post test) adj probability).ti,ab.
96.	(predictive value* or PPV or NPV).ti,ab.

97.	likelihood ratio*.ti,ab.
98.	((area under adj4 curve) or AUC).ti,ab.
99.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
100.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness)).ti,ab.
101.	diagnostic accuracy/
102.	diagnostic test accuracy study/
103.	gold standard.ab.
104.	or/93-103
105.	56 and (66 or 77 or 92 or 104)

Cochrane Library (Wiley) search terms

#1.	MeSH descriptor: [Atrial Fibrillation] explode all trees
#2.	((atrial or atria or atrium or auricular) near/3 fibrillat*):ti,ab
#3.	AF:ti,ab
#4.	#1 or #2 or #3
#5.	(stroke or strokes):ti,ab
#6.	((cerebro* or cerebral*) near/2 (accident* or apoplexy)):ti,ab
#7.	(CVA or poststroke or poststrokes):ti,ab
#8.	MeSH descriptor: [Intracranial Hemorrhages] explode all trees
#9.	(brain near/2 (attack* or hemorrhag* or haemorrhag* or bleed* or infarct*)):ti,ab
#10.	((intracerebral or intracranial or cerebral* or cerebro* or cerebrum or cerebellum or subarachnoid* or choroidal or basal ganglia or subdural) near/3 (hemorrhag* or haemorrhag* or bleed*)):ti,ab
#11.	MeSH descriptor: [Brain Infarction] explode all trees
#12.	MeSH descriptor: [Thromboembolism] this term only
#13.	MeSH descriptor: [Intracranial Embolism and Thrombosis] explode all trees
#14.	MeSH descriptor: [Carotid Artery Thrombosis] explode all trees
#15.	((brain or brainstem or cerebr* or cerebell* or vertebrobasil* or verte brobasil* or hemisphere* or intracran* or intracerebral or infratentorial or supratentorial or mca* or anterior circulation or carotid or transient or lacunar) near/3 (infarct* or thrombo* or emboli* or occlus* or hypoxi*)):ti,ab
#16.	((thrombo* or emboli*) near/2 event*):ti,ab
#17.	troponin:ti,ab
#18.	(OR #5-#17)
#19.	MeSH descriptor: [Risk] explode all trees
#20.	(risk near/3 (assess* or scheme* or rating* or tool* or rule* or index* or indices or score* or scoring or scale* or model* or system* or algorithm* or stratif* or criteria or calculat*)):ti,ab
#21.	MeSH descriptor: [Decision Support Systems, Clinical] this term only
#22.	MeSH descriptor: [Decision Support Techniques] this term only
#23.	((decision or assess* or screen*) near/3 (tool* or rule* or instrument* or index* or test* or technique* or analy* or system* or model*)):ti,ab
#24.	(logistic* near/1 model*)
#25.	MeSH descriptor: [Prognosis] explode all trees
#26.	MeSH descriptor: [Predictive Value of Tests] 3 tree(s) exploded
#27.	(prognos* or predict*):ti,ab

#28.	(OR #19-#27)
#29.	#18 AND #28
#30.	chads*:ti,ab
#31.	cha2ds2*:ti,ab
#32.	cha(2)ds(2)-vasc:ti,ab
#33.	("Anticoagulation and Risk Factors in Atrial Fibrillation" or ATRIA):ti,ab
#34.	Q stroke:ti,ab
#35.	ABC Stroke:ti,ab
#36.	(OR #29-#35)
#37.	#4 AND #36

B.2 Health Economics literature search strategy

Health economic evidence was identified by conducting a broad search relating to the Atrial Fibrillation population in NHS Economic Evaluation Database (NHS EED – this ceased to be updated after March 2015) and the Health Technology Assessment database (HTA - this ceased to be updated after March 2018). NHS EED and HTA databases are hosted by the Centre for Research and Dissemination (CRD). Additional health economics searches were run in Medline and Embase.

Table 15: Database date parameters and filters used

Database	Dates searched	Search filter used
Medline	2003– 10 September 2020	Exclusions Health economics studies
Embase	2003– 10 September 2020	Exclusions Health economics studies
Centre for Research and Dissemination (CRD)	NHSEED - 2003 to March 2015 HTA - 2003 to 31 March 2018	None

Medline (Ovid) search terms

1.	exp atrial fibrillation/
2.	((atrial or atria or atrium or auricular) adj3 fibrillat*).ti,ab.
3.	AF.ti,ab.
4.	1 or 2 or 3
5.	letter/
6.	editorial/
7.	news/
8.	exp historical article/
9.	Anecdotes as Topic/
10.	comment/
11.	case report/
12.	(letter or comment*).ti.
13.	or/5-12
14.	randomized controlled trial/ or random*.ti,ab.
15.	13 not 14

16.	animals/ not humans/
17.	exp Animals, Laboratory/
18.	exp Animal Experimentation/
19.	exp Models, Animal/
20.	exp Rodentia/
21.	(rat or rats or mouse or mice).ti.
22.	or/15-21
23.	4 not 22
24.	limit 23 to English language
25.	economics/
26.	value of life/
27.	exp "costs and cost analysis"/
28.	exp Economics, Hospital/
29.	exp Economics, medical/
30.	Economics, nursing/
31.	economics, pharmaceutical/
32.	exp "Fees and Charges"/
33.	exp budgets/
34.	budget*.ti,ab.
35.	cost*.ti.
36.	(economic* or pharmaco?economic*).ti.
37.	(price* or pricing*).ti,ab.
38.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
39.	(financ* or fee or fees).ti,ab.
40.	(value adj2 (money or monetary)).ti,ab.
41.	or/25-40
42.	24 and 41

Embase (Ovid) search terms

1.	exp atrial fibrillation/
2.	((atrial or atria or atrium or auricular) adj3 fibrillat*).ti,ab.
3.	AF.ti,ab.
4.	1 or 2 or 3
5.	letter.pt. or letter/
6.	note.pt.
7.	editorial.pt.
8.	case report/ or case study/
9.	(letter or comment*).ti.
10.	or/5-9
11.	randomized controlled trial/ or random*.ti,ab.
12.	10 not 11
13.	animal/ not human/

14.	nonhuman/
15.	exp Animal Experiment/
16.	exp Experimental Animal/
17.	animal model/
18.	exp Rodent/
19.	(rat or rats or mouse or mice).ti.
20.	or/12-19
21.	4 not 20
22.	limit 21 to English language
23.	health economics/
24.	exp economic evaluation/
25.	exp health care cost/
26.	exp fee/
27.	budget/
28.	funding/
29.	budget*.ti,ab.
30.	cost*.ti.
31.	(economic* or pharmaco?economic*).ti.
32.	(price* or pricing*).ti,ab.
33.	(cost* adj2 (effectiv* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
34.	(financ* or fee or fees).ti,ab.
35.	(value adj2 (money or monetary)).ti,ab.
36.	or/23-35
37.	22 and 36

NHS EED and HTA (CRD) search terms

#1.	MeSH DESCRIPTOR Atrial Fibrillation EXPLODE ALL TREES
#2.	(((atrial or atria or atrium or auricular) adj3 fibrillat*))
#3.	(AF)
#4.	(#1 or #2 or #3)

Appendix C: Clinical article selection

Figure 1: Flow chart of clinical study selection for the review of the most clinically and cost-effective risk stratification tool for predicting stroke or thromboembolic events in people with atrial fibrillation

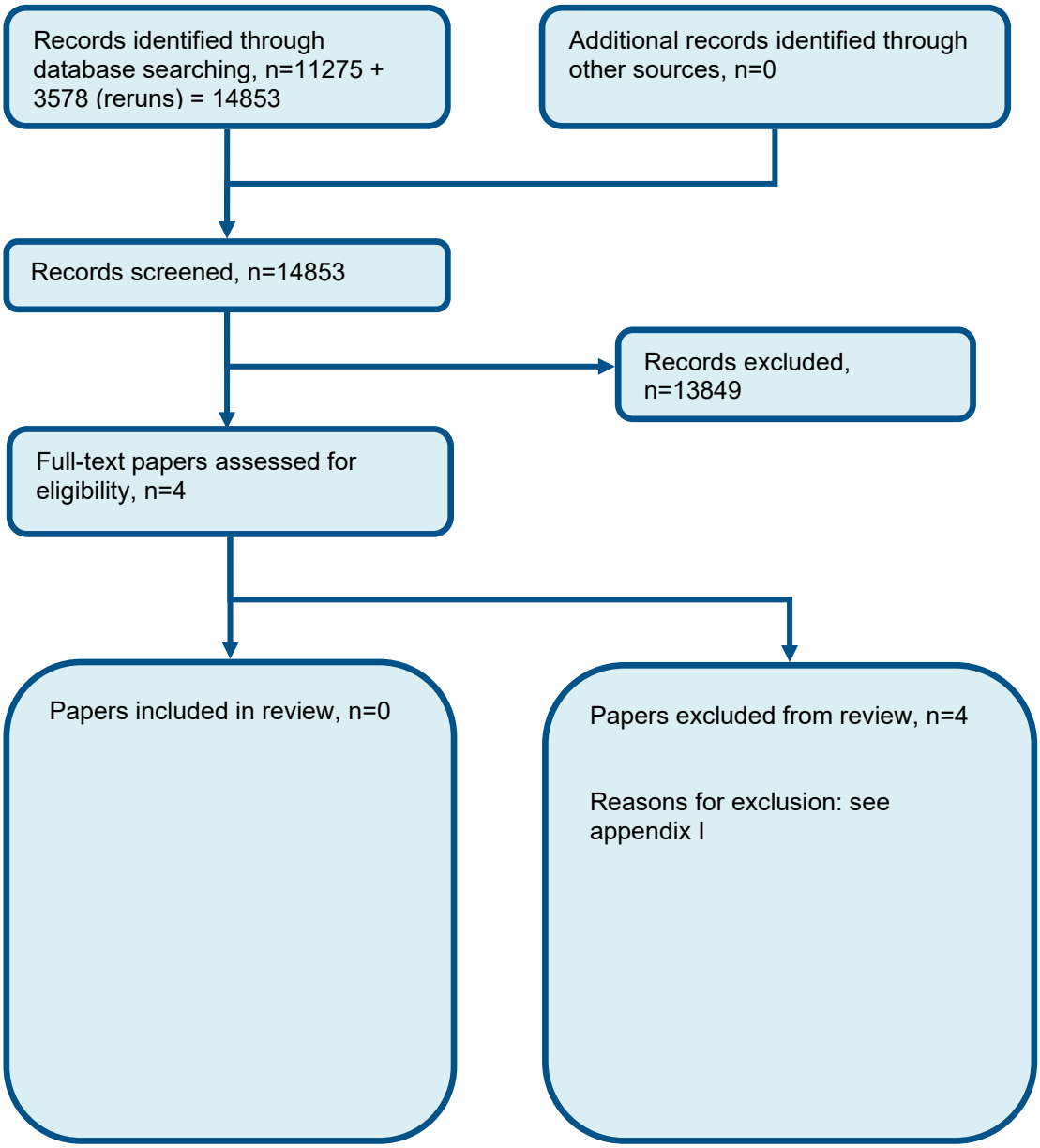
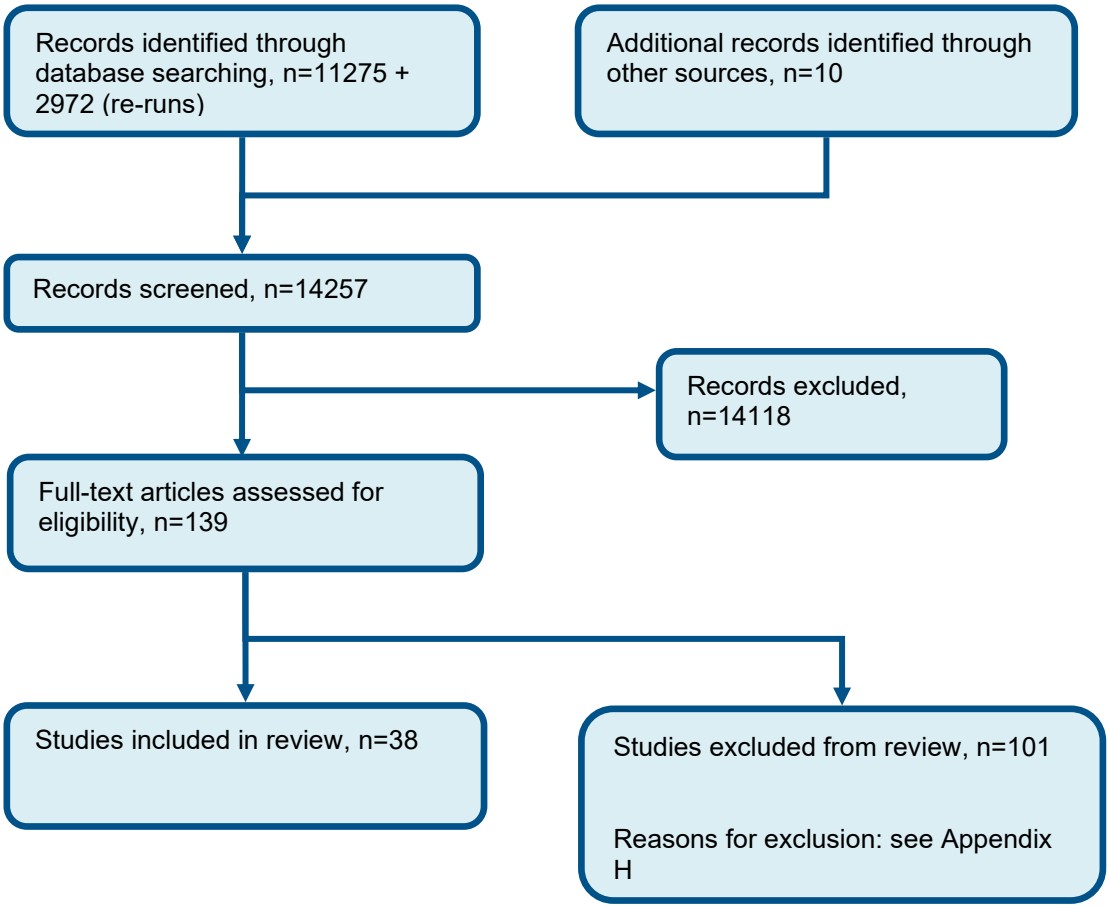


Figure 2: Flow chart of clinical article selection for the review of 'risk tools for prediction of stroke'.



Appendix D: FULL GRADE TABLES (including individual study results)

Table 16: Clinical evidence profile: Discriminative capacity of stroke prediction tools featured in the studies (see table 3).

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
CHADS2	27	884,951 (one study n is unknown)	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.65(0.62-0.67)[2] 0.69(0.69-0.70)[10] 0.66(0.66-0.67)[33] 0.82(0.80-0.84)[36] 0.58(0.50-0.67)[38] 0.61(0.59-0.65)[F] 0.63(0.61-0.66)[M][45] 0.74(0.72-0.75)[55] 0.64(0.56-0.71)[64] 0.67(0.58-0.75)[67] 0.57(0.40-0.74)[68] 0.66(0.65-0.68)[75] 0.70(0.70-0.70)[76] 0.81(0.80-0.83)[85] 0.63(0.62-0.65)[86] 0.66(0.64-0.69)[109] 0.51(0.49-0.52)[110] 0.68(0.61-0.75)[113] 0.64(0.53-0.73)[117] 0.68(0.67-0.69)[119] 0.66(0.64-0.68)[122]	VERY LOW

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
							0.65(0.60-0.69)[125] 0.87(0.81-0.93) [128] 0.70(0.68-0.73) [88] 0.62(0.60-0.65)CCAE[110] 0.56(0.55-0.58)MDLD[110] 0.64(0.61-0.68)CUMC[110] 0.72(0.53-0.91)AUSOM[110] 0.51(0.41-0.62)STRIDE[110] **POOLED EFFECT: Random Effects: 0.66(0.64-0.69); I^2=98%** **Results that could not be pooled due to lack of variance measures:** 0.67[27] 0.62[123] 0.7[35] 0.64[110]Optum claims0.65[110]Optum EHR	
Modified CHADS2 (Van Staa, 2010)	1	79,884	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.69(0.67-0.71)[122]	VERY LOW

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
Revised CHADS2 (Friberg 2012)	2	91,574	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.62(0.61-0.62)[33] 0.55(0.37-0.73)[68] **POOLED EFFECT: Fixed Effects: 0.62(0.61-0.63); I^2=0%**	LOW
R2 CHADS2 (Abumaileq 2015, Yoshizawa, 2017, Piccini, 2013)	3	16846	Very serious risk of bias[a]	Very serious risk of inconsistency[b]	No serious indirectness	Serious imprecision	0.65(0.53-0.78)[3] 0.85(0.79-0.91)[128] 0.70(0.67-0.73) [88] **POOLED EFFECT: Random Effects: 0.74(0.62-0.86); I^2=92%**	VERY LOW
R2CHADS2 (71 points) (McAlister, 2017)	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.66(0.64-0.67)[75]	LOW
CHADS2 KDIGO (McAlister, 2017)	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.65(0.64-0.66)[75]	LOW
CHADS2 Alb (McAlister, 2017)	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.65(0.64-0.67)[75]	LOW
CHADS2 eGFR	1	58,451	Very serious	No serious inconsistency	No serious indirectness	No serious imprecision	0.67(0.65-0.68)[75]	LOW

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
(McAlister, 2017)			risk of bias[a]					
CHADS2 with vWF	1	994	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.69(0.60-0.77)[67]	VERY LOW
CHADSVASC 2009	27	987032 (in one study n unknown)	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.67(0.65-0.69)[2] 0.69(0.53-0.85)[3] 0.69(0.69-0.70)[10] 0.69(0.68-0.69)[20] 0.69(0.63-0.76)[30] 0.67(0.66-0.68)[33] 0.72(0.64-0.81)[38] 0.62(0.59-0.65)[F] 0.67(0.65-0.69)[M][45] 0.71(0.69-0.73)[55] 0.66(0.59-0.72)[64] 0.64(0.56-0.71)[67] 0.58(0.44-0.73)[68] 0.66(0.65-0.67)[75] 0.62(0.61-0.63)[76] 0.89(0.88-0.90[85] 0.66(0.65-0.68)[84] 0.68(0.66-0.70)[109] 0.53(0.51-0.54)[110] 0.67(0.61-0.74)[113] 0.68(0.67-0.69)[119] 0.67(0.65-0.69)[122] 0.62(0.57-0.66)[125]	VERY LOW

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
							0.60(0.51-0.68)[126] 0.89(0.85-0.95)[128] 0.64(0.58-0.70)[73] ARIC cohort 0.68(0.52-0.84) [73] MESA cohort 0.65(0.62-0.67)CCAE[110] 0.58(0.56-0.59)MDLD[110] 0.66(0.63-0.69)CUMC[110] 0.81(0.71-0.90)AUSOM[110] 0.55(0.44-0.65)STRIDE[110] **POOLED EFFECT: Random Effects: 0.67(0.65-0.69); I^2=99%** **Results that could not be pooled due to lack of variance measures:** 0.61[124] 0.65[110]Optum claims 0.67[110]Optum EHR	
P2-CHADSVASC	2	2929	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.67(0.60-0.75)[73] ARIC cohort 0.75(0.60-0.91)[73] MESA cohort **POOLED EFFECT: Fixed effect 0.68 (0.62-0.75) I^2=0%**	VERY LOW
Age modified CHADSVASC	1	124,271	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.71(0.70-0.71)[20]	MODERATE

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
(Chao 2016)								
mCHADS VASC (modified in Tomita 2015)	1	997	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.60(0.51-0.68)[117]	LOW
Modified CHADSVASC (no stroke/TIA)[108]	1	11433	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.65(0.57-0.72)(non-white)[108] 0.68(0.64-0.72)(white)[108]	VERY LOW
mCHADS VA – (Modified in Tomita 2015)	1	997	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.62(0.53-0.71)[117]	VERY LOW
Q STROKE	2	320,043	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.65(0.62-0.67) [Female] [45]0.71(0.69-0.73)[Male] [45] 0.62(0.60-0.64)CCAE[110] 0.55(0.53-0.56)MDLD[110] 0.57(0.53-0.61)CUMC[110] 0.68(0.42-0.94)AUSOM[110] 0.47(0.36-0.57)STRIDE[110] **POOLED EFFECT: Random Effects: 0.61(0.56-0.66); I^2=96%**	LOW

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
							Results that could not be pooled due to lack of variance measures: 0.58[110]Optum claims 0.6[110]Optum EHR	
ATRIA	7	572,012 (one study unknown n)	Very serious risk of bias[a]	Very serious inconsistency	No serious indirectness	serious imprecision	0.64(0.49-0.80)[3] 0.71(0.70-0.71)[10] 0.67(0.66-0.68)[75] 0.76(0.755-0.765)[76] 0.70(0.67-0.72)[109] 0.70(0.69-0.71)[119] 0.63(0.61-0.66)CCAE[110] 0.58(0.56-0.59)MDLD[110] 0.67(0.64-0.70)CUMC[110] 0.79(0.63-0.94)AUSOM[110] 0.53(0.43-0.63)STRIDE[110] **POOLED EFFECT: Random Effects: 0.67 (0.64-0.70); I^2=98%** **Results that could not be pooled due to lack of variance measures:** 0.67[110]Optum claims 0.67[110]Optum EHR	VERY LOW

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
AFI 1994	7	182,064	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.58(0.58-0.59)[33] 0.68(0.65-0.71)[36] 0.60(0.39-0.73)[68] 0.60(0.58-0.61)[122] **POOLED EFFECT: Random Effects: 0.62(0.57-0.66); I^2=92%** **Results that could not be pooled due to lack of variance measures:** 0.61[27] 0.63[35] 0.61[123]	VERY LOW
AFI 1998	1	79,884	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.61(0.60-0.62)[122]	LOW
SPAF 1995	5	90,490	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	Serious imprecision	0.74(0.71-0.76)[36] 0.63(0.61-0.65)[122] **POOLED EFFECT: Random Effects: 0.68(0.58-0.79); I^2=97%** **Results that could not be pooled due to lack of variance measures:** 0.65[27] 0.64[35] 0.62[123]	VERY LOW

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
SPAF 1999	2	91,574	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Very serious imprecision[c]	0.63(0.62-0.64)[33] 0.51(0.33-0.67)[68] **POOLED EFFECT: Random Effects: 0.60(0.49-0.70); I^2=50%**	VERY LOW
FRAMINGHAM	7	492685	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.67(0.66-0.68)[33] 0.61(0.42-0.79)[68] 0.65(0.63-0.68)[122] 0.64(0.61-0.66)CCAE[110] 0.57(0.56-0.59)MDLD[110] 0.66(0.63-0.69)CUMC[110] 0.76(0.59-0.93)AUSOM[110] 0.62(0.51-0.72)STRIDE[110] **POOLED EFFECT: Fixed Effects: 0.64(0.60-0.67); I^2=97%** **Results that could not be pooled due to lack of variance measures:** 0.69[27] 0.69[35] 0.66[123] 0.65[110]Optum claims 0.66[110]Optum EHR	LOW
ACCP 2001	2	82,464	Very serious	No serious inconsistency	No serious indirectness	No serious imprecision	0.58[35] 0.62(0.60-0.62)[122] **Range:0.58 to 0.62**	LOW

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
			risk of bias[a]				**Median: 0.60**	
ACCP 2004	2	85,472	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.6[27] 0.61(0.60-0.62)[122] **Range: 0.60 to 0.61** **Median: 0.605**	LOW
ACCP 2008	2	80,968	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.56(0.39-0.73)[68] 0.64(0.62-0.65)[122] **POOLED EFFECT: Fixed Effects: 0.64(0.62-0.66); I^2=0%**	LOW
ACC/AHA/ESC guidelines 2006	3	171,458	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.62(0.61-0.62)[33] 0.55(0.38-0.72)[68] 0.64(0.62-0.66)[122] **POOLED EFFECT: Fixed Effects: 0.62(0.61-0.63); I^2=47%**	LOW
NICE	3	171,458	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	No serious imprecision	0.61(0.60-0.62)[33] 0.57(0.42-0.72)[68] 0.64(0.62-0.65)[122] **POOLED EFFECT: Random Effects: 0.62(0.59-0.65); I^2=72%**	VERY LOW
Hart 1998	1	79,884	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.62(0.60-0.64)[122]	LOW

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
Van Walraven	1	79,884	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.55(0.54-0.58)[122]	LOW
Van Latum	1	79,884	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.57(0.55-0.59)[122]	LOW
CHADSVASC with vWF	1	994	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.68(0.59-0.76)[67]	VERY LOW
GARFIELD	1	2301	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.70(0.63-0.77)[30]	VERY LOW
SAMe-TT2R2	1	3483	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.51(0.49-0.53)[69]	LOW
Sum of CrCl <60 mL/min and prior stroke/TIA[88]	1	16,360	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.61 (0.58-0.64)[88]	LOW

Pooling (meta-analysis) was carried out if there were at least two studies per risk tool with confidence intervals. RevMan was used to carry out the analyses. If pooling was not possible for risk tools with >1 data point then the range and median value of the study point estimates were recorded. If there were only one data point then only the result from the study was recorded.

a) Risk of bias was assessed using the PROBAST checklist (see Appendix F).Risk of bias was serious for some risk tools because none of the studies reported any blinding of assessors for risk tool data and outcome status, and most did not report loss to follow up, although follow up and number of events were appropriate. Risk of bias was very serious for the rest of the risk tools because many studies with the aforementioned limitations also had insufficient numbers of events (<100) and/or inappropriately short follow up times (<5 years) to be able to accurately predict risk.
b) Where data were pooled, an I^2 of 50-74% was deemed serious inconsistency and an I^2 of 75% or above was deemed very serious inconsistency. If no pooling were possible, inconsistency was assessed by inspection of the degree of overlap of confidence intervals between studies: if one of more Cis did not overlap then a rating of serious inconsistency was given. Reasons for heterogeneity between studies may include geographical/cultural/ethnic differences. Clinically the studies appeared reasonably homogeneous, with similar rates of hypertension, diabetes and former stroke.
c) The judgement of precision was based on the spread of confidence interval across two clinical thresholds: C statistics of 0.5 and 0.7. The threshold of 0.5 marked the boundary between no predictive value better than chance and a predictive value better than chance. The threshold of 0.7 marked the boundary above which the committee might consider recommendations. If the 95% Cis crossed one of these thresholds a rating of serious imprecision was given and if they crossed both of these thresholds a rating of very serious imprecision as given.

Table 17: Clinical evidence profile: sensitivity and specificity of stroke prediction tools featured in the studies (see table 3). For pooled data the 95% CIs of individual studies can be found in the Forest plots in the appendices. For individual or non-pooled data the 95% CIs are given below. The pooled sensitivity/specificity values have been calculated using Bayesian methodology and are expressed as medians (95% credible intervals).

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
CHADS2 at threshold of ≥1	6	172,747	**At threshold for risk of ≥1** 0.769[49] 0.842[10] 0.840[2] 0.978[39] 0.570[71]	**At threshold for risk of ≥1** 0.389[49] 0.205[10] 0.306[2] 0.072[39] 0.537[71]	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
					specificity				

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
			0.869[127] **Pooled sensitivity: 0.874(0.676-0.960)**	0.307[127] **Pooled specificity: 0.228(0.131-0.501)**	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
CHADS2 at threshold of ≥2	5	165,058	**At threshold for risk of ≥2** 0.743[10] 0.365[2] 0.790[39] 0.320[71] 0.638[127] **Pooled sensitivity: 0.582(0.308-0.811)**	**At threshold for risk of ≥2** 0.409[10] 0.787[2] 0.344[39] 0.814[71] 0.634[127] **Pooled specificity: 0.625(0.363-0.835)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
CHADS2 at threshold of ≥3	5	165,058	**At threshold for risk of ≥3** 0.495[10] 0.133[2] 0.550[39] 0.148[71] 0.405[127] **Pooled sensitivity: 0.316(0.129-0.593)**	**At threshold for risk of ≥3** 0.707[10] 0.935[2] 0.649[39] 0.932[71] 0.844[127] **Pooled specificity: 0.845(0.641-0.944)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	No serious imprecision	VERY LOW
Revised CHADS2 (Friberg 2012)	1	90,490	0.980[36] at standard threshold [no raw data	0.150[36] at standard threshold [no raw data in	**Sensitivity**				
					Very serious	NA	No serious indirectness	NA	LOW

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
			in paper, and no 95% CIs reported]	paper, and no 95% CIs reported]	risk of bias[a]				
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
R2CHADS2 (71 points) (McAlister, 2017)	1	7340	0.800[83] no specified threshold	0.511[83] no specified threshold	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
CHADS2 KDIGO (McAlister, 2017)	1	7340	0.726[83] no specified threshold	0.575[83] no specified threshold	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
CHADS2 Alb (McAlister, 2017)	1	7340	0.821[83] no specified threshold	0.488[83] no specified threshold	**Sensitivity**				
					Very serious	NA	No serious indirectness	NA	LOW

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					risk of bias[a]				
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
CHADS2 eGFR (McAlister, 2017)	1	7340	0.693[83] no specified threshold	0.640[83] no specified threshold	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
CHADSVASC 2009 at threshold of ≥1	9	440,691	**At threshold for risk of ≥1** 0.966[49] 0.987[10] 1.000[3] 0.967[21] 0.890[71] 1.000[75] 0.927[127] 0.964[130]	**At threshold for risk of ≥1** 0.164[49] 0.086[10] 0.034[3] 0.057[21] 0.160[71] 0.090[75] 0.174[127] 0.162[130]	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	No serious imprecision	VERY LOW
					Specificity				

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
			0.999[138] **Pooled sensitivity: 0.977(0.947-0.992)**	0.025[138] **Pooled specificity: 0.092(0.051-0.156)**	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
CHADSVASC 2009 at threshold of ≥2	9	438983	**At threshold for risk of ≥2** 0.957[10] 0.952[2] 1.000[3] 0.868[21] 0.695[71] 0.960[75] 0.840[127] 0.895[130] 0.982[138] **Pooled sensitivity: 0.923(0.850-0.964)**	**At threshold for risk of ≥2** 0.195[10] 0.168[2] 0.158[3] 0.169[21] 0.450[71] 0.249[75] 0.372[127] 0.297[130] 0.088[138] **Pooled specificity: 0.223(0.144-0.328)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	No serious imprecision	VERY LOW
CHADSVASC 2009 at threshold of ≥3	8	569,938	**At threshold for risk of ≥3** 0.864[10] 0.742[2] 0.693[21] 0.390[71]	**At threshold for risk of ≥3** 0.3395[10] 0.476[2] 0.323[21] 0.710[71]	**Sensitivity**				
					Serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
			0.840[75] 0.681[127] 0.716[130] 0.933[138] **Pooled sensitivity: 0.809(0.631-0.913)**	0.420[75] 0.558[127] 0.484[130] 0.177[138] **Pooled specificity: 0.431(0.287-0.582)**	**Specificity**				
					Serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
CHADSVASC 2009 at threshold of ≥4	8	438,829	**At threshold for risk of ≥4** 0.511[130] 0.845[138] 0.41[2] 0.69[10] 0.480[21] 0.200[71] 0.520[75] 0.490[127] **Pooled sensitivity: 0.524(0.347-0.695)**	**At threshold for risk of ≥4** 0.671[130] 0.318[138] 0.770[2] 0.530[10] 0.500[21] 0.870[71] 0.630[75] 0.740[127] **Pooled specificity: 0.646(0.477-0.781)**	**Sensitivity**				
					Serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
Modified CHADSVASC (no stroke/TIA)[122] at threshold for risk of >2	1	11,433	**At threshold for risk of ≥2** 0.821(0.759-0.872)[122]	**At threshold for risk of >2** 0.393(0.384-0.402)[122]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very serious	NA	No serious indirectness	No serious imprecision	LOW

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					risk of bias[a]				
Modified CHADSVASC (no stroke/TIA)[122] at threshold for risk of >3	1	11,433	**At threshold for risk of ≥3** 0.631(0.559-0.699)[122]	**At threshold for risk of ≥3** 0.612(0.603-0.621)[122]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	Serious imprecision	VERY LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
Modified CHADSVASC (no stroke/TIA)[122] at threshold for risk of >4	1	11,433	**At threshold for risk of ≥4** 0.359(0.292-0.431)[122]	**At threshold for risk of ≥4** 0.798(0.791-0.805)[122]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
Q STROKE with optimal cut-off at top 63%	1	7689	0.825 (0.798-0.849)[49] with optimal cut-off at top 63%	0.395(0.383-0.407)[49] with optimal cut-off at top 63%	**Sensitivity**				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
					Specificity				

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
Q STROKE with at top 90%	1	7689	0.992(0.984-0.997)[49] with cut-off at top 90%	0.112(0.105-0.119)[49] with cut-off at top 90%	**Sensitivity**				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
					Specificity				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
Q STROKE with at top 85%	1	7689	0.979(0.967-0.987)[49] with cut-off at top 85%	0.167(0.158-0.176[49] with cut-off at top 85%	**Sensitivity**				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
					Specificity				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
Q STROKE with at top 80%	1	7689	0.958(0.943-0.971)[49] with cut-off at top 80%	0.221(0.211-0.231)[49] with cut-off at top 80%	**Sensitivity**				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
					Specificity				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
	1	7689			**Sensitivity**				

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Q STROKE with at top 70%			0.890(0.868-0.909)[49] with cut-off at top 70%	0.325(0.314-0.336)[49] with cut-off at top 70%	Serious risk of bias[a]	NA	No serious indirectness	Serious imprecision[c]	LOW
					Specificity				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	MOD
ATRIA at threshold for risk of ≥1	2	158004	**At threshold for risk of ≥1** 0.994[66][no raw data] 0.985(0.983-0.987)[10] **Median[d]: 0.985(0.983-0.987)**	**At threshold for risk of ≥1** 0.0820[66][no raw data] 0.091(0.089-0.168)[10] **Median[d]: 0.091(0.089-0.168)[10]**	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	Serious imprecision[c]	VERY LOW
ATRIA at threshold for risk of ≥2	1	152149	**At threshold for risk of ≥2** 0.967(0.964-0.970)[10]	**At threshold for risk of ≥2** 0.166(0.164-0.168)[10]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
	1	152149			**Sensitivity**				

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
ATRIA at threshold for risk of ≥3			**At threshold for risk of ≥3** 0.958(0.955-0.962)[10]	**At threshold for risk of ≥3** 0.192(0.189-0.194)[10]	Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
ATRIA at threshold for risk of ≥4	1	152149	**At threshold for risk of ≥4** 0.936(0.931-0.940)[10]	**At threshold for risk of ≥4** 0.241(0.238-0.243)[10]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
ATRIA at threshold for risk of ≥5	1	152149	**At threshold for risk of ≥5** 0.894(0.888-0.899)[10]	**At threshold for risk of ≥5** 0.309(0.307-0.312)[10]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
	3	158158			**Sensitivity**				

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
ATRIA at threshold for risk of ≥6			**At threshold for risk of ≥6** 0.748[66][no raw data] 0.831(0.390-0.395)[10] 0.444(0.137-0.788)[3] **Median[d]: 0.444(0.137-0.788)**	**At threshold for risk of ≥6** 0.610[66] [no raw data] 0.393(0.390-0.395)[10] 0.510(0.426-0.594)[3] **Median[d]: 0.510(0.426-0.594)**	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency	No serious indirectness	Serious imprecision[c]	VERY LOW
ATRIA at threshold for risk of ≥7	2	152303	**At threshold for risk of ≥7** 0.698(0.689-0.706)[10] 0.444(0.137-0.788)[3] **Median[d]: 0.444(0.137-0.788)**	**At threshold for risk of ≥7** 0.527(0.524-0.529)[10] 0.607(0.522-0.687)[3] **Median[d]: 0.607(0.522-0.687)**	**Sensitivity**				
					Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	LOW
AFI 1994	1	90,490	0.990[36] at standard threshold[no raw data in paper, and no 95% CIs reported]	0.090[36] at standard threshold[no raw data in paper, and no 95% CIs reported]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
SPAF 1999	1	90,490	0.890[36] at standard threshold[no raw data in paper, and no 95% CIs reported]	0.290[36] at standard threshold[no raw data in paper, and no 95% CIs reported]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
FRAMINGHAM	1	90,490	0.920[36] at standard threshold[no raw data in paper, and no 95% CIs reported]	0.260[36] at standard threshold[no raw data in paper, and no 95% CIs reported]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
ACC/AHA/ESC guidelines 2006	1	90,490	0.980[36] at standard threshold[no raw data in paper, and no 95% CIs reported]	0.150[36] at standard threshold[no raw data in paper, and no 95% CIs reported]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW

Prediction tool	No of studies	n	Sensitivity	Specificity	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
NICE	1	90,490	1.000[36] at standard threshold[no raw data in paper, and no 95% CIs reported]	0.090[36] at standard threshold[no raw data in paper, and no 95% CIs reported]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW

Pooling (meta-analysis) was carried out if there were at least three studies per risk tool with confidence intervals. RevMan and WinBugs were used to carry out the analyses. If pooling was not possible for risk tools with >1 data point then the range and median value of the study point estimates were recorded. If there were only one data point then only the result from the study was recorded.
a) Risk of bias was assessed using the PROBAST checklist. Risk of bias was serious for some risk tools because none of the studies reported any blinding of assessors for risk tool data and outcome status, and most did not report loss to follow up, although follow up and number of events were appropriate. Risk of bias was very serious for the rest of the risk tools because many studies with the aforementioned limitations also had insufficient numbers of events (<100) and/or inappropriately short follow up times (<5 years) to be able to accurately predict risk.
b) Where data were pooled, inconsistency was assessed by visual inspection of the sensitivity/specificity plots, or data (if 2 studies). The evidence was downgraded by 1 increment if there was no overlap of 95% confidence intervals. For single studies no evaluation was made and 'not applicable' was recorded.
c) Imprecision was assessed based on inspection of the confidence region in the meta-analysis or, where meta-analysis has not been conducted, assessed according to the range of confidence intervals in the individual studies. The evidence was downgraded by 1 increment when the confidence interval around the point estimate crossed one of the clinical thresholds (0.90 or 0.60 for sensitivity and 0.5 and 0.1 for specificity), and downgraded by 2 increments when the confidence interval around the point estimate crossed both of the clinical thresholds. The upper clinical threshold marked the point above which recommendations would be possible, and the lower clinical threshold marked the point below which the tool would be regarded as of little clinical use.
d)For unpooled data the median value was given (of data with 95% CIs). If there were an even number of data points in the unpooled data, the data point chosen in the central pair was the one with lower sensitivity, with its paired specificity.

Table 18: Clinical evidence profile: D statistics of prediction tools featured in the studies (see table 3)

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	D statistic (95%CI)	Quality
Q Stroke [female]	1	3180	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.820(0.660-0.990) [Female][49]	MODERATE
Q Stroke [male]	1	4509	serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[b]	1.150(1.000 to 1.300) [Male][49]	LOW
CHADS2 [female]	1	3180	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.640(0.490-0.810) [Female][49]	MODERATE
CHADS2 [male]	1	4509	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.810(0.660 to 0.960) [Male][49]	MODERATE
CHADSVASC [female]	1	3180	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.670(0.510-0.830) [Female][49]	MODERATE
CHADSVASC [male]	1	4509	serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[b]	0.970(0.820 to 1.120) [Male][49]	LOW

a) Risk of bias was assessed using the PROBAST checklist (see Appendix F).Risk of bias was serious for most risk tools because none of the studies reported any blinding of assessors for risk tool data and outcome status, and most did not report loss to follow up, although follow up and number of events were appropriate. Risk of bias was very serious for the Framingham risk tool because the study concerned also had insufficient numbers of events (<100) and/or inappropriately short follow up times (<5 years) to be able to accurately predict risk.
b) The judgement of precision was based on the spread of confidence intervals around the clinically important point at 1.1. If the CIs crossed 1.1 then they were graded as seriously imprecise

Table 19: Clinical evidence profile: calibration statistics of prediction tools featured in the studies (see table 3)

Prediction tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	R^2 (95%CI)	Hosmer-Lemeshow statistics	Quality
Q Stroke [female]	1	3180	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision[b]	0.140(0.092-0.187)[Female] [49]	-	MODERATE
Q Stroke [male]	1	4509	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision[b]	0.241(0.193-0.289)[Male][49]	-	MODERATE
CHADS2 [female]	1	3180	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision[b]	0.091(0.049-0.132)[Female] [49]	-	MODERATE
CHADS2 [male]	1	4509	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision[b]	0.135(0.091-0.179) [Male][49]	-	MODERATE
CHADSVASC [female]	1	3180	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision[b]	0.096(0.055-0.138)[Female] [49]	-	MODERATE
CHADSVASC [male]	1	4509	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision[b]	0.183(0.137-0.228) [Male][49]	-	MODERATE
Framingham	1	705	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	-	7.6[137] (values <20 indicate good calibration. No CIs or p value provided in study.	LOW

a) Risk of bias was assessed using the PROBAST checklist (see Appendix F).Risk of bias was serious for most risk tools because none of the studies reported any blinding of assessors for risk tool data and outcome status, and most did not report loss to follow up, although follow up and number of events were appropriate. Risk of bias was very serious for the

Framingham risk tool because the study concerned also had insufficient numbers of events (<100) and/or inappropriately short follow up times (<5 years) to be able to accurately predict risk.
b) The judgement of precision was based on the spread of confidence intervals around the clinically important point at 0.5. If the CIs crossed 0.5 then they were graded as seriously imprecise.

Table 19: Clinical evidence profile: NRI of prediction tools featured in the studies (see table 3) with CHADS2 as the comparator

Prediction tool comparison	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	[NRI(95% CI)](95%CI)	Quality
ATRIA versus CHADS2	4	259,504	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	No serious imprecision	+0.160 (0.140-0.170)[10] (This mainly resulted from up-classification (that is, CHADS2 tended to lead to more false negatives) +0.137 (0.120 to 0.153)[133] (Mainly due to down-classification) +0.240 (0.170 to 0.310)[123] +0.008 (-0.010 to 0.026)[83] **POOLED EFFECT: Random effects NRI +0.130 (+0.050 to +0.220); I^2=98%**	VERY LOW
R2CHADS2 (71 point) versus CHADS2	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	-0.015 (-0.036 to 0.006)[83]	VERY LOW
R2CHADS2 versus CHADS2	1	16,360	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.226(0.125 to 0.307)[100]	LOW
CHADS2 KDIGO versus CHADS2	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.026(-0.049 to -0.002)[83]	LOW
CHADS2 Alb versus CHADS2	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	-0.018 (-0.026 to 0.028)[83]	VERY LOW
CHADS2 eGFR versus CHADS2	1	58,451	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.006 (-0.017 to 0.030)[83]	VERY LOW

CHASDS2 with vascular disease versus CHADS2	1	2002	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.400 (0.000 to +0.800)[95]	MODERATE

Pooling (meta-analysis) was carried out if there were at least two studies per risk tool with confidence intervals. RevMan was used to carry out the analyses. If pooling was not possible for risk tools with >1 data point then the range and median value of the study point estimates were recorded. If there were only one data point then only the result from the study was recorded.
a) Risk of bias was assessed using the PROBAST checklist (see Appendix F).Risk of bias was serious for some risk tools because none of the studies reported any blinding of assessors for risk tool data and outcome status, and most did not report loss to follow up, although follow up and number of events were appropriate. Risk of bias was very serious for the rest of the risk tools because many studies with the aforementioned limitations also had insufficient numbers of events (<100) and/or inappropriately short follow up times (<5 years) to be able to accurately predict risk.
b) Where data were pooled, an I^2 of 50-74% was deemed serious inconsistency and an I^2 of 75% or above was deemed very serious inconsistency. If no pooling were possible, inconsistency was assessed by inspection of the degree of overlap of confidence intervals between studies: if one of more Cis did not overlap then a rating of serious inconsistency was given. Reasons for heterogeneity between studies may include geographical/cultural/ethnic differences. Clinically the studies appeared reasonably homogeneous, with similar rates of hypertension, diabetes and former stroke.
c) The judgement of precision was based on the spread of confidence intervals. If the lower 95% CI passed across 0 then this was graded as seriously imprecise

Table 20: Clinical evidence profile: NRI of prediction tools featured in the studies (see table 3) with CHADSVASC (or CHADSVASC derivatives) as the comparator

Prediction tool comparison	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	[NRI(95% CI)](95%CI)	Quality
ATRIA versus CHADSVASC	3	210,053	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	No serious imprecision	+0.210(0.200-0.230)[10] (This mainly resulted from down-classification (that is, that CHADSVASC tended to lead to more false positives) +0.233 (0.219 to 0.248)[133](wholly due to down classification) +0.250 (0.210 to 0.300)[123] **POOLED EFFECT: Random effects NRI +0.230 (+0.200 to +0.250); I^2=79%**	VERY LOW

Age-modified CHADSVASC versus CHADSVASC	1	124,271	serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.039 (0.0216 to 0.0459)[21]	MODERATE
CHADS2 versus CHADSVASC	8	210,854	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	-0.211 (-0.340 to -0.090) [2] -0.166 (-0.291 to -0.039)[41] +0.005(+0.011 to +0.021)[83] +0.017 (0.000 to +4.200)[60] +0.030 (+0.010 to +0.060)[71] -0.142 (-0.230 to -0.060)[97] +0.237 (0.000 to 0.470)[139] **POOLED EFFECT: Random effects NRI -0.020 (-0.060 to +0.020); I^2=84%** **Not pooled because of lack of 95% CIs:** +0.070[36]	VERY LOW
Revised CHADS2 versus CHADSVASC	1	90,490	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Not applicable	+0.070[36]	LOW
Framingham versus CHADSVASC	1	90,490	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Not applicable	+0.120[36]	LOW
SPAF 1999 versus CHADSVASC	1	90,490	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Not applicable	+0.120[36]	LOW
ACC/AHA/ESC versus CHADSVASC	1	90,490	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Not applicable	+0.070[36]	LOW
NICE versus CHADSVASC	1	90,490	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Not applicable	+0.000[36]	LOW

AFI 1994 versus CHADSVASC	1	90,490	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Not applicable	+0.000[36]	LOW
CHADS2 versus mCHADSVASC	1	997	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	-0.100(-0.280 to 0.080)[131]	VERY LOW
CHADS2 versus mCHADSVA	1	997	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	-0.030 (-0.210 to 0.160)[131]	VERY LOW
mCHADSVASC versus mCHADSVA	1	997	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.110(0.010 to 0.200)[131]	LOW
P2-CHADSVASC versus CHADSVASC	2	2929	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	No serious imprecision	+0.250(0.130-0.390)[81 2029] ARIC cohort +0.510(0.180-0.860)[81 2029] MESA cohort **POOLED EFFECT: Random effects NRI +0.330 (+0.100 to +0.570); I^2=53%**	VERY LOW

Pooling (meta-analysis) was carried out if there were at least two studies per risk tool with confidence intervals. RevMan was used to carry out the analyses. If pooling was not possible for risk tools with >1 data point then the range and median value of the study point estimates were recorded. If there were only one data point then only the result from the study was recorded.
a) Risk of bias was assessed using the PROBAST checklist (see Appendix F).Risk of bias was serious for some risk tools because none of the studies reported any blinding of assessors for risk tool data and outcome status, and most did not report loss to follow up, although follow up and number of events were appropriate. Risk of bias was very serious for the rest of the risk tools because many studies with the aforementioned limitations also had insufficient numbers of events (<100) and/or inappropriately short follow up times (<5 years) to be able to accurately predict risk.
b) Where data were pooled, an I^2 of 50-74% was deemed serious inconsistency and an I^2 of 75% or above was deemed very serious inconsistency. If no pooling were possible, inconsistency was assessed by inspection of the degree of overlap of confidence intervals between studies: if one of more Cis did not overlap then a rating of serious inconsistency was given. Reasons for heterogeneity between studies may include geographical/cultural/ethnic differences. Clinically the studies appeared reasonably homogeneous, with similar rates of hypertension, diabetes and former stroke.
c) The judgement of precision was based on the spread of confidence intervals. If the lower 95% CI passed across 0 then this was graded as seriously imprecise

Appendix E: Forest plots

Note that Forest Plots have not been presented for prediction tools with only a single study.

E.1 C statistics

Figure 3: C statistic in CHADS2

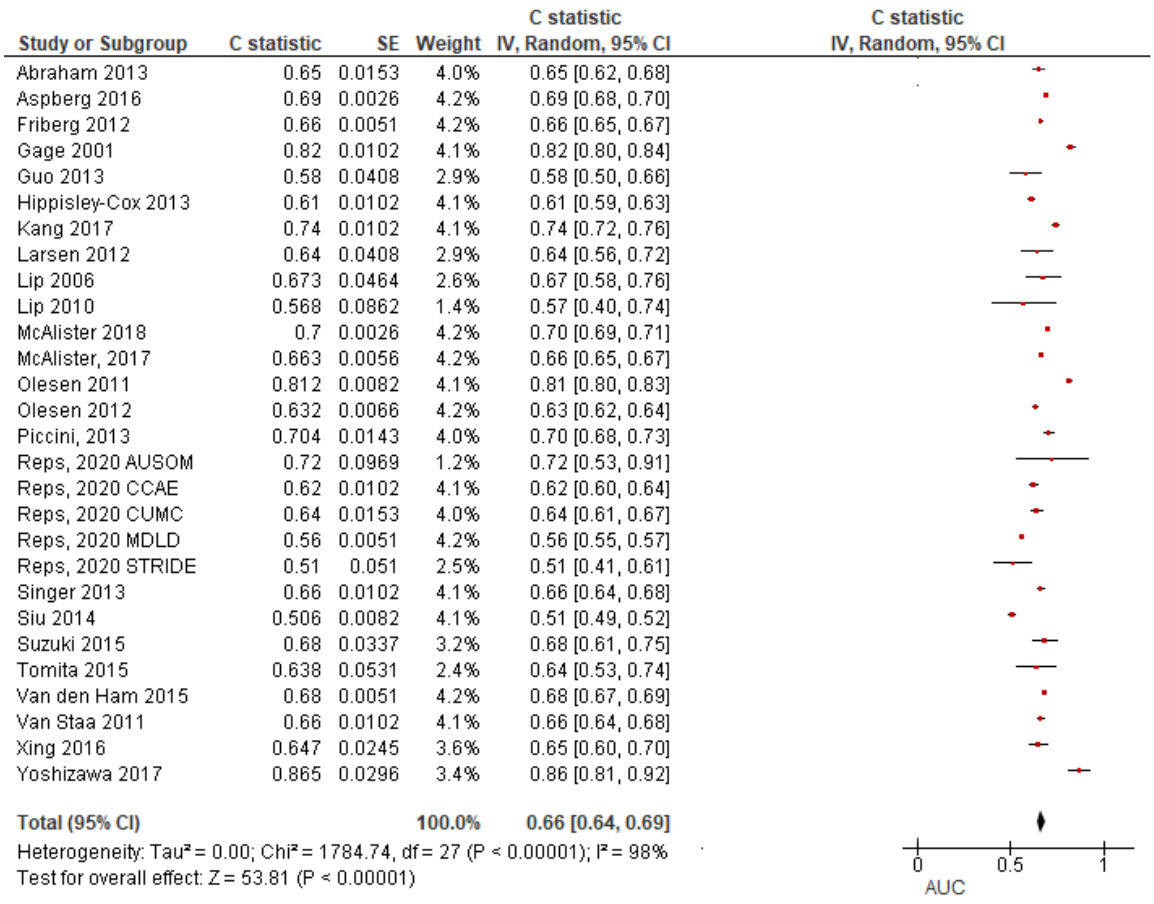


Figure 4: C statistic in Revised CHADS2

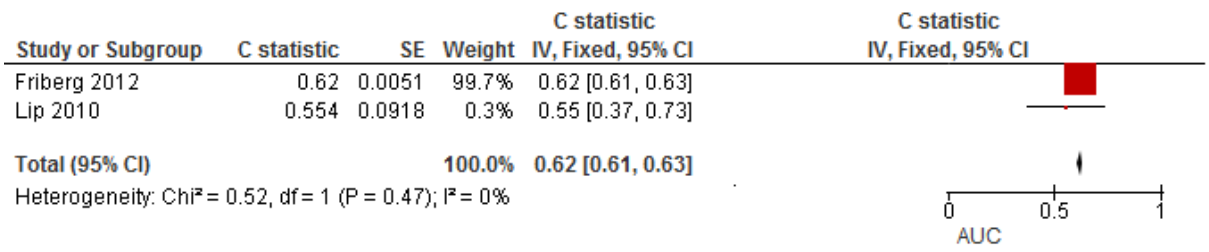


Figure 5: C statistic in R2CHADS2

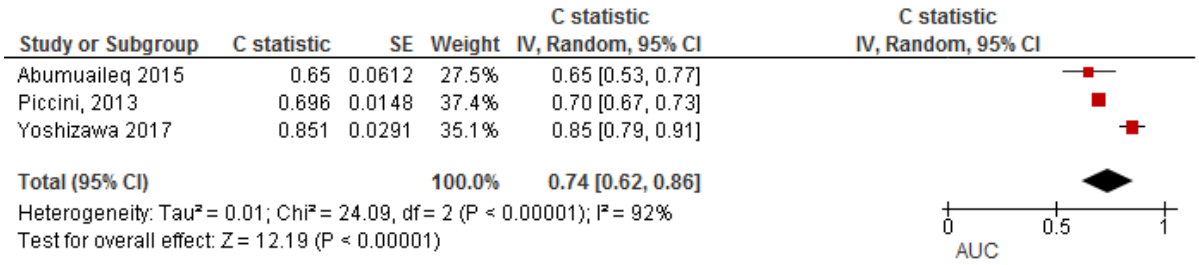


Figure 6: C statistic in CHADSVASC

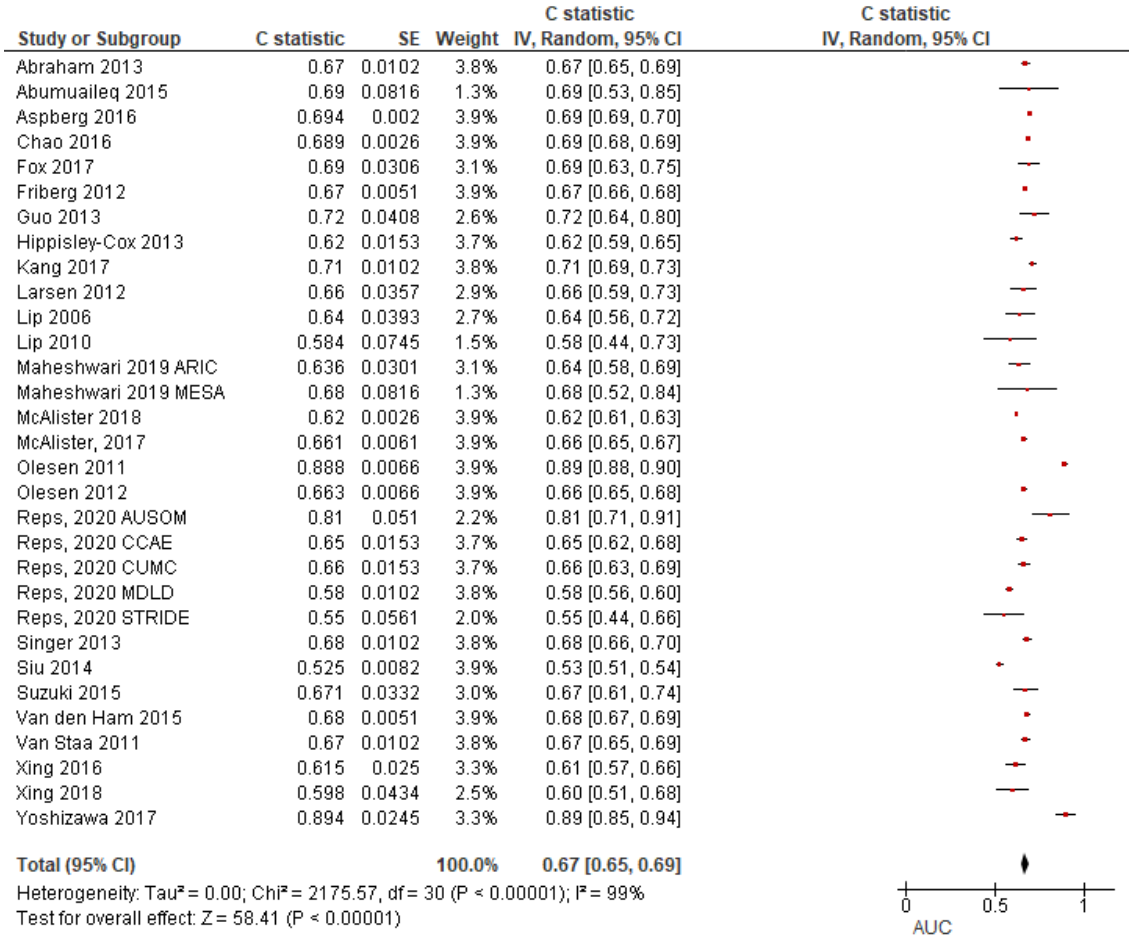


Figure 7: C statistic in P2-CHADSVASC

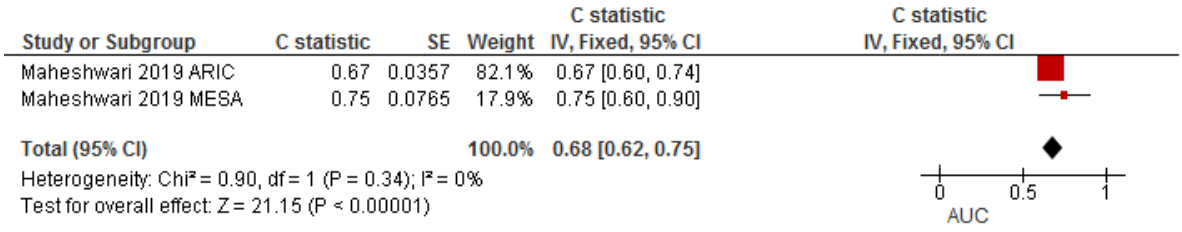


Figure 8: C statistic in ATRIA

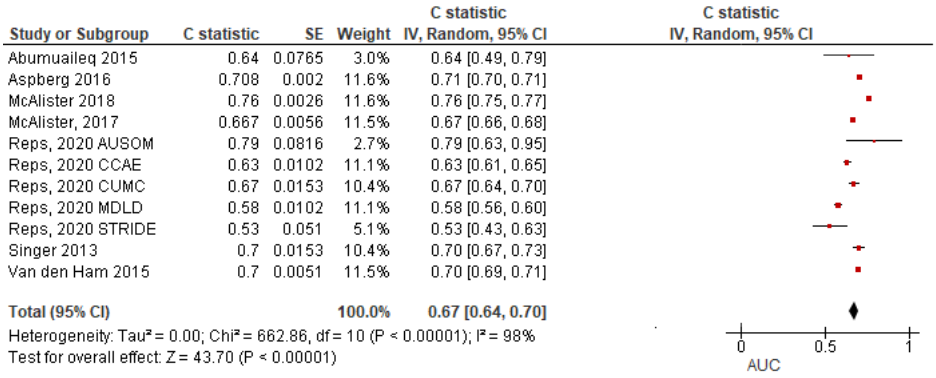


Figure 9: C statistic in AFI 1994

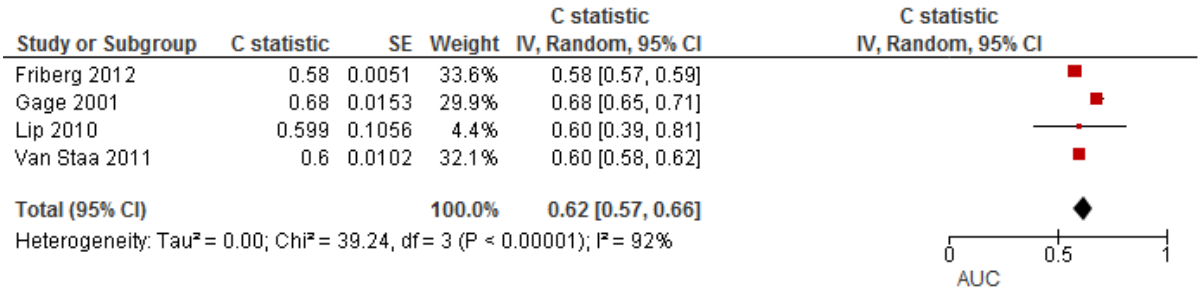


Figure 10: C statistic in SPAF 1995

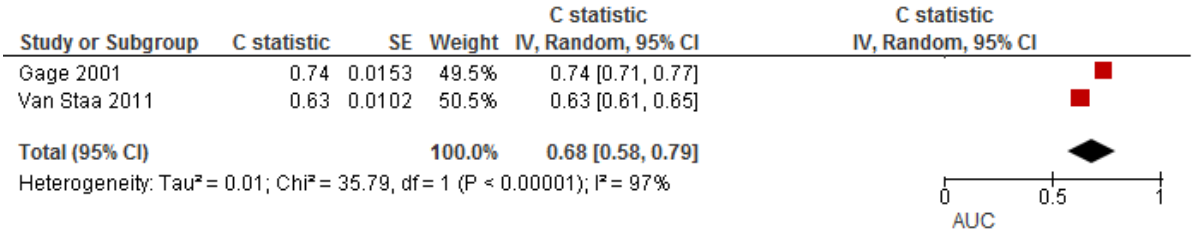


Figure 11: C statistic in SPAF 1999

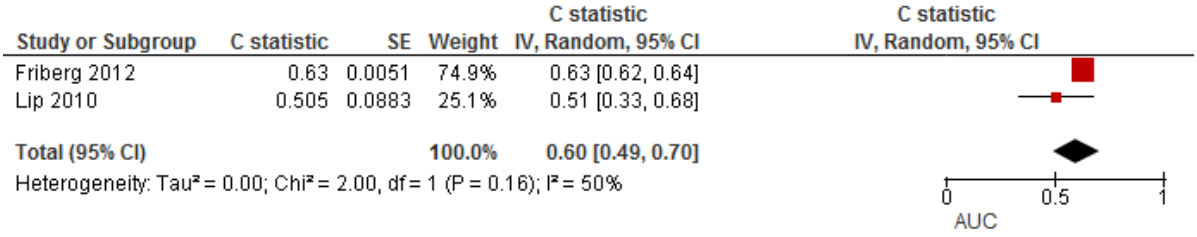


Figure 12: C statistic in FRAMINGHAM

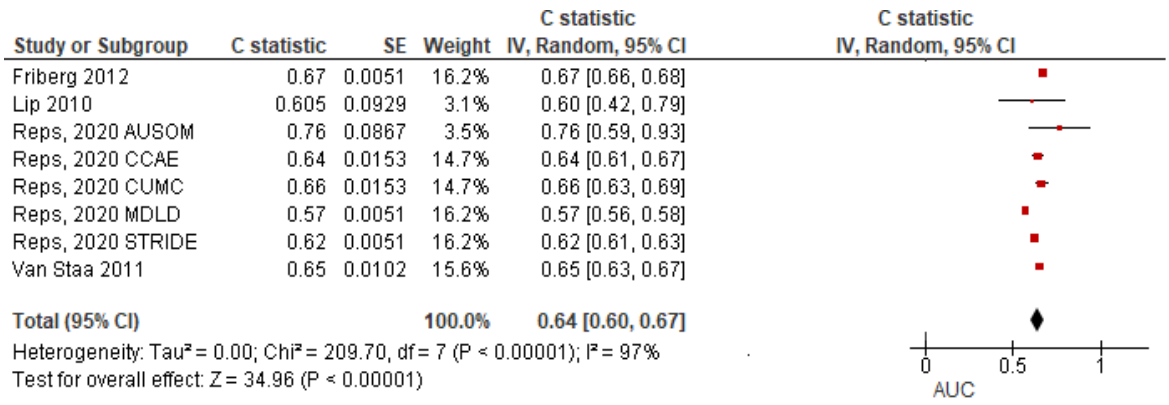


Figure 13: C statistic in Q STROKE

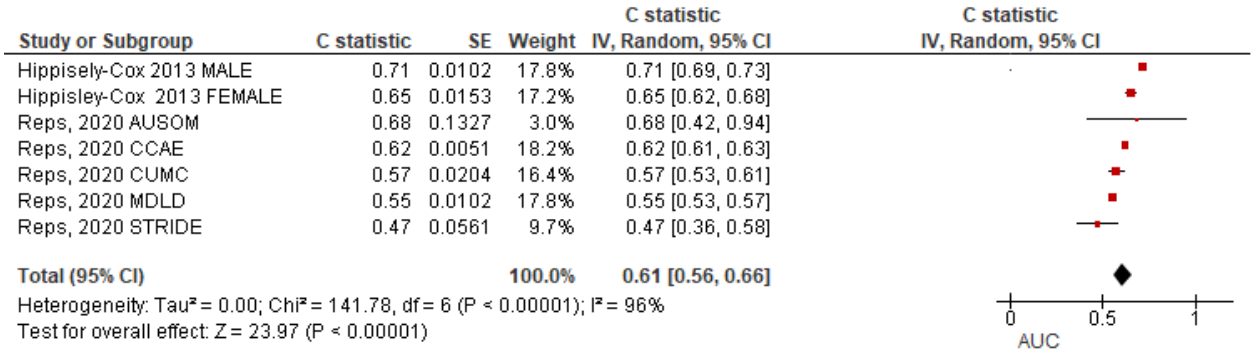


Figure 14: C statistic in ACCP 2008

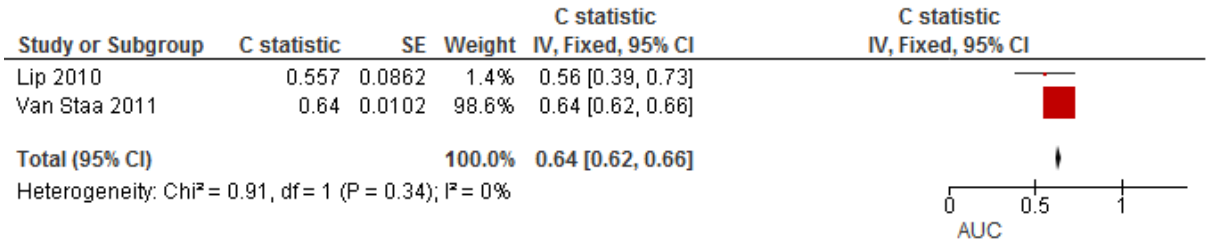


Figure 15: C statistic in ACH/AHA/ESC

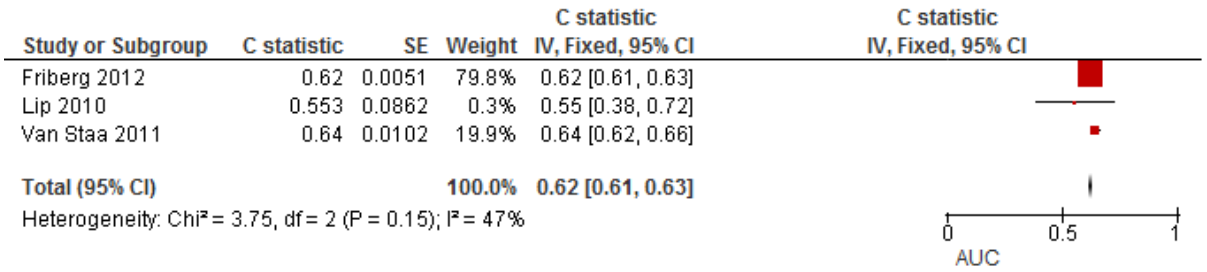
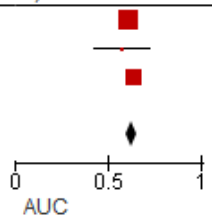


Figure 16: C statistic in NICE

Study or Subgroup	C statistic	SE	Weight	C statistic IV, Random, 95% CI
Friberg 2012	0.61	0.0051	53.2%	0.61 [0.60, 0.62]
Lip 2010	0.573	0.0765	3.1%	0.57 [0.42, 0.72]
Van Staa 2011	0.64	0.0102	43.7%	0.64 [0.62, 0.66]
Total (95% CI)			**100.0%**	**0.62 [0.59, 0.65]**

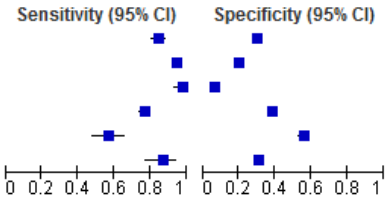
Heterogeneity: Tau² = 0.00; Chi² = 7.24, df = 2 (P = 0.03); I² = 72%

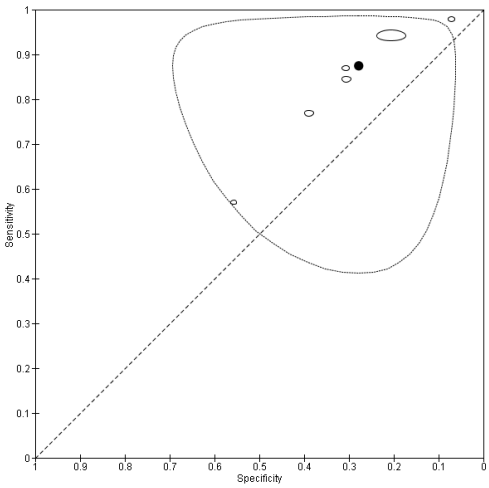


E.2 Sensitivity/specificity (pooled data only)

CHADS at threshold of ≥1

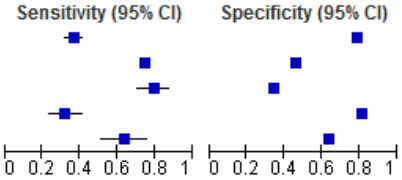
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abraham 2013	386	3835	71	1689	0.84 [0.81, 0.88]	0.31 [0.29, 0.32]
Aspberg 2016	10412	112082	641	29018	0.94 [0.94, 0.95]	0.21 [0.20, 0.21]
Gage 2001	92	1521	2	118	0.98 [0.93, 1.00]	0.07 [0.06, 0.09]
HippisleyCox 2013	684	4153	206	2646	0.77 [0.74, 0.80]	0.39 [0.38, 0.40]
Larsen 2012	73	653	55	822	0.57 [0.48, 0.66]	0.56 [0.53, 0.58]
Suzuki 2015	60	2438	9	1081	0.87 [0.77, 0.94]	0.31 [0.29, 0.32]

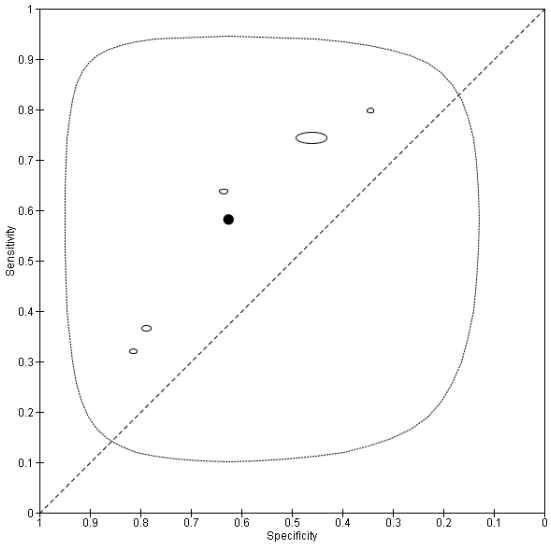




CHADS at threshold ≥2

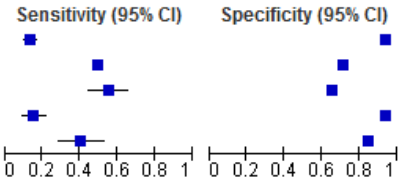
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abraham 2013	167	1175	290	4349	0.37 [0.32, 0.41]	0.79 [0.78, 0.80]
Aspberg 2016	8218	76067	2835	65033	0.74 [0.74, 0.75]	0.46 [0.46, 0.46]
Gage 2001	75	1075	19	564	0.80 [0.70, 0.87]	0.34 [0.32, 0.37]
Larsen 2012	41	274	87	1201	0.32 [0.24, 0.41]	0.81 [0.79, 0.83]
Suzuki 2015	44	1285	25	2234	0.64 [0.51, 0.75]	0.63 [0.62, 0.65]

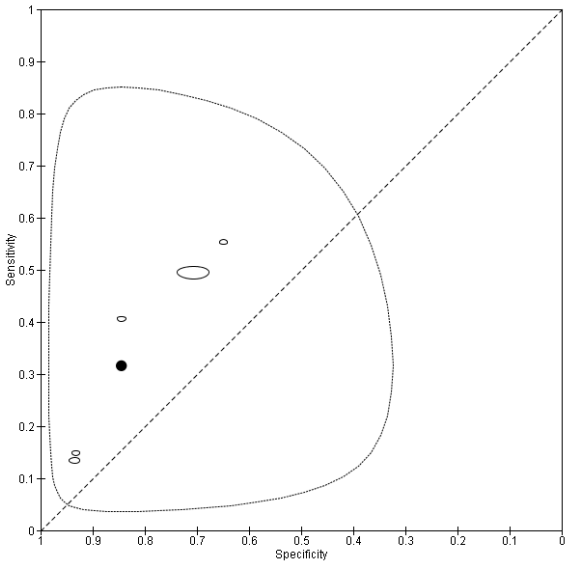




CHADS at threshold ≥3

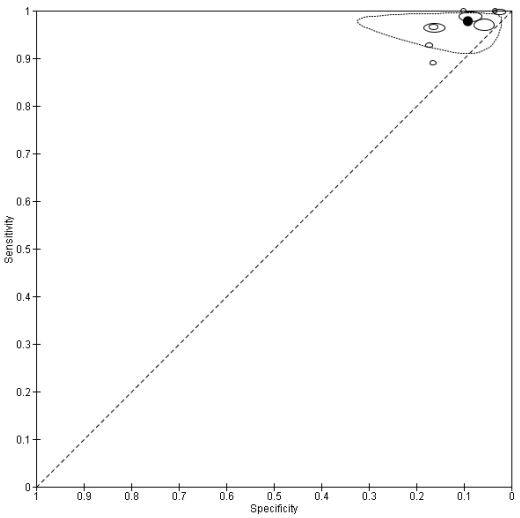
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abraham 2013	61	359	396	5165	0.13 [0.10, 0.17]	0.94 [0.93, 0.94]
Aspberg 2016	5468	41292	5585	99808	0.49 [0.49, 0.50]	0.71 [0.70, 0.71]
Gage 2001	52	575	42	1064	0.55 [0.45, 0.66]	0.65 [0.63, 0.67]
Larsen 2012	19	100	109	1375	0.15 [0.09, 0.22]	0.93 [0.92, 0.94]
Suzuki 2015	28	547	41	2972	0.41 [0.29, 0.53]	0.84 [0.83, 0.86]





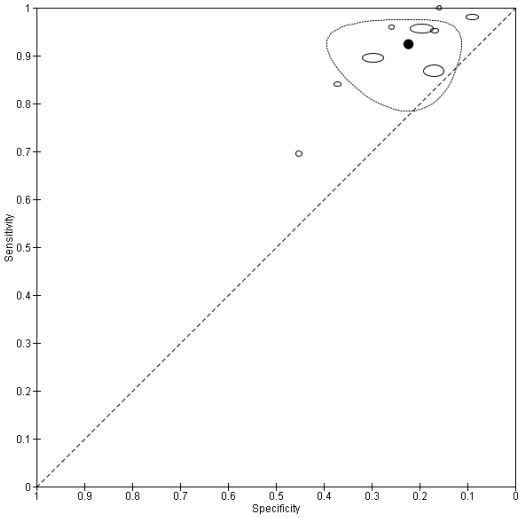
CHADSVASC at threshold ≥1

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abumaileq 2015	9	140	0	5	1.00 [0.66, 1.00]	0.03 [0.01, 0.08]
Aspberg 2016	10911	128976	142	12124	0.99 [0.98, 0.99]	0.09 [0.08, 0.09]
Chao 2016	20373	97411	635	5852	0.97 [0.97, 0.97]	0.06 [0.06, 0.06]
HippisleyCox 2013	860	5684	30	1115	0.97 [0.95, 0.98]	0.16 [0.16, 0.17]
Larsen 2012	114	1231	14	243	0.89 [0.82, 0.94]	0.16 [0.15, 0.18]
Lip 2010	25	919	0	103	1.00 [0.86, 1.00]	0.10 [0.08, 0.12]
Suzuki 2015	64	2907	5	612	0.93 [0.84, 0.98]	0.17 [0.16, 0.19]
Tomasdottir 2019	7145	94265	270	18278	0.96 [0.96, 0.97]	0.16 [0.16, 0.16]
Wicke 2019	1549	27959	4	717	1.00 [0.99, 1.00]	0.03 [0.02, 0.03]



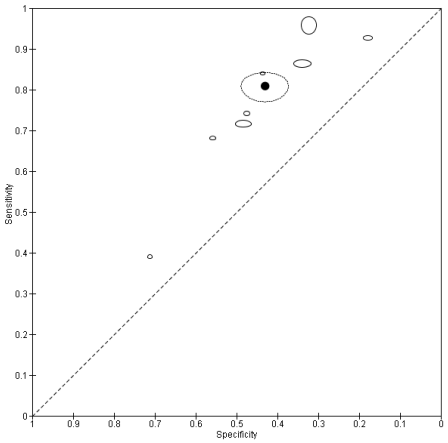
CHADSVASC at threshold ≥2

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abraham 2013	435	4595	22	929	0.95 [0.93, 0.97]	0.17 [0.16, 0.18]
Abumaileq 2015	9	122	0	23	1.00 [0.66, 1.00]	0.16 [0.10, 0.23]
Aspberg 2016	10574	113619	479	27481	0.96 [0.95, 0.96]	0.19 [0.19, 0.20]
Chao 2016	18235	85721	2773	17542	0.87 [0.86, 0.87]	0.17 [0.17, 0.17]
Larsen 2012	89	807	39	667	0.70 [0.61, 0.77]	0.45 [0.43, 0.48]
Lip 2010	24	758	1	264	0.96 [0.80, 1.00]	0.26 [0.23, 0.29]
Suzuki 2015	58	2211	11	1308	0.84 [0.73, 0.92]	0.37 [0.36, 0.39]
Tomasdottir 2019	6637	79164	778	33379	0.90 [0.89, 0.90]	0.30 [0.29, 0.30]
Wicke 2019	1523	26104	30	2572	0.98 [0.97, 0.99]	0.09 [0.09, 0.09]



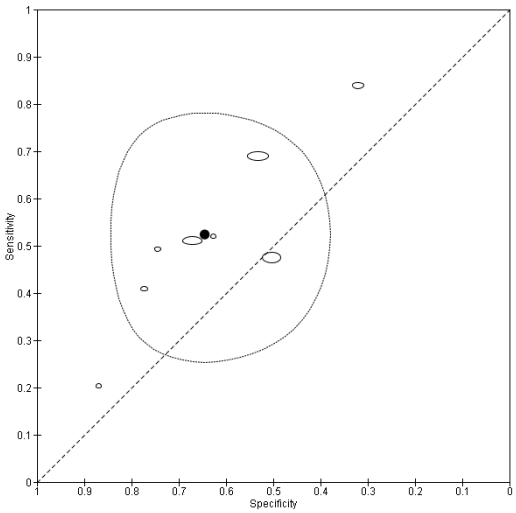
CHADSVASC at threshold ≥3

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abraham 2013	339	2897	118	2627	0.74 [0.70, 0.78]	0.48 [0.46, 0.49]
Aspberg 2016	9546	93184	1507	47916	0.86 [0.86, 0.87]	0.34 [0.34, 0.34]
Chao 2016	145676	69851	6441	33412	0.96 [0.96, 0.96]	0.32 [0.32, 0.33]
Larsen 2012	50	424	78	1050	0.39 [0.31, 0.48]	0.71 [0.69, 0.74]
Lip 2010	21	577	4	445	0.84 [0.64, 0.95]	0.44 [0.40, 0.47]
Suzuki 2015	47	1555	22	1964	0.68 [0.56, 0.79]	0.56 [0.54, 0.57]
Tomasdottir 2019	5311	58123	2104	54420	0.72 [0.71, 0.73]	0.48 [0.48, 0.49]
Wicke 2019	1440	23551	113	5125	0.93 [0.91, 0.94]	0.18 [0.17, 0.18]



CHADSVASC at threshold ≥4

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abraham 2013	187	1250	270	4274	0.41 [0.36, 0.46]	0.77 [0.76, 0.78]
Aspberg 2016	7619	65912	3434	75188	0.69 [0.68, 0.70]	0.53 [0.53, 0.54]
Chao 2016	9990	51314	11018	51949	0.48 [0.47, 0.48]	0.50 [0.50, 0.51]
Larsen 2012	26	192	102	1282	0.20 [0.14, 0.28]	0.87 [0.85, 0.89]
Lip 2010	13	382	12	640	0.52 [0.31, 0.72]	0.63 [0.60, 0.66]
Suzuki 2015	34	899	35	2620	0.49 [0.37, 0.62]	0.74 [0.73, 0.76]
Tomasdottir 2019	3787	37065	3628	75478	0.51 [0.50, 0.52]	0.67 [0.67, 0.67]
Wicke 2019	1303	19480	250	9196	0.84 [0.82, 0.86]	0.32 [0.32, 0.33]



E.3 NRI

Figure 17: ATRIA versus CHADS2

Study or Subgroup	NRI	SE	Weight	NRI IV, Random, 95% CI
Aspberg 2016	0.16	0.0102	25.9%	0.16 [0.14, 0.18]
McAlister, 2017	0.0082	0.0093	26.0%	0.01 [-0.01, 0.03]
Singer 2013	0.24	0.0357	22.2%	0.24 [0.17, 0.31]
Van den Ham 2015	0.137	0.0087	26.0%	0.14 [0.12, 0.15]
Total (95% CI)			**100.0%**	**0.13 [0.05, 0.22]**

Heterogeneity: Tau² = 0.01; Chi² = 165.13, df = 3 (P < 0.00001); I² = 98%
Test for overall effect: Z = 3.12 (P = 0.002)

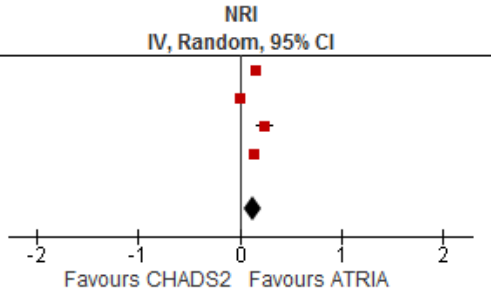


Figure 18: ATRIA versus CHADSVASC

Study or Subgroup	NRI	SE	Weight	NRI IV, Random, 95% CI
Aspberg 2016	0.21	0.0051	43.0%	0.21 [0.20, 0.22]
Singer 2013	0.25	0.0204	17.6%	0.25 [0.21, 0.29]
Van den Ham 2015	0.233	0.0071	39.4%	0.23 [0.22, 0.25]
Total (95% CI)			**100.0%**	**0.23 [0.20, 0.25]**

Heterogeneity: Tau² = 0.00; Chi² = 9.31, df = 2 (P = 0.010); I² = 79%
Test for overall effect: Z = 20.93 (P < 0.00001)

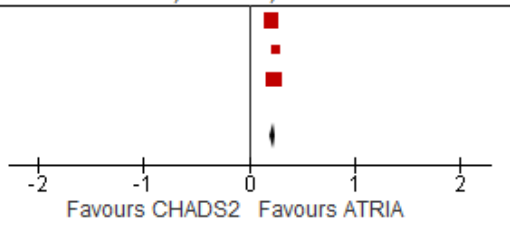


Figure 19: CHADS2 versus CHADSVASC

Study or Subgroup	NRI	SE	Weight	NRI IV, Random, 95% CI
Abraham 2013	-0.211	0.0658	6.8%	-0.21 [-0.34, -0.08]
Guo 2013	-0.166	0.0638	7.2%	-0.17 [-0.29, -0.04]
Kang 2017	0.017	0.0128	23.3%	0.02 [-0.01, 0.04]
Larsen 2012	0.03	0.0102	24.2%	0.03 [0.01, 0.05]
McAlister, 2017	0.0054	0.0081	24.7%	0.01 [-0.01, 0.02]
Olesen 2012	-0.142	0.0449	11.3%	-0.14 [-0.23, -0.05]
Xing 2016	0.237	0.1209	2.5%	0.24 [0.00, 0.47]
Total (95% CI)			**100.0%**	**-0.02 [-0.06, 0.02]**

Heterogeneity: Tau² = 0.00; Chi² = 38.34, df = 6 (P < 0.00001); I² = 84%
Test for overall effect: Z = 1.19 (P = 0.23)

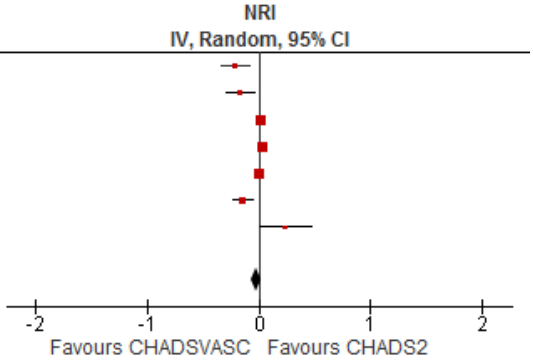
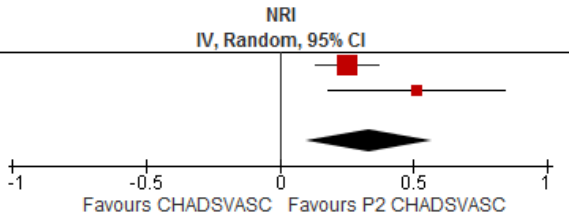


Figure 20: P2-CHADSVASC versus CHADSVASC

Study or Subgroup	NRI	SE	Weight	NRI IV, Random, 95% CI
Maheshwari 2019 ARIC	0.25	0.0612	68.2%	0.25 [0.13, 0.37]
Maheshwari 2019 MESA	0.51	0.1684	31.8%	0.51 [0.18, 0.84]
Total (95% CI)			**100.0%**	**0.33 [0.10, 0.57]**

Heterogeneity: Tau² = 0.02; Chi² = 2.11, df = 1 (P = 0.15); I² = 53%
Test for overall effect: Z = 2.75 (P = 0.006)



Appendix F: Clinical evidence tables

Table 21. Abraham, 2013[2]

Reference	Abraham, 2013[2]
Study type	Retrospective cohort analysis of risk prediction tools
Study sample	161,809 post-menopausal women aged 50-79 years were prospectively enrolled in the Women's Health Initiative (WHI) cohort. Events from 1993 through September 2010 were used for this retrospective analysis. The initial study population consisted of women who reported a history of atrial fibrillation or had an electrocardiogram with documented atrial fibrillation at baseline (n ¼ 7108). From this group, 291 were excluded with valvular heart disease or hyperthyroidism, 85 with missing values for either CHADS2 or CHA2DS2-VASc, and 790 on warfarin or other OACs at WHI randomization or enrolment. There were 1127 excluded, leaving a final sample of 5981, of which 2390 were participants in one of the clinical trials and 3591 were enrolled in the observational study; 5901 women with atrial fibrillation were identified by self-report, 24 by electrocardiogram, and 56 had both.
Inclusion criteria	Study participants were members of the Women's Health Initiative (WHI) cohort: a prospective, multiarm clinical trial and observational study that focused on the causes and prevention of cardiovascular disease, cancer, and osteoporosis in women
Exclusion criteria	Major exclusion criteria were predicted survival <3 years, alcohol or drug dependency, dementia, severe mental illness, and participation in another clinical trial, valvular heart disease, hyperthyroidism, warfarin or other OACs use.
Risk tools	CHADS2 and CHADSVASC
Outcome	Intensity of follow-up visits varied based on enrolment arm, ranging from every 6 months (clinical trials) to every 3 years (observational study). When a potential outcome was identified, medical records were obtained and stroke (including self-reports) and transient ischemic attack (only the first event) were centrally adjudicated. Up to 17 year follow up
Results	457 events CHA2DS2-VASc had a higher c statistic than CHADS2: 0.67 (95% CI,0.65-0.69) versus 0.65 (95% CI, 0.62-0.67), P <.01 When using CHA2DS2-VASc at 5-year follow-up, the NRI (vs CHADS2) was +0.211, P <.001.
Why the group were not anticoagulated	Not a low risk group as 457/5981 with an event at follow up. However the group were somewhat different to a group of warfarin or other OACs users, in terms of a lower risk of: CHF, prior stroke/TIA, and CABG.

Table 22. Abumuaileq, 2015[3]

Reference	Abumuaileq, 2015[3]
Study type	External validation
Study sample	154 consecutive patients with NVAF, and uncoagulated. All the consultations which were registered in the emergency department of a tertiary hospital between January 2008 and June 2010 enabled identification of all consecutive patients ≥18 years of age with AF documented by electrocardiographic records (n = 1873). After excluding patients with prosthetic valve (n=473), rheumatic heart disease (n = 46) and/or patients with active cancer (n = 61), there were 1293 patients with NVAF. After excluding patients on anticoagulation (n = 1135) and those patients lost to follow up (n = 4) there were 154 consecutive patients with NVAF. Mean age was 74 years, mean SBP was 129, 30% were current smokers, 21% had DM, 6.5% had HF, 15% CHD. 85% CHADSVASC score of 2 points or more
Inclusion criteria	Non-valvular AF
Exclusion criteria	Patients on anticoagulation, prosthetic valve, rheumatic heart disease, active cancer
Risk tools	CHA2DS2-VASc, R2CHADS2 and ATRIA
Outcome	9 TE events The primary endpoint for the present study was the development of TE event during follow-up. A TE complication was defined as the occurrence of ischemic stroke, TIA or peripheral embolism (including fatal TE events). Diagnosis of stroke or transient ischemic attack required an acute neurological deficit lasting for more or less than 24 h, respectively, which could not be explained by other causes and with at least 1 image test (computed tomography or magnetic resonance) compatible with the diagnosis, as well as confirmation from a neurologist. A diagnosis of peripheral embolism was defined as non-central nervous system embolism leading to an abrupt vascular insufficiency associated with clinical or radiographic evidence of arterial occlusion in absence of another mechanism such as atherosclerosis, instrumentation or trauma. 11 month follow up
Results	9 TE events at follow up The C statistics for each tool were as follows: CHADSVASC: 0.69 (0.53 – 0.85) R2CHADS2: 0.65 (0.53 – 0.78) ATRIA: 0.64 (0.49 – 0.80) At the conventional thresholds, CHADSVASC had 100% sensitivity
Why the group were not anticoagulated	The non-anticoagulated patients were almost all on antiplatelets, compared to a very small proportion of the anticoagulated patients. They also had a lower prevalence of HF, previous stroke and were more likely to be smokers.

Table 23. Aspberg, 2016[10]

Reference	Aspberg, 2016[10]
Study type	External validation
Study sample	115,153 participants with AF and no anticoagulant therapy. The SAF Cohort is based on information from two nationwide Swedish health care registers, the National Patient Register and the Prescribed Drug Register. The National Patient Register contains individual information on all hospitalizations and all visits to hospital outpatient clinics in Sweden since 1987. The cases of AF were identified in the National Patient Register. They were defined by ICD code 1409 with or without any of the specifying sub-codes A to F. Prior stroke 13%, Age 70.7% >65 years, 49.3% female, 15.8% DM, 28% HF, 6% Renal failure, 44% hypertension. 2005 to 2010
Inclusion criteria	All patients with a diagnosis of AF between 1 July 2005 and 31 December 2010 were included. Atrial fibrillation was defined by the ICD-10 code (I489 with or without any of the specifying sub codes A–F). Thus, both AF and atrial flutter were included.
Exclusion criteria	The analyses were restricted to patients who did not use anticoagulant therapy during the follow-up period. Patients who were taken care of in the primary care or in other open clinics not affiliated with a hospital during follow-up were not included.
Risk tools	ATRIA CHADS CHADSVASC
Outcome	Acute ischaemic stroke was the sole outcome event (defined by ICD-10 code I63), excluding TIAs or other kind of thromboembolism sometimes considered in previous studies. The outcome diagnosis, ischaemic stroke, was retrieved from the National Patient Register. A blanking period of 14 days after the index date was used to avoid including events that were registered twice or more due to transfer between hospitals, or reflecting events during the hospital stay possibly occurring prior to the AF diagnosis. The patients were censored at the date when the outcome event occurred, at the date of death, or at end of follow-up (31 December 2010). Follow up 5 years (maximum)
Results	11,053 strokes at follow up (3.25% per year) The total number of patients with a diagnosis of AF during the defined time period was 307 351. After exclusion of patients with mitral stenosis or valvular surgery (13 039) or death within 14 days from the index date (10 343), 283 969 patients remained. Further exclusion of patients given warfarin or other OACs therapy during the follow-up or having a diagnosis of ischaemic stroke within 2 weeks of inclusion, left 152 153 patients for analysis. These patients contributed 340 223 person-years of follow-up, with a mean follow-up time of 2.23 years. The total number of strokes observed during follow-up was 11 053 for an overall ischaemic stroke rate of 3.25%/year. C Index Using the entire point score range:

Reference	Aspberg, 2016[10]
	ATRIA: 0.708 (0.704–0.713) CHADS2: 0.690 (0.685–0.695) CHA2DS2-VASc: 0.694 (0.690–0.700). Using the categorical, published cut-points for low, moderate, and high ischaemic stroke risk ATRIA: 0.668 (0.664–0.672) CHADS2: 0.663 (0.658–0.668) CHADSVASC: 0.593 (0.591–0.595). However, the C-indices were quite similar when the cut-points in the categorical score were altered to better fit the Swedish cohort's ischaemic stroke rates. ATRIA then had a C-index of 0.633 (0.630–0.635), CHADS2 0.649 (0.646–0.653), and CHA2DS2-VASc 0.634 (0.631–0.637). Using published cut-points for the categorical scores, Net reclassification Improvement (NRI) favoured ATRIA: 0.16 (0.14–0.17) vs. CHADS2 and 0.21 (0.20–0.23) vs. CHA2DS2-VASc. These improvements resulted from predominant up-reclassification of the CHADS2 score (with up-reclassification of events outweighing up-reclassification of non-events) exclusive down-reclassification of the CHA2DS2-VASc score (with down-reclassification of non-events outweighing down-reclassification of events). Net reclassification improvement decreased to near zero when using the optimized cut-points, ATRIA -0.088 -0.022 to 0.0041) vs. CHADS2 and -0.00086 (-0.0094 to 0.0076) vs. CHA2DS2-VASc.
Why the group were not anticoagulated	Pre-warfarin or other OACs recommendations. No evidence of low risk or 'special' group.

Table 24. Chao, 2016[21]

Reference	Chao, 2016[21]
Study type	Retrospective cohort study
Study sample	124, 271 patients with AF (diagnosed using ICD-9-CM code from the National health Insurance Research database in Taiwan, who had not received warfarin or other OACs or any antiplatelet agents. Age 72, 54% male, 56.8% hypertensive, 23% DM, 38% CHF, 28% previous stroke/TIA. Median CHADSVASC score 3.

Reference	Chao, 2016[21]
Inclusion criteria	AF as defined above
Exclusion criteria	Warfarin or other OACs or any antiplatelet agents
Risk tools	CHADSVASC Age modified CHADSVASC (as original CHADSVASC, but modified by extending the first age criterion from 65-74 to 50-74)
Outcome	Ischaemic stroke, with concomitant imaging studies of the brain (CT/MRI) Follow up to 10 years
Results	21,0008 patients had events, for an annual risk of 3.9% C indexes for IS CHADSVASC: 0.689 (0.684-0.694) mCHADSVASC: 0.708 (0.703-0.712) DeLong test showed that there was a significant difference (p<0.0001) NRI mCHADSVASC v CHADSVASC: +0.039 (0.0216 to 0.0459), p<0.0001
Why the group were not anticoagulated	Unclear but clinical data suggested this was not a low risk or special group.

Table 25. Fang, 2008[30]

Reference	Fang, 2008[30]
Study type	Retrospective Cohort study
Study sample	The ATRIA (AnTicoagulation and Risk Factors In Atrial Fibrillation) study is a cohort of 10,932 adults with diagnosed non-valvular AF and who were not taking Warfarin or other OACs. The study sample relevant to this review were a sub-set of 5,588 patients who were known not to have used anticoagulants from baseline to a fixed follow up of 12 months. Sample data are not given for this sub-group, but the characteristics of the larger sub-group were 46% aged >75, 43% women, 8.3% with prior stroke, 50% with hypertension, 29% with HF and 16% with DM. 81.3% were at moderate or high risk of stroke.
Inclusion criteria	Patients with a diagnosis of AF between July 1, 1996, and December 31, 1997, found via automated inpatient, outpatient, and electrocardiographic databases. The cohort was followed up through September 2003, a median follow-up of 6.0 years (interquartile range 3.1 to 6.7 years).

Reference	Fang, 2008[30]
Exclusion criteria	Mitral stenosis, documented valvular repair or replacement, transient post-operative AF, or concurrent hyperthyroidism. Warfarin or other OACs exposure among patients was determined from computerized records from pharmacy, laboratory, and ambulatory visits. The analyses were restricted to the 10,932 patients who had periods of time when they appeared not to be taking warfarin or other OACs.
Risk tools	AFI 1994, SPAF 1995, CHADS2, Framingham and ACCP 2004
Outcome	Database searched for incident thromboembolic events, either ischemic stroke or other peripheral embolism. The validity of potential events was adjudicated by an outcomes committee of 3 physicians using a formal study protocol. If there was no consensus on the validity of an event, an expert neurologist adjudicated the event. Valid ischemic strokes were defined as neurological deficits of sudden onset that persisted for more than 24 h and were not explained by other etiologies. Valid non-stroke peripheral emboli were defined as emboli identified by radiographic imaging, intraoperative examination, or pathological findings, and without underlying atherosclerotic disease in the affected artery. Outcome events that occurred during hospitalization or as a complication from a diagnostic or interventional procedure were excluded. 6 year follow up
Results	685 TEs (643 ISs) C statistics for each tool: AFI 1994 0.61 SPAF 1995 0.65 CHADS2 0.67 Framingham 0.69 ACCP 2004 0.60
Why the group were not anticoagulated	Unclear if the non-anticoagulated sample were a special group. No details provided as to why they remained anticoagulated. The 685 events and % at moderate/high risk according to risk tools suggests not a low-risk group.

Table 26. Fox, 2017[33]

Reference	Fox, 2017[33]
Study type	Retrospective Cohort study
Study sample	2301 patients with AF that were not on OACs. These patients were part of a larger cohort of 10.132 patients enrolled on the UK-based ORBIT-AF registry. Details of the characteristics of these 2301 patients are not reported.
Inclusion criteria	People with incident or prevalent AF
Exclusion criteria	Not reported

Reference	Fox, 2017[33]
Risk tools	GARFIELD AF Risk CHADSVASC
Outcome	Stroke/SE defined as the combined end point of IS, SE and TIA. Follow up not reported
Results	Untreated cohort (n=2301) C statistics at 1 year (number of events =27) GARFIELD: 0.76(0.68-0.84) CHADSVASC: 0.67(0.61-0.77) C statistics at 3 years (number of events = 51) GARFIELD: 0.70(0.63-0.77) CHADSVASC: 0.69(0.63-0.76)
Why the group were not anticoagulated	Unclear.

Table 27. Friberg, 2012[36]

Reference	Friberg et al. 2012[36]
Study type	Retrospective cohort study.
Study sample	90, 490 patients with AF (defined by ICD-10 code 1489 with or without subscales A-F) identified from the Swedish National Discharge Registry. Demographic data stated to be in supplementary file but not available in that file.
Inclusion criteria	All individuals with a diagnosis of AF, between July 2005 and December 2008 who were known to not have used Warfarin or other OACs during the 1.4 year mean follow up.
Exclusion criteria	Silent AF and patients with AF taken care of in a primary care setting not affiliated to a hospital; valvular AF, mitral stenosis, valvular surgery.
Risk tools	CHADSVASC, CHADS2, SPAF 1999, ACC/AHA/ESC, Framingham, NICE
Outcome	First occurrence of Ischaemic stroke (defined by ICD-10 code 163). A blanking period of 14 days was also used, that excluded events occurring in first 14 days.
Results	7334 TE events; 5359 IS events C statistics, sensitivity, specificity and NRI for Ischaemic stroke:

Reference	Friberg et al. 2012[36]				
		C statistic (95% Cis)	sensitivity	specificity	NRI
	CHADSVASC (continuous)	0.67(0.66-0.68)	-	-	-
	CHADSVASC	0.56(0.56-0.57)	1	0.06	Reference
	CHADS2 (continuous)	0.66(0.66-0.67)	-	-	-
	Revised CHADS2	0.62(0.61-0.62)	0.98	0.15	0.07
	CHADS2	0.65(0.64-0.65)	0.98	0.15	0.07
	Framingham (cont)	0.67(0.66-0.68)	-	-	-
	Framingham	0.64(0.64-0.65)	0.92	0.26	0.12
	SPAF 1999	0.63(0.62-0.64)	0.89	0.29	0.12
	ACC/AHA/ESC 2006	0.62(0.61-0.62)	0.98	0.15	0.07
	NICE 2006	0.61(0.60-0.62)	1	0.09	0.00
	AFI 1994	0.58(0.58-0.59)	0.99	0.09	0.00
Why the group were not anticoagulated	Unclear. Limited demographic information but high number of events suggesting not low risk.				

Table 28. Gage, 2001[39]

Reference	Gage, 2001[39]
Study type	Retrospective cohort study
Study sample	1733 patients from the US National Registry of AF cohort. Mean age 81, 58% women, 56% CHF, 56% hypertension, 23% DM, 25% history of cerebral ischaemia. 1204 were not prescribed any antithrombotic therapy and 529 (31%) were prescribed aspirin. CHADS2 score of 2.1.
Inclusion criteria	Chronic or recurrent AF – confirmed by ECG or documentation.

Reference	Gage, 2001[39]
Exclusion criteria	Acute AF or death during hospitalisation
Risk tools	CHADS2 (created in this study by amalgamating the AFI and SPAF schemes), API, SPAF
Outcome	Hospitalisation for ischeamic stroke as determined by Medicare claims. ICD-9-CM codes used to identify. 1.2 year FU
Results	94 IS events (74 strokes) AFI 1994 0.68 (0.65 to 0.71) SPAF 1995: 0.74 (0.71 to 0.76) CHADS: 0.82 (0.80 to 0.84)
Why the group were not anticoagulated	Unclear but clinical data suggested this was not a low risk or special group.

Table 29. Gage, 2004[38]

Reference	Gage, 2004[38]
Study type	Retrospective cohort study
Study sample	2580 patients with nonvalvular AF who were on aspirin therapy (doses of 75 – 325mg/d) but not on warfarin or other OACs therapy [or on an ineffective dose of 1.25mg/d (n=171) or low dose of 2 mg/d (n=290)]. Data taken from 6 prospective RCTS. 37% women, mean age 72, 46% hypertension, 25% HF, 13% DM, 22% prior stroke or TIA, 18% prior MI/angina. 59% moderate or high risk.
Inclusion criteria	Nonvalvular AF (not defined) on aspirin therapy
Exclusion criteria	Participants included in any derivation cohorts
Risk tools	AFI 1994, SPAF, ACCP 2001, CHADS2, Framingham
Outcome	Suspected stroke, confirmed by CT in 98% of incident neurological events. Strokes defined as neurological deficits that persisted > 24 hours and not associated with an intracranial haemorrhage. Mean follow up 1.9 years
Results	207 IS events C statistics AFI 1994 0.63 (sd 0.01) SPAF 1995 0.64 (0.01) ACCP 2001 0.58 (0.01) CHADS2 0.70 (0.02)

Reference	Gage, 2004[38]
	Framingham 0.69 (0.02) If prior stroke excluded: AFI 0.61 (sd 0.02) SPAF 1995 0.61 (0.02) ACCP 0.58 (0.02) CHADS2 0.63 (0.03) Framingham 0.62 (0.03)
Why the group were not anticoagulated	Unclear

Table 30. Guo, 2013[41]

Reference	Guo, 2013[41]
Study type	Retrospective cohort study
Study sample	885 patients with pre-existing diagnosis of permanent, persistent or paroxysmal AF at General Hospital in China between 2007 and 2010. Not using Warfarin or other OACs, Mean age 77, 27% female, 75% hypertensive, 39% DM, 23% HF, 63% CAD, 20.9% prior stroke, renal failure 9.6%. 81.2% high risk on CHADSVASC.
Inclusion criteria	Development of new onset AF during admission (defined on ECG or Holter recording) and recorded as an ICD-10 code.
Exclusion criteria	Warfarin or other OACs
Risk tools	CHADS2 CHADSVASC
Outcome	Major adverse events (stroke/TE). IS defined as focal neurological deficit of sudden onset lasting >24 hours diagnosed clinically by a neurologist. A TE was IS, PE or peripheral embolism. Follow up mean 1.9 years
Results	55 IS, 2 PEs, 12 DVTs and 16 other STEs (Total 85 TE events) C statistic for TEs CHADS2: 0.58 (0.50 to 0.67) CHADSVASC: 0.72 (0.64 to 0.81)

Reference	Guo, 2013[41]
	NRI CHADSVASC v CHADS: +0.166 (0.039 to 0.291), p=0.009 IDI +0.011 (0.001 to 0.017)
Why the group were not anticoagulated	Unclear but clinical data suggested this was not a low risk or special group.

Table 31. Hippisley-Cox, 2013[49]

Reference	Hippisley-Cox, 2013[49]
Study type	Internal validation study as this was a joint derivation and validation study, conducted by the same researchers. However the pool of people for the validation study was quite distinct (see below).
Study sample	7689 people on 225 NHS database from GPs who had atrial fibrillation (not defined) at baseline. This was a different random group from the derivation cohort, the derivation cohort being based on 451 completely different NHS practices. Demographic data given for entire dataset, but not for the AF sub-set which the data in this extraction is based on. 71% classed as high risk on CHADS2.
Inclusion criteria	People aged 25-84 years at the study entry date, drawn from patients registered with eligible practices between 1 January 1998 and 1 Aug 2012; diagnosis of AF
Exclusion criteria	patients with a prior recorded diagnosis of stroke or transient ischaemic attack at baseline because of the difficulty of distinguishing a new stroke from a review of an existing stroke in GP records. patients without a Townsend deprivation score related to a valid postcode. patients who were taking anticoagulants (as defined by chapter 2.8.2 of the British National Formulary) at baseline Did not exclude patients prescribed aspirin at baseline as aspirin is generally not considered to be effective at preventing stroke in patients with atrial fibrillation. Incident users of anticoagulants during follow-up not excluded " in order to ensure the baseline population was representative of patients who might subsequently be prescribed anticoagulants".
Risk tools	QStroke (the paper also describes the methodology and results of the derivation of this tool, but not relevant to this review) CHADS CHADSVASC

Reference	Hippisley-Cox, 2013[49]
Outcome	First recorded diagnosis of either stroke or transient ischemic attacks, excluding haemorrhagic stroke. The Read codes used for case identification on the GP computer record were those agreed and used in the Quality and Outcomes Framework for General Practice. The ICD-10 codes used for case identification on the Office for National Statistics death certificate were cerebral infarction (I63) and stroke not specified as haemorrhage or infarction (I64). 10 year follow up
Results	Of the 7689 eligible patients, 890 had a stroke or TIA at follow-up. Sensitivity and specificity CHADS2 (score >2): sen 76.9%, spec: 38.9% CHADSVASC (score >2): sen 96.6%, spec: 16.4% Q STROKE (top 63%): sen 82.5%, spec: 39.5% Q STROKE (top 90%): sen 99.2%, spec: 11.2% Q STROKE (top 85%): sen 97.9%, spec: 16.7% Q STROKE (top 80%): sen 95.8%, spec: 22.1% Q STROKE (top 70%): sen 89.0%, spec: 32.5% Harrell's C statistic Female Q stroke (95% CIs): 0.65(0.62-0.67) CHADSVASC: 0.62(0.59-0.65) CHADS: 0.61(0.59-0.65) Male Q stroke (95% CIs): 0.71(0.69-0.73) CHADSVASC: 0.67(0.65-0.69) CHADS: 0.63(0.61-0.66) R2 Female Q stroke (95% CIs): 14(9.2-18.7) CHADSVASC: 9.6(5.5-13.8) CHADS: 9.1(4.9-13.2)

Reference	Hippisley-Cox, 2013[49]
	Male Q stroke (95% CIs): 24.1(19.3-28.9) CHADSVASC: 18.3(13.7-22.8) CHADS: 13.5(9.1-17.9) D statistic Female Q stroke (95% CIs): 0.82(0.66-0.99) CHADSVASC: 0.67(0.51-0.83) CHADS: 0.64(0.49-0.81) Male Q stroke (95% CIs): 1.15(1-1.3) CHADSVASC: 0.97(0.82-1.12) CHADS: 0.81(0.66-0.96) NRI Data related to reclassification were given but there were insufficient information on true events and non-events to allow calculation of the NRI (NRI results not provided in the paper)
Why the group were not anticoagulated	Unclear but appeared to be not low risk based on the proportion of people with strokes at follow up

Table 32. Kang, 2017[60]

Reference	Kang, 2017[60]
Study type	Retrospective cohort
Study sample	10,846 patients with newly diagnosed NVAF naïve to oral anticoagulants from the Korean National health Insurance Service national Sample Cohort. Mean age 63.7 years, 47% women, previous stroke 16.7%, CHF 25%, DM 21%, IHD 48%, CHADS more

Reference	Kang, 2017[60]
	than or equal to 4: 16%, CHADSVASC more than or equal to 6 10%. 30,138 person-years of follow up (mean follow up time: 2.8years)
Inclusion criteria	Non-valvular AF – defined as having AF is 1 or more AF diagnoses made during hospitalisation on 2 or more diagnoses made at outpatient clinics.
Exclusion criteria	Rheumatic mitral stenosis, mechanical or bioprosthetic hearts valve, mitral valve repair. Any AF diagnosis during first year following inception of the database to ensure washout period of >1 year Any patients prescribed OACs within 1 month after initial diagnosis of AF (aim was to establish accuracy of tools in people not having OACs at all)
Risk tools	CHADSVASC, CHADS2
Outcome	Ischeamic stroke. Stroke was defined according to ICD-10 codes (I63-64) for diagnoses made during hospitalization and according to brain imaging such as computed tomography and magnetic resonance imaging. Patients were censored when they were prescribed oral vitamin K antagonists 1.17 years mean follow up
Results	888 events in 29,466 person-years at risk The 2 scoring systems were shown to be useful in discriminating the risk of ischemic stroke C statistic, 0.74; 95% confidence intervals [CI]: 0.72–0.75 for CHADS2; 0.71; 95% CI:0.69-0.73, for CHA2DS2-VASc; Harrell's c-index, 0.79 for CHADS2 and 0.78 for CHA2DS2-VASc. The CHA2DS2-VASc score had a lower NRI than the CHADS2 score −1.7%; 95% CI: −4.2 to 0%; P=0.03.
Why the group were not anticoagulated	Unclear – stated as a limitation of study that reasons for prescribing OAC were not identified.

Table 33. Kim, 2017[66]

Reference	Kim, 2017[66]
Study type	Retrospective cohort

Reference	Kim, 2017[66]
Study sample	5855 OAC naïve AF patients identified from the Korea NHIS sample cohort database from 2002 to 2008. Mean age 64, 48% women, CHADSVASC means core 3.28, 24.5% prior stroke, 13% MI, 32% HF, 76% hypertension, 20% DM.
Inclusion criteria	Patients with at least 1 in-patient or 2 out-patient diagnoses of AF.
Exclusion criteria	Valvular AF; patients receiving OACs at baseline; <20 years
Risk tools	CHADS2, CHADSVASC, ATRIA
Outcome	The primary end point was incident ischemic stroke (including ischemic stroke–related death) during the 5 years of follow-up period (from January 2009 to December 2013). Any diagnosis of ischemic stroke with concomitant brain imaging studies, including computed tomography or MRI, was defined as incident ischemic stroke. Mean 4.21 years follow up.
Results	819 strokes CHADS sen 85.7, spec 46.8 CHADSVASC sen 98.8, spec 16.9 ATRIA (0-5) sen 74.8, spec 61 ATRIA (0) sen 99.4, spec 8.2 No C statistics given.
Why the group were not anticoagulated	Unclear

Table 34. Larsen, 2012[71]

Reference	Larsen, 2012[71]
Study type	Retrospective cohort study
Study sample	1603 non-anticoagulated patients with incident AF (defined by ICD-08 [pre 1994] or ICD-10 codes) from a Danish cohort of 57,053 middle aged people. Age 67, 40% women, mean follow up 5.4 years, CHF 24.4%, 30% hypertension, 10% DM, 6% stroke history. 7% CHADS2 of 5 or above, 6% CHADSVASC score of 5 or above.
Inclusion criteria	The study population was defined as incident cases of atrial fibrillation after recruitment who had not emigrated before being diagnosed with atrial fibrillation.

Reference	Larsen, 2012[71]
Exclusion criteria	Cases diagnosed simultaneously with stroke, thromboembolism, and transient ischemic attack or patients who died on the same day they were diagnosed with atrial fibrillation were excluded for analysis. Based on the Danish prescription registry, all atrial fibrillation patients having had prescriptions of anticoagulant agents, warfarin or other OACs, or phenprocoumon (ATC code B01AA) within 180 days to the outcome event or end of follow-up were excluded.
Risk tools	CHADS2 and CHADSVASC
Outcome	Stroke (not defined) 5.4 year FU
Results	1.9 strokes per 100 person years At mean 5.4 year follow up, C statistics: CHADS2: 0.64 (0.56 – 0.71) CHADSVASC 0.66 (0.59 – 0.72) At 1 year follow up, C statistics: CHADS2: 0.68 (0.59 – 0.76) CHADSVASC 0.69 (0.60 – 0.77) At 5 year follow up, NRI: CHADSVASC vs CHADS2: -3% (-6% to -1%)
Why the group were not anticoagulated	Unclear but cohort were not clearly low risk (56% had CHADSVASC score of 2 or more at baseline).

Table 35. Lip, 2006[74]

Reference	Lip, 2006[74]
Study type	Retrospective cohort study of data from the RCT SPAF III study
Study sample	994 patients with NVAF, not on adjusted dose warfarin or other OACs therapy (all on aspirin, or aspirin plus low dose 'inefficacious' warfarin or other OACs). Mean age 69.3, 75% male, 53% hypertension, 14% diabetes, 19% recent HF, 13% previous TIA/stroke, 10% previous MI, 6% PVD, 9% LV systolic dysfunction, 8% current smokers. 43 IS events. 73.9% not low risk according to CHADS2.

Reference	Lip, 2006[74]
Inclusion criteria	NVAF (not defined in paper)
Exclusion criteria	Patients randomised to adjusted-dose warfarin or other OACs
Risk tools	CHADS2, Birmingham, CHADS2 with vWF (Plasma von Willebrand Factor Levels) incorporated into the scale, Birmingham with vWF incorporated in to the scale
Outcome	Ischaemic Stroke (not defined) 1.6 year mean FU
Results	2.32% IS rate C statistics for IS Birmingham: 0.640 (0.563 to 0.713) CHADS2: 0.673 (0.582 to 0.754) Birmingham with vWF: 0.679 (0.591 to 0.756) CHADS2 with vWF: 0.691 (0.600 to 0.772)
Why the group were not anticoagulated	Unclear but clinical data suggested this was not a low risk or special group.

Table 36. Lip, 2010[75]

Reference	Lip, 2010[75]
Study type	Retrospective cohort study
Study sample	1084 patients without mitral stenosis or previous heart surgery and who did not use VKA or heparin at discharge of qualifying visit and for whom TE outcome was known at 1 year (results from a sub-group analysis of those without ANY use of VKA/heparin during follow up are also given below but exact size and characteristics of this sub-group are not given. However from the %s given the n was around 850). Age 66 years, 41% women, previous stroke 4.2%, TIA 4.3%, DM 17.3%, hypertension 67%, HF 23.5%, antiplatelets 74%, LVEF 53%. 17% classed as high risk and 61.9% as intermediate risk on CHADS2
Inclusion criteria	>18 years, ECG/Holter evidence of AF.
Exclusion criteria	Mitral stenosis, previous heart surgery, use of VKAs or Heparin at discharge.
Risk tools	AFI 1994, SPAF 1999, CHADS2 (2001), CHADS2 Revised, Framingham, NICE, ACCA/AHA/ESC, ACCP 2008 and Birmingham
Outcome	Thromboembolic events: IS (focal neurological event lasting >24 hours diagnosed by neurologist), PE or peripheral embolism
Results	C statistic data below are for those known to have been free from warfarin or other OACs throughout follow up (n=850 approx.)

Reference	Lip, 2010[75]
	AFI 1994: 0.599(0.392-0.726) SPAF 1999: 0.505(0.332-0.677) CHADS2: 0.568(0.399-0.737) Revised CHADS2: 0.554(0.374-0.734) Framingham: 0.605(0.423-0.787) NICE: 0.573(0.423-0.723) ACC/AHA/ESC: 0.553(0.384-0.722) ACCP 2008: 0.557(0.388-0.725) Birmingham: 0.584(0.438-0.731) Data for those that were just VKA free at baseline were qualitatively very similar (not shown here)
Why the group were not anticoagulated	Unclear but the cohort are not low risk.

Table 37. Lip, 2014[76]

Reference	Lip, 2014[76]
Study type	Retrospective cohort study
Study sample	3,483 patients with AF (n=242 had valvular AF) who were not receiving OACs. Mean age 70, 43% female, 48% HF, 33% CAD, 17% previous MI, 5% previous CABG, 40% hypertensive, 7% previous stroke, 9% renal insufficiency. Mean CHADSVASC score 3.1.
Inclusion criteria	Patients given a diagnosis of NVAF or atrial flutter between 2000 and 2010 at Cardiology department in France.
Exclusion criteria	OACs
Risk tools	SAMe-TT2R2 score
Outcome	Stroke/ TEs (not defined) Up to 10 years follow up
Results	273 stroke/TE events Harrel C statistic for stroke/TEs SAMe-TT2R2 score (cont): 0.509 (0.492 to 0.526) SAMe-TT2R2 score (3 cats): 0.514 (0.497 to 0.531)

Reference	Lip, 2014[76]
	SAMe-TT2R2 score (2 cats): 0.530 (0.513 to 0.547)
Why the group were not anticoagulated	Unclear but clinical data suggested this was not a low risk or special group.

Table 38. Maheshwari, 2019[81]

Reference	Maheshwari, 2019[81]
Study type	Retrospective cohort study
Study sample	2229 participants from the ARIC study (Atherosclerosis Risk in Communities) and 700 participants from MESA (Multi-Ethnic Study of Atherosclerosis) with incident AF who were not on anticoagulants within 1 year of AF diagnosis; ARIC cohort: age 73; female 47%; DM 30%; hypertension 75%; previous MI 24%; HF 38%; PAD 9%; past stroke/TIA 15%; CHADSVASC 3.6; black 19%, white 81%; MESA cohort: age 76; female 45%; DM 18%; hypertension 68%; previous MI 6%; HF 8%; PAD 2%; past stroke/TIA 6%; CHADSVASC 3.0; black 20%, white 49%; Chines 13%; Hispanic 17%
Inclusion criteria	People enrolled on the ARIC study from 1987-2013 and the MESA study from 2000-2014
Exclusion criteria	From the ARIC study participants with missing ECG data (n=242), missing P-wave indices at baseline (n=45), prevalent AF (n=37), and those who were not white or black from all study sites and non-white from Minneapolis and Washington County (because of the small sample size; n=103) were excluded, resulting in a baseline cohort of 15 365 participants. We then identified 2625 cases of incident AF after the baseline study visit. Because of the potential bias introduced by anticoagulant use when studying stroke risk, participants with anticoagulant use within 1 year of AF diagnosis (n=172) were excluded. We also excluded those without follow-up beyond AF date (n=224), resulting in a final cohort of 2229 participants with incident AF. From the MESA study, participants with prevalent AF (n=66) or missing ECG or P-wave indices at baseline (n=49) were excluded, and we identified 876 cases of incident AF. We then excluded those without follow-up beyond the date of AF diagnosis (n=117), oral anticoagulant use within 1 year of AF diagnosis (n=54), and those with invalid P-wave axis measurements (n=5), resulting in a final cohort of 700 participants with incident AF.
Risk tools	CHADSVASC P2-CHADSVASC
Outcome	1 year ischaemic stroke
Results	ARIC data Number of ischaemic strokes: 47 at 1 year; 163 at 5 years

Reference	Maheshwari, 2019[81]
	C statistic CHADSVASC 0.60(0.51-0.69) (1 yr) CHADSVASC 0.636 (0.577-0.695) 5 yrs (in online supplement of paper) P2-CHADSVASC 0.67(0.60-0.75) (1 yr) NRI (P2-CHADSVASC v CHASDSVASC at 1 yr) +0.25(0.13-0.86) MESA data Number of ischaemic strokes: 31 at 3.3yrs **C statistic** CHADSVASC 0.68(0.52-0.84) (1 yr) P2-CHADSVASC 0.75(0.60-0.91) (1 yr) NRI (P2-CHADSVASC v CHASDSVASC at 1 yr) +0.51(0.18-0.86)
Why the group were not anticoagulated	Unclear

Table 39. McAlister, 2017[83]

Reference	McAlister, 2017[83]
Study type	Retrospective cohort study
Study sample	58,451 people from Alberta Canada with incident NVAF, and no anticoagulant use. eGFR < 60 24.4%; previous stroke 10.8%; previous bleed 11.2%; age >65 52.6%; female 47%; previous MI: 11.3%; HF: 21.8%; DM: 21.6%; PVD: 3.5%; hypertensive: 64.1%
Inclusion criteria	AF defined by ICD ninth revision clinical modification code 427.3 and ICD 10th revision code I48 in any fields of the Alberta health administrative databases;
Exclusion criteria	History of aortic or mitral valve disease; valve surgery; end stage renal disease; AF incident in previous 5 years; anticoagulation started in the first 3 months after index NVAF diagnosis.
Risk tools	CHADS2, CHADSVASC, R2CHADS2 (71 point), ATRIA, CHADS2KDIGO, CHADS2Alb, CHADS2 eGFR
Outcome	Stroke/TE – not defined

Reference	McAlister, 2017[83]					
	Mean FU: 2.5 years					
Results	7,340 patients had TES.					
	Tool	Sen	Spec	C statistic	NRI	
	CHADS2	0.83	0.524	0.663(0.652-0.657	Reference	
	CHADSVASC	0.825	0.496	0.661(0.649-0.672)	-0.0054(-0.0213 to 0.0105)	
	R2CHADS2 (71 point)	0.80	0.511	0.656(0.644-0.667)	-0.0150(-0.0363 to 0.0063)	
	ATRIA	0.811	0.524	0.667(0.656-0.679)	+0.0082(-0.0100 to 0.0264)	
	CHADS2 KDIGO	0.726	0.575	0.650 (0.638-0.663)	-0.0255(-0.0491 to -0.0019)	
	CHADS2 Alb	0.821	0.488	0.654(0.643-0.666)	-0.0178(-0.0256 to 0.0282)	
	CHADS2 eGFR	0.693	0.640	0.666(0.653-0.680)	0.0062(-0.0171 to 0.0295)	
Why the group were not anticoagulated	Unclear but clinical characteristics and rate of events (12.6%) suggest cohort was neither low risk nor 'special'.					

Table 40. McAlister, 2018[84]

Reference	McAlister, 2018[84]
Study type	Prospective cohorts study

Reference	McAlister, 2018[84]
Study sample	This was a sample of people (of an unknown size) with AF (defined as: ICD-9CM 427.3 or ICD-10CA I48) and who were not treated with OACs. No details are given of their characteristics. They were drawn from a larger cohort of 147,952 adult Canadians with AF.
Inclusion criteria	AF
Exclusion criteria	None reported
Risk tools	CHADS2 CHADSVASC ATRIA
Outcome	First TE: First stroke, TIA or systemic arterial TE
Results	C statistics at 1 year for first TE (newly diagnosed [incident]) CHADS2: 0.73(0.72-0.73) CHADSVASC: 0.64(0.64-0.64) ATRIA: 0.78(0.78-0.79) C statistics at 1 year for first TE (prevalent patients) CHADS2: 0.70(0.70-0.70) CHADSVASC: 0.62(0.62-0.62) ATRIA: 0.76(0.75-0.76)
Why the group were not anticoagulated	Unclear

Table 41. Olesen, 2011[96]

Reference	Olesen, 2011[96]
Study type	Retrospective cohort study
Study sample	73,538 people with NVAF who did not receive VKA or heparin. This cohort had almost identical stroke risk scores to others on VKA/heparin (CHADSVASC of 2 or more was 80.5%, comparing to 80.6% for another group with VKA prescription). Age >75 60%, female 51%, DM 9%, previous TE 18%, Vascular disease 18%, antiplatelets 35%. Follow up to 10 years
Inclusion criteria	NVAF or atrial flutter (defined by ICD codes ICD-8 [pre 1994] and ICD-10)

Reference	Olesen, 2011[96]
Exclusion criteria	VKA or heparin; death or TE in 7 days after baseline; no mitral or aortic valve disease or surgery. Note however that at 10 years 15,344 (20%) had received at least 1 prescription for Warfarin or other OACs, but an unknown sensitivity analysis showed this did not change results.
Risk tools	CHADS2 and CHADSVASC
Outcome	Admission to hospital, or death, from TE (defined by codes I26,63,64 and 74).
Results	Number of events not provided. C statistics at 1 year for TE: CHADS2: 0.711 CHADSVASC: 0.850 C statistics at 5 years: CHADS2: 0.796 CHADSVASC: 0.880 C statistics at 10 years: CHADS2: 0.812 CHADSVASC: 0.880
Why the group were not anticoagulated	Prior to routine VKA – not a low risk group.

Table 42. Olesen, 2012[95]

Reference	Olesen, 2012[95]
Study type	Retrospective cohort study
Study sample	2002 people aged <65 years with NVAF or atrial flutter. Age 54.9, 39% HF, 11% DM, 5% previous stroke, 17% vascular disease, 28.5% female. 38% scored >2 on CHADSVASC. Of these, 924 were not on OACs (results below are only for these), but no demographic data for these provided.
Inclusion criteria	NVAF

Reference	Olesen, 2012[95]
Exclusion criteria	OACs
Risk tools	CHADSVASC CHADS2 CHADS2 with vascular disease added
Outcome	Stroke and thromboembolism (from documentation) Follow up to 10 years
Results	14 events of TE No accuracy data for CHADSVASC provided. NRI for CHADS2 with vascular disease vs CHADS2 +0.4 (0 to 0.8) IDI +0.031, with an area under the ROC improvement of 0.046 (p<0.001)
Why the group were not anticoagulated	Unclear if this was not a low risk or special group.

Table 43. Olesen 2012b[97]

Reference	Olesen 2012b[97]
Study type	Retrospective cohort study
Study sample	47,576 patients with atrial fibrillation (defined by ICD code I48 from Danish National Patient Registry), not on OACs. Mean age 69.4, CHF 2%, hypertension 17%, DM 2%, previous stroke 0%, vascular disease 12%, female 46.3%, aspirin 26%. 63% CHADSVASC score of 2 or more. All had CHADS2 scores of 0 or 1.
Inclusion criteria	NVAF patients
Exclusion criteria	OACs
Risk tools	CHADSVASC CHADS2
Outcome	Hospitalisation or death from stroke/TE. ICD codes ICD-10: G458, G459, I63,I64,I74) 12 year follow up period.
Results	At 12 years there were 4599 events

Reference	Olesen 2012b[97]
	C statistics (12 years) CHADS2: 0.632 (0.619-0.646) CHADSVASC: 0.663 (0.650-0.676) NRI (1 year) CHADSVASC v CHADS2: +0.142, p<0.001 IDI (1 year) IDI was 0.003 and the area under the receiver operating curve was improved by 0.042 (p<0.001)
Why the group were not anticoagulated	This appears to have been a lower risk group than normal, based on the baseline figures.

Table 44. Piccini, 2013[100]

Reference	Piccini, 2013 [100]
Study type	External validation retrospective cohort study
Study sample	Sub-group from the ATRIA cohort that were NOT taking OACS (n=16,360). No information given on characteristics in Piccini, 2013.
Inclusion criteria	NVAF patients
Exclusion criteria	OACs
Risk tools	CHADS2 R2CHADS2 score – CHADS2 with creatinine clearance incorporated (2 points for CrCl <60mL/min) Sum of CrCl<60 ml and prior stroke/TIA
Outcome	Stroke – a composite of all stroke (both ischemic and haemorrhagic) and systemic embolism. Stroke was defined as a new, sudden focal neurological deficit resulting from a presumed cerebrovascular cause that persisted beyond 24 hours and was not attributable to other identifiable causes such as tumour or seizure. Events that involved symptoms that lasted <24 hours were considered TIAs. Brain imaging was sought in each case to distinguish haemorrhagic from ischemic stroke. Non-CNS systemic embolism was defined as abrupt vascular insufficiency associated with clinical or radiographic evidence of arterial occlusion in the absence of another likely mechanism (e.g., atherosclerosis instrumentation, or trauma). In the presence of atherosclerotic peripheral arterial disease, diagnosis of embolism required angiographic demonstration of abrupt arterial occlusion. All suspected primary events were adjudicated by an independent clinical end-point committee that included a stroke neurologist.
Results	**[Sub-group NOT taking OACS only]** **C statistics**

Reference	Piccini, 2013 [100]
	R2CHADS: 0.696 (0.667-0.726) CHADS2: 0.74 (0.676-0.732) Sum of CrCl<60 ml and prior stroke/TIA: 0.625(0.594-0.656) **NRI (vs CHADS2)** R2CHADS: 0.226(0.125 to 0.307) **NRI (vs R2CHADS)** Sum of CrCl<60 ml and prior stroke/TIA:-0.024 (-0.077 to + 0.029)
Why the group were not anticoagulated	Unclear.

Table 45. Singer, 2013[123]

Reference	Singer, 2013[123]
Study type	Derivation and internal/external validation study of the ATRIA scheme
Study sample	Validation cohort: 25, 306 patients with NVAF contributing 26, 263 person-years of follow up off warfarin or other OACs (mean follow up 1 year). TE rate of 1.9% per year (496 stroke or other TE events). Baseline data is only given for the overall (% patient-years) but likely that the validation cohort were similar: female 43%, HF 26%. Hypertension 56%, CAD 29%, PAD 3%, DM 17%, eGFR <60: 35.8%
Inclusion criteria	AF confirmed by ECG or physician diagnosis in the medical record (>1 inpatient or >2 outpatient), aged >21. Included also people with mitral stenosis and a history of valve replacement in mitral or aortic positions (1.5% of external validation cohort)
Exclusion criteria	Warfarin or other OACs.
Risk tools	ATRIA, CHADS2, CHADSVASC
Outcome	IS, defined as sudden onset of a neurologic deficit lasting >24 hours and not attributable to other causes. Other TEs: sudden occlusion to an artery to a major organ documented by imaging, surgery or pathology and not due to concomitant atherosclerosis or other causes. Mean FU 1 year
Results	496 TEs C index for stroke/ other TE ATRIA: 0.70 (0.67 to 0.72) (bootstrapped) CHADS2: 0.66 (0.64 to 0.69)

Reference	Singer, 2013[123]
	CHADSVASC: 0.68 (0.66 to 0.70) NRI Atria v CHADS2: 0.24(0.17-0.31) Atria v CHADSVASC: 0.25(0.21-0.30)
Why the group were not anticoagulated	Unclear but clinical data suggested this was not a low risk or special group.

Table 46. Schwartz, 2019[122]

Reference	Schwartz, 2019[122]
Study type	Retrospective cohort study
Study sample	Data from 11,443 patients with AF who were NOT on DOACs or VKAs were retrieved from the Northwestern Healthcare system's Enterprise Database Warehouse. The data allowed identification of stroke outcomes, and calculation of prior CHADSVASC scores. Mean age 67.6 for white patients and 63.1 for non-white patients. Mean CHADSVASC was 2.4 in whites and 2.2 in non-whites
Inclusion criteria	AF patients with no history of stroke; No use of VKAs or DOACs
Exclusion criteria	Patients with missing admission date, unknown race, prescription for dual-antiplatelet agents, and creatine clearance <30 ml/min
Risk tools used	CHADSVASC
Outcome definition	Incident Stroke using ICD-9 codes and ICD-10 codes
Results	**CHADSVASC** **Follow up 971 days post diagnosis; number of stroke events: 205** C statistic ('whites'): 0.681 (0.640-0.721) C statistic ('non-whites'): 0.646(0.572-0.720) Accuracy (derived from table 2 in the paper, summating the data in 'whites' and 'non-whites' to produce the overall accuracy figures Threshold of >1, sensitivity 0.8293, spec 0.3931 (TP 170, TN 35, FP 6820, TN 4418). Threshold of >2, sensitivity 0.649, spec 0.6127 (TP 133, TN 72, FP 4352, TN 6885). Threshold of >3, sensitivity 0.3902, spec 0.7987 (TP 80, TN 125, FP 2262, TN 8976).

Reference	Schwartz, 2019[122]
Why the group were not anticoagulated	**Not reported**

Table 47. Siu, 2014[124]

Reference	Siu, 2014[124]
Study type	Retrospective cohort study
Study sample	3881 patients with NVAF (not defined) who did not receive OACs. Mean age 77, 53.5% female, 47.5% hypertensive, 18% DM, 1.7% renal failure on dialysis, 19% HF, 8% CAD, 1.3% PAD, 17% prior stroke/TIA. Mean CHADSVASC 3.3.
Inclusion criteria	Non valvular AF
Exclusion criteria	Significant valvular heart disease, previous valvular surgery.
Risk tools	CHADS2 CHADSVASC
Outcome	Mean 3.19 year follow up Stroke (not defined)
Results	847 strokes during follow up. C statistics for stroke CHADS2: 0.506 (0.490-0.522) CHADSVASC: 0.525 (0.509-0.541) CHADSVASC sensitivity of 0.98 at cut-off of 1.
Why the group were not anticoagulated	Unclear but clinical data suggested this was not a low risk or special group.

Table 48. Suzuki, 2015[127]

Reference	Suzuki, 2015[127]
Study type	Retrospective cohort study
Study sample	3588 patients with AF without anticoagulation. Taken from 3 Japanese databases. Age 68.1, 34% female, 50% hypertension, 15% DM, 8.5% previous stroke or TIA, 15% HF, 11% CAD, 42% antiplatelet use. No data on CHADSVASC scores at baseline

Reference	Suzuki, 2015[127]
Inclusion criteria	AF patients (confirmed by 12 lead ECG or Holter monitoring) on one of 3 Japanese patient registries.
Exclusion criteria	Anticoagulation at time of registration.
Risk tools	CHADS2 and CHADSVASC
Outcome	Ischeamic stroke (not defined) Average 1.4 years follow up
Results	69 strokes in follow-up period (in 5.188 person-years) CHADS2: 0.680 (0.614 – 0.746) CHADSVASC: 0.671 (0.606 – 0.736)
Why the group were not anticoagulated	Unclear but data on baseline CHADSVASC score not given. However incidence rate of stroke was 1.3%, about half the expected value, suggesting a lower than expected level of risk.

Table 49. Tomasdottir, 2019[130]

Reference	Tomita, 2015[130]			
Study type	Retrospective cohort study			
Study sample	231 077 (48.1% women) non-selected patients with AF not receiving oral anticoagulation from 2006 to 2014. Data from cross-linked national Swedish registers. Age 75 (men), 82 (women); HF 28.5%; hypertension 48.4%; DM 17.2%; Stroke/TIA/SE 18.7%; Vascular disease 24.1%			
Inclusion criteria	All patients with an AF diagnosis registered between 2 December 2005 and 31 December 2014			
Exclusion criteria	Using OACs within 6 months of start of study (if during follow up were censored at that point); < 18 yrs; mitral stenosis or prosthetic heart valve			
Risk tools	CHADSVASC			
Outcome	Ischaemic stroke at mean follow up of 2.5 years			
Results	**Sensitivity and specificity of CHADSVASC at different thresholds (calculated from data in figure 3 in paper)**			
	Women			
	threshold	**sensitivity**	**specificity**	
	≥1	1	0	
	≥2	0.984296	0.083649	

Reference	Tomita, 2015[130]			
	≥3	0.946173	0.177238	
	≥4	0.785679	0.361596	
	≥5	0.546864	0.592322	
	≥6	0.328691	0.773967	
	≥7	0.142617	0.903489	
	≥8	0.039802	0.969533	
	≥9	0.007309	0.994277	
	Men			
	threshold	**sensitivity**	**specificity**	
	≥1	0.963587	0.162409	
	≥2	0.895078	0.296589	
	≥3	0.716251	0.483549	
	≥4	0.510722	0.670659	
	≥5	0.306001	0.820655	
	≥6	0.141065	0.921506	
	≥7	0.04356	0.973948	
	≥8	0.007687	0.994429	
	No other predictive data reported			
Why the group were not anticoagulated	Unclear			

Table 50. Tomita, 2015[131]

Reference	Tomita, 2015[131]
Study type	Retrospective cohort study
Study sample	294 women and 703 men with NVAF and no warfarin or other OACs treatment. Mean mCHADSVASC scores of 1.9 (male) and 3.3 (female). , Mean age 687% history of stroke/TIA, 58% antiplatelet use, 29% paroxysmal AF. 2 year follow up. 5 lost to FU
Inclusion criteria	AF (not defined)
Exclusion criteria	OACs
Risk tools	mCHADSVASC excluding female sex from the scheme = mCHADSVA mCHADSVASC CHADS2 Note: the m refers to the fact that these did not include PAD.
Outcome	TE events – not defined
Results	30 IS events C statistic CHADS2: 0.638 (0.534-0.730) mCHADSVASC: 0.595 (0.504 – 0.680) mCHADSVA: 0.624 (0.531-0.709) NRI CHADS2 versus mCHADSVASC -0.1(-0.28 to 0.08) CHADS2 versus mCHADSVA -0.03 (-0.21 to 0.16) mCHADSVASC versus mCHADSVA +0.11(0.001 to 0.20)
Why the group were not anticoagulated	Unclear but clinical data suggested this was not a low risk or special group.

Table 51. Van den Ham, 2015[133]

Reference	Van den Ham, 2015[133]
Study type	Retrospective cohort study

Reference	Van den Ham, 2015[133]
Study sample	60, 594 patients with NVAF untreated with warfarin or other OACs. Mean age 74.4 years, female 48.7%, 50% past or present smokers; 12% DM, 17.5% CHF, 54.6% hypertension, 15% previous stroke/TIA, 31% vascular disease, 28% renal dysfunction (eGFR <60 ml/min/1.73m2). mean follow up time was 2.1 years
Inclusion criteria	People with a first AF diagnosis (not defined) aged 18 years or older
Exclusion criteria	Rheumatic mitral stenosis, prosthetic heart valve; use of anticoagulants
Risk tools	ATRIA, CHADSVASC and CHADS2
Outcome	Ischeamic stroke (defined by codes in CPRD, HES or both) Mean 2.1 year follow up
Results	3751 IS events in follow up period of 125,296 person-years The C statistics for the continuous risk scores were 0.70 (95% CI:0.69to0.71) for the ATRIA risk score, 0.68 (95% CI: 0.67 to 0.69) for the CHADS2,and 0.68 (95% CI: 0.67 to 0.69) for the CHA2DS2-VASc risk score. The categorical risk scores, using the published low/moderate/high risk cut-offs resulted in C statistics of 0.66 (95% CI: 0.66 to 0.67) for the ATRIA, 0.65 (95% CI: 0.64 to 0.66) for the CHADS2, and 0.59 (95% CI: 0.59 to 0.60) for the CHA2DS2-VASc risk score. The NRI was 0.137 (95% CI: 0.120 to 0.153) or 0.233 (95% CI: 0.219 to 0.248) when using the ATRIA versus the CHADS2 or CHA2DS2-VASc risk scores, respectively. These improvements resulted mainly from downward reclassification from the CHADS2 score and entirely from downward reclassification from the CHADSVASC score.
Why the group were not anticoagulated	Unclear. Annualised stroke rate was 3% indicating these were not low risk patients.

Table 52. Van Staa, 2011[136]

Reference	Van Staa, 2011[136]
Study type	Retrospective cohort study

Reference	Van Staa, 2011[136]
Study sample	79,884 patients with AF (documented record). Age 73.3, female 49.7%, 54.6% current or past smoker, CHADS score more than or equal to 3: 20%, CHF 29%, DM 17%, Hypertension 50%, previous stroke or TIA 18%.
Inclusion criteria	AF aged >18 in the General practice Research Database, up to warfarin or other OACs inception or INR monitoring at a mean of 2.4 years; incident and prevalent AF
Exclusion criteria	Rheumatic valve disease
Risk tools	15 covered: see below
Outcome	Stroke as recorded in the GPRD, hospitalisation for stroke as recorded in the HES, and mortality resulting from stroke as recorded on death certificates.
Results	79,884 strokes recorded C statistics for stroke (GP recorded or registry) AFI 1994: 0.60(0.58-0.61) AFI 1998: 0.61(0.60-0.62) ACCP 2001: 0.62(0.60-0.62) ACCP2004: 0.61(0.60-0.62) NICE 2006: 0.64(0.62-0.65) ACC/AHA/ESC 2006: 0.64(0.62-0.66) ACCP 2008: 0.64(0.62-0.65) CHADSVASC (3 cats): 0.60(0.59-0.61) CHADSVASC (risk score): 0.67(0.65-0.69) CHADS2 (3 cats): 0.65(0.63-0.67) CHADS2 (risk score): 0.66(0.64-0.68) Mod CHADS2 (3 cats): 0.63(0.61-0.65) Mod CHADS2 (risk score): 0.69(0.67-0.71) SPAF 1995: 0.63(0.61-0.65) Hart 1999: 0.62(0.60-0.64) Van Walraven 2002: 0.55(0.54-0.58) Van Latum1995: 0.57(0.55-0.59) Framingham 2003 (3 cats): 0.62(0.60-0.64) Framingham 2003 (3 cats): 0.65(0.63-0.68)

Reference	Van Staa, 2011[136]
Why the group were not anticoagulated	Followed up to point of anticoagulation (2.4 years). Appears to be a cohort with normal levels of stroke risk based on CHADS score.

Table 53. Wang, 2003[137]

Reference	Wang, 2003[137]
Study type	Developmental study with internal validation using bootstrapping.
Study sample	705 participants with new onset AF (on ECG or based on hospital charts or physician office records) with no OAC treatment at baseline. Mean follow up of 4 years.
Inclusion criteria	Mean age 75, 48% women, SBP: 146, hypertension therapy 50%, DM 15%, smoking 18%, prior CHF or MI 34%, prior CVA/TIA 14%.
Exclusion criteria	AF prior to the first Framingham examination in the offspring cohort (n=1) or prior to 1960 in the original cohort (n=23); missing covariate data; stroke/TIA or death within 30 days of AF diagnosis; rheumatic mitral stenosis.
Risk tools	Framingham, CHADS2, SPAF 1995, AFI 1994
Outcome	Stroke – decided by a panel of 3 Framingham investigators, including a neurologist, based on a review of all medical records and clinical data, and an examination by the neurologist.
Results	83 strokes recorded C statistics for stroke Framingham:0.66 (sd=0.03) [Internal validation using bootstrapping samples] CHADS2: 0.62 SPAF 1995: 0.62 AFI 1994: 0.61 Calibration (for Framingham only) Ranking participants into quintiles according to their stroke-risk score yielded predicted 5-year stroke rates of7%(lowest quintile), 10%, 14%, 20%, and 33% (highest quintile). These predicted rates corresponded closely with actual 5-year stroke rates in each quintile: 8%, 9%, 13%, 20%, and 29%. The stroke-risk score and stroke or death–risk score had Hosmer-Lemeshow statistics of 7.6 and 6.5, respectively; values of 20 or less indicate good calibration.
Why the group were not anticoagulated	Unclear but clinical data suggested this was not a low risk or special group.

Table 54. Wicke, 2019[138]

Reference	Wicke, 2019[138]			
Study type	Retrospective cohort study			
Study sample	A broadly representative population with AF who were not on OACs from southern Germany (n=30,299). Claims data from a statutory health insurance (AOK Baden Wuerttemberg), the largest insurance fund in the German state of Baden-Wuerttemberg (population in 2014 was 10.7 million), were used. For the year 2014, the data contained information on 3.8 million individuals, which equals to about 35% of the state's population. Age 76.4; 46.6% male; CHADSVASC score 4.25; hypertension 85%; CHF 40.2%; stroke/TIA 7.96%; DM 10.1%;			
Inclusion criteria	All patients aged 18 years or older with a diagnosis of AF recorded in 2014. To increase diagnostic specificity, outpatient diagnoses of AF were required to be coded in at least two quarters of the year 2014. For hospital diagnoses, only one coding was required. Not on OACs.			
Exclusion criteria	Coded rheumatic mitral valve disease or artificial heart valves and those that died in 2014. On OACs in 2014 – identified based on ATC codes of prescription data.			
Risk tools	CHADSVASC (calculated via the ICD-10 codings on the data for 2014)			
Outcome	All hospitalisations for ischaemic stroke (ICD-10 code I63) recorded on the database in 2015 and 2016. This has been downgraded for indirectness as this will have a lower prevalence than any ischaemic stroke			
Results	961 hospitalisations due to stroke experienced by the 30,299 patients during the 2 year follow up. C statistic: 0.608			
	Threshold	**sensitivity**	**specificity**	
	≥1	0.998959	0.0246	
	≥2	0.98231	0.088322	
	≥3	0.933403	0.17678	
	≥4	0.844953	0.317651	
	≥5	0.621228	0.528769	
	≥6	0.368366	0.752255	
	≥7	0.16025	0.910209	
	≥8	0.048907	0.977142	
	≥9	0.014583	0.993975	

Reference	Wicke, 2019[138]
Why the group were not anticoagulated	Unclear but clinical data suggested this was not a low risk or a group at higher than average risk of bleeding

Table 55. Xing, 2016[139]

Reference	Xing, 2016[139]
Study type	Retrospective cohort study from 2011 to 2013.
Study sample	413 patients with NVAF, and not on oral anticoagulants for previous 6 months. mean age 81, 71% male, median CHADSVASC score 4.77. Hypertension 77.5%, previous stroke/TIA 36.8%, DM 36.1%, antiplatelets 68%. . Mean follow up 2 years.
Inclusion criteria	NVAF (diagnosed by 12 lead ECG or Holter), aged >65
Exclusion criteria	Oral anticoagulants in past 6 months, valvular AF, rheumatic mitral stenosis, mechanical or bioprosthetic heart valves, mitral valve repair, haemodialysis.
Risk tools	CHADS, CHADSVASC
Outcome	Ischeamic stroke – new sudden focal neurological deficit resulting from a supposed CV cause that persisted >24 hours and not attributable to other causes. Brain imaging also used to differentiate from haemorrhage. 2 years FU
Results	59 developed IS/TE C statistics CHADS2: 0.647 (0.599 – 0.693) CHADSVASC: 0.615 (0.566 – 0.662) De Long's test showed that CHADS2 was significantly better (NRI 0.237, p=0.0498)
Why the group were not anticoagulated	Unclear but not a high risk group, and no evidence that this cohort was specifically different in terms of other factors.

Table 56. Xing, 2018[140]

Reference	Xing, 2018[140]
Study type	Retrospective cohort study
Study sample	389 consecutive patients with AF (may overlap with Xing 2016). Age 83.7, 77% female, 82% hypertension, 56% vascular disease, 36% DM, 36% previous IS, 25% HF, Cr 100 mg/dL, EF 62%, CHADSVASC 4.87. Mean follow up of 2.57 years. 49 IS/TE events
Inclusion criteria	AF diagnosed by EMG, Holter monitoring or history.
Exclusion criteria	Mechanical prosthetic heart valves, PE, recent DVT and intraventricular thrombus. OACs in previous 3 months.
Risk tools	CHADSVASC
Outcome	Ischeamic stroke – new sudden focal neurological deficit resulting from a supposed CV cause that persisted >24 hours and not attributable to other causes. Brain imaging also used to differentiate from haemorrhage. 2.57 years
Results	49 IS/TE events C statistic at follow up for IS/TE: CHADSVASC: 0.598 (0.513 – 0.683)
Why the group were not anticoagulated	Not clear but appear not to be low risk. High level of previous strokes.

Table 57. Yoshizawa, 2017[142]

Reference	Yoshizawa, 2017[142] (same study as Komatsu 2014[68], except that Yoshizawa additionally contains results for R2CHADS as well as CHADS2 and CHADSVASC)
Study type	Retrospective cohort study
Study sample	332 consecutive cases in people with paroxysmal or permanent AF (confirmed by ECG) who were not receiving anticoagulant therapy, without cardiac valvular disease estimated by TTE. Patients on rhythm control therapy. Patients not receiving OACs because this was prior to guidelines promoting their use. Followed up for mean 53 months (but up to 120 months). Age 65, male/female: 224:108, hypertension 43%, DM 13%, smoking 27%, underlying heart disease 20% (IHD 11.4%, non-ischeamic 8.6%), 18 month Hx of AF, 33% on aspirin, 0% on warfarin or other OACs. CHADSVASC score 2 points or more: 59%.
Inclusion criteria	See above
Exclusion criteria	The study excluded patients with the following conditions: severe bradyarrhythmia (sick sinus syndrome, atrioventricular block, or intraventricular conduction defect); hepatorenal dysfunction; women in whom pregnancy was likely; or patients receiving warfarin or other OACs anticoagulation therapy.
Risk tools	R2CHADS, CHADS VASC and CHADS2

Reference	Yoshizawa, 2017[142] (same study as Komatsu 2014[68], except that Yoshizawa additionally contains results for R2CHADS as well as CHADS2 and CHADSVASC)
Outcome	IS/STE. Cerebral TE confirmed based on clinical symptoms and the presence of a 3mm or larger infarct area on CT/MRI. Mean 53 months FU
Results	2.1% rate of IS/TE per year C statistic R2CHADS: 0.851(0.794-0.908 CHADS2: 0.866(0.807-0.925) CHADSVASC: 0.894(0.846-0.951)
Why the group were not anticoagulated	Historical reasons. Not low risk as most (59%) had CHADSVASC scores of 2 points or more.

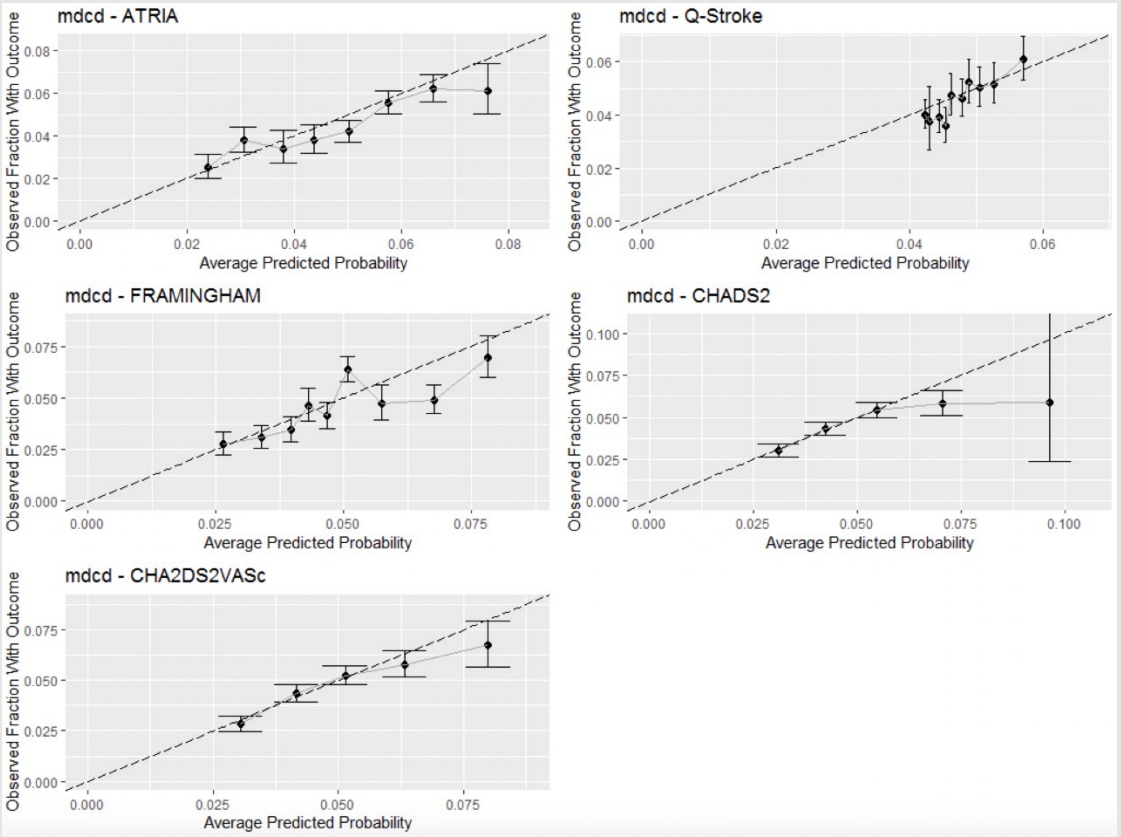
Table 58. Reps, 2020[110]

Reference	Reps, 2020[110]
Study type	Observational retrospective
Study sample	312,274 people with AF from 7 different observational healthcare datasets (CCAE, MDCD, CUMC, AUSOM, STRIDE, Optum claims, Optum EHR) were relevant to this review. There is no evidence that these were overlapping datasets. Two samples were examined – one aged 65-95 and one of all ages; only the latter were included in this review. In addition we only extracted data for the sub-group where *no anticoagulants* were used prior to the study, and *none were used during follow up*, and patients had to have no prior stroke. Since each dataset was in a different population these were regarded as separate sources of data, and thus C statistics can be regarded as separate datapoints for meta-analysis. Characteristics of each relevant database were as follows: CCAE: n=46054, strokes during FU = 589; USA MDCD: n=29546, strokes during FU = 1386; USA CUMC: n=4546, strokes during FU = 243; USA AUSOM: n=256, strokes during FU = 7; S. KOREA STRIDE: n=2786, strokes during FU = 34; USA Optum claims: n=100,757, strokes during FU=3758; USA Optum EHR: n=128,409, strokes during FU=1567; USA

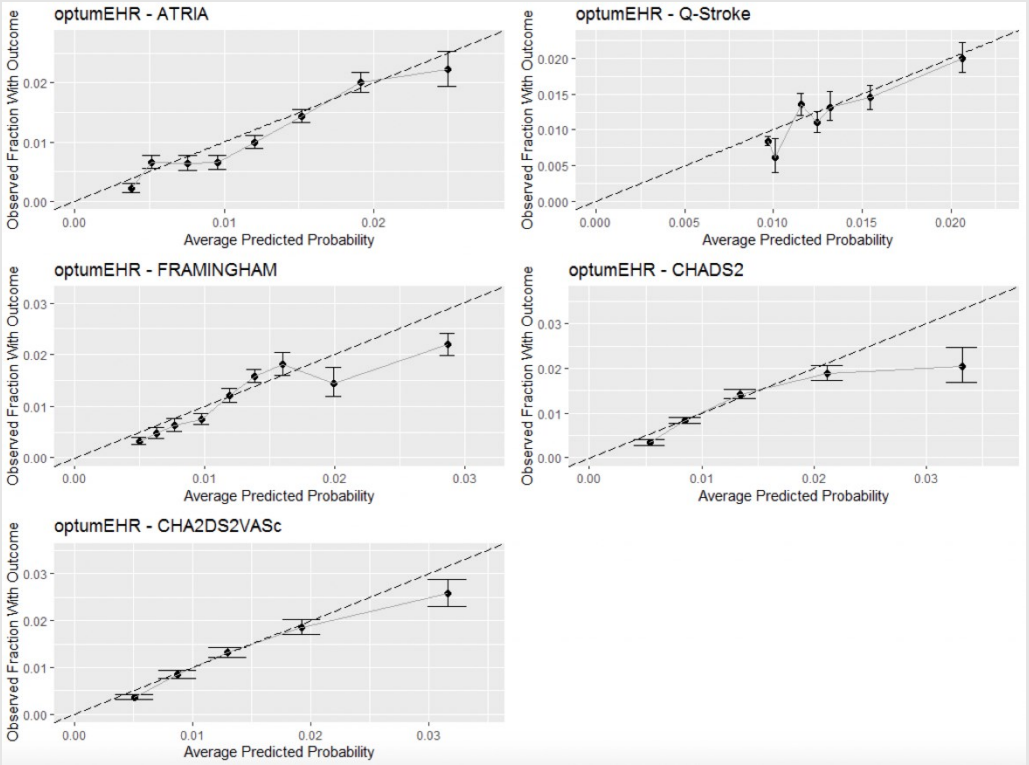
Reference	Reps, 2020[110]
Inclusion criteria	Females with either 2 AF records on database OR 1 AF event in inpatient setting OR 1 AF with ECG within 30 days prior; at least 730 days prior database observation; no prior stroke and no prior anticoagulant, and no anticoagulants in follow up.
Exclusion criteria	See above
Risk tools	ATRIA CHADS2 CHADSVASC FRAMINGHAM Q-Stroke
Outcome	Authors predicted stroke occurring 1 day until 365 days after the initial atrial fibrillation start date. The stroke out-come was defined as an ischemic or hemorrhagic stroke recorded with an inpatient or ER visit
Results	7584 stroke events C statistics (using data for all ages, no prior anticoagulants, and censored for AC use during 1 year follow up). Only the databases where 95% CIs were provided have been reported here. *CCAE database* ATRIA: 0.63(0.61-0.66) CHADS2: 0.62(0.60-0.65) CHADSVASC: 0.65(0.62-0.67) FRAMINGHAM: 0.64(0.61-0.66) Q-Stroke: 0.62(0.60-0.64) *MDLD database* ATRIA: 0.58(0.56-0.59) CHADS2: 0.56(0.55-0.58) CHADSVASC: 0.58(0.56-0.59) FRAMINGHAM: 0.57(0.56-0.59) Q-Stroke: 0.55(0.53-0.56) *CUMCdatabase* ATRIA: 0.67(0.64-0.70) CHADS2: 0.64(0.61-0.68) CHADSVASC: 0.66(0.63-0.69)

Reference	Reps, 2020[110]
	FRAMINGHAM: 0.66(0.63-0.69) Q-Stroke: 0.57(0.53-0.61) *AUSOM database* ATRIA: 0.79(0.63-0.94) CHADS2: 0.72(0.53-0.91) CHADSVASC: 0.81(0.71-0.90) FRAMINGHAM: 0.76(0.59-0.93) Q-Stroke: 0.68(0.42-0.94) *STRIDE database* ATRIA: 0.53(0.43-0.63) CHADS2: 0.51(0.41-0.62) CHADSVASC: 0.55(0.44-0.65) FRAMINGHAM: 0.62(0.51-0.72) Q-Stroke: 0.47(0.36-0.57) *Optum claims database (no CIs provided)* ATRIA: 0.67 CHADS2: 0.64 CHADSVASC: 0.65 FRAMINGHAM: 0.65 Q-Stroke: 0.58 *Optum EHR database (no CIs provided)* ATRIA: 0.67 CHADS2: 0.65 CHADSVASC: 0.67 FRAMINGHAM: 0.66 Q-Stroke: 0.6

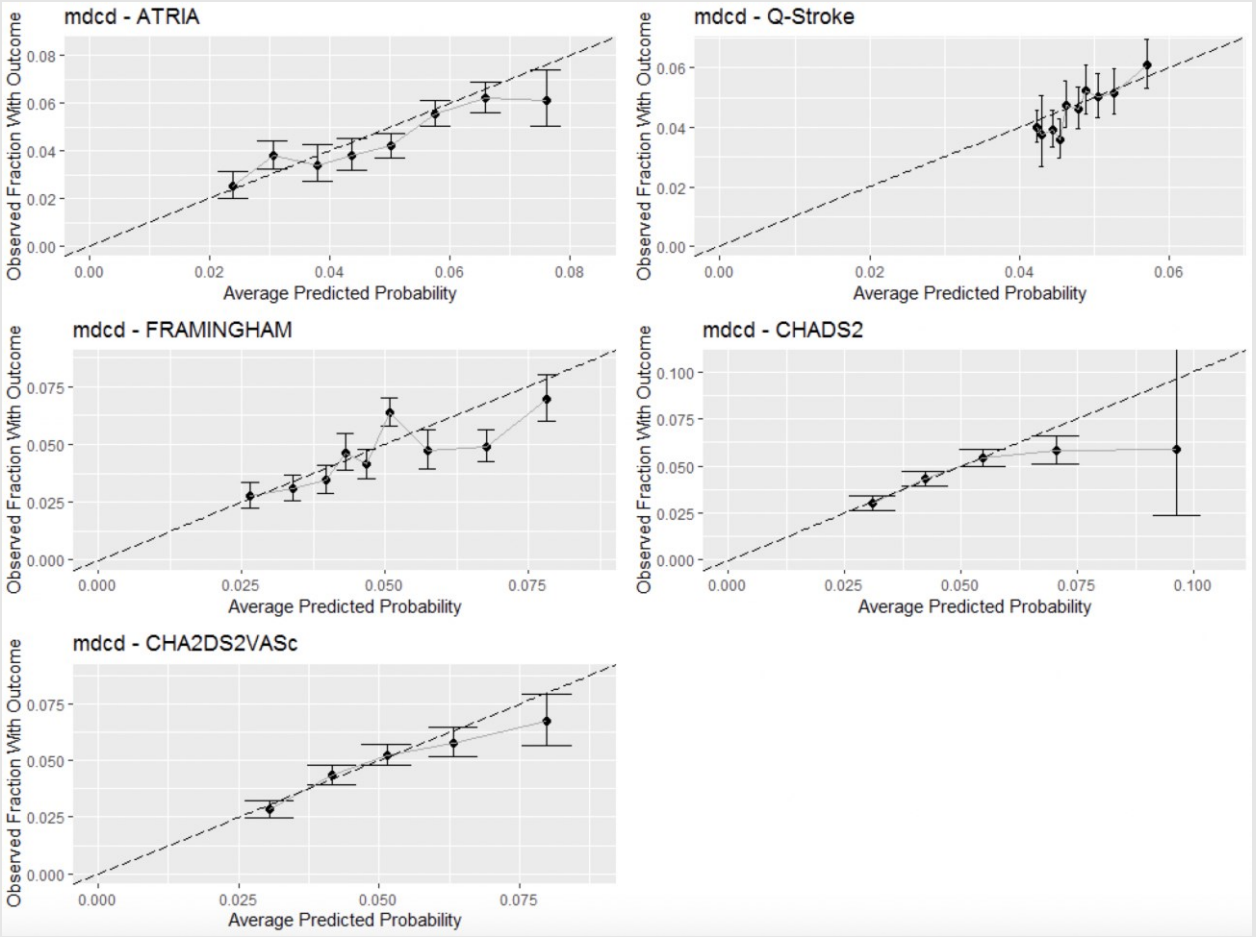
Reference	Reps, 2020[110]
	Calibration: Calibration plots of each of the 5 tools were produced for each of the separate databases:

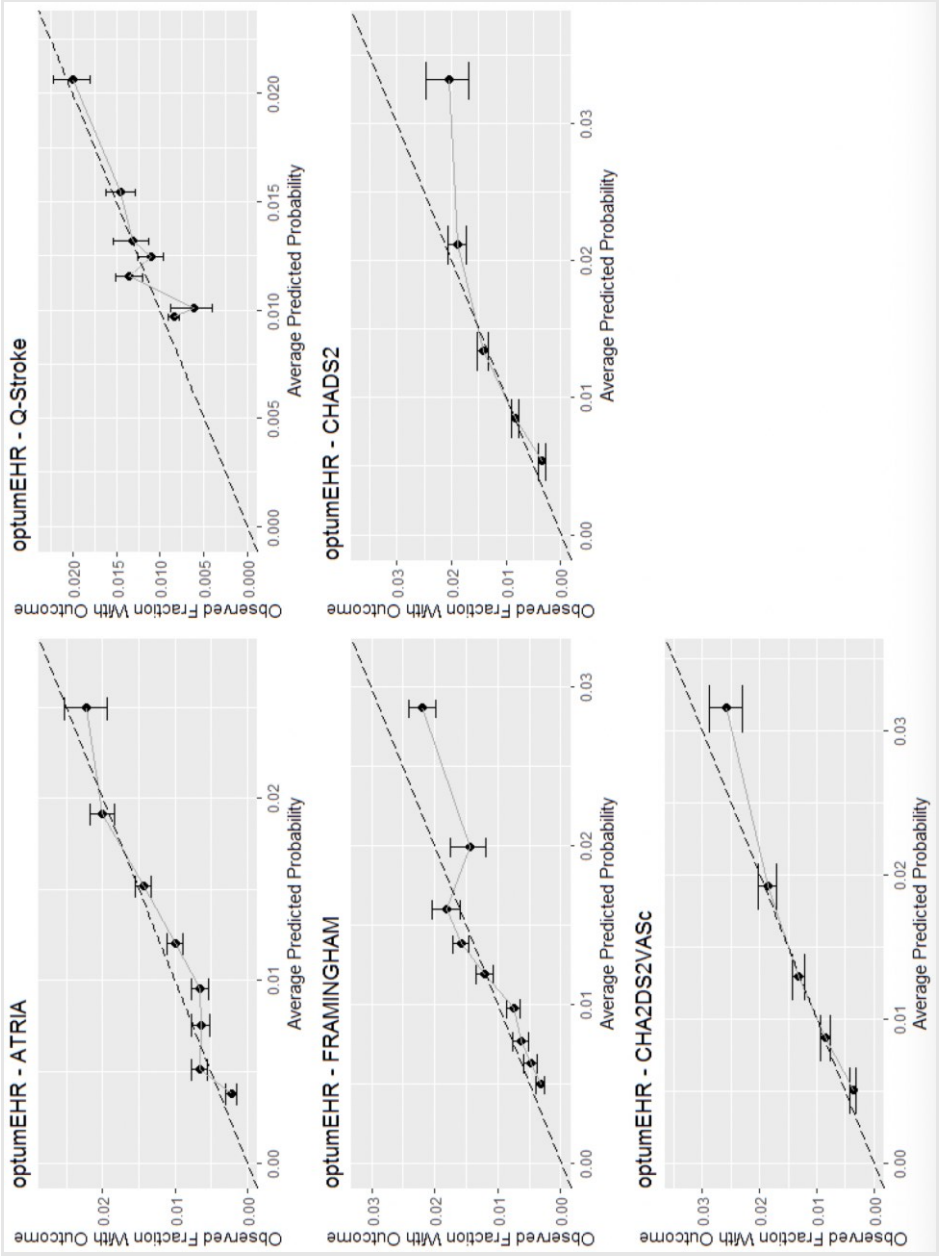


Reference	Reps, 2020[110]

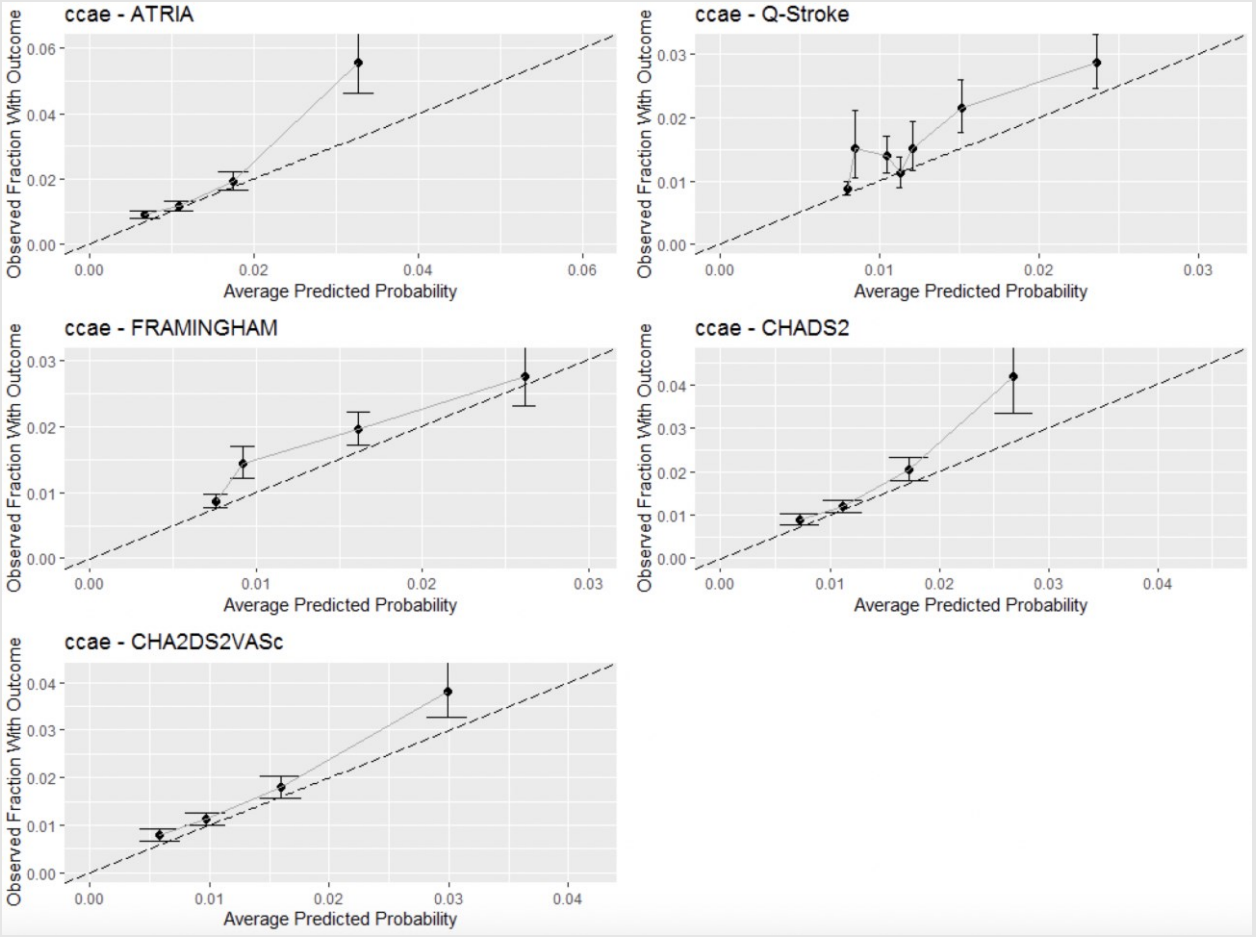


Reference	Reps, 2020[110]

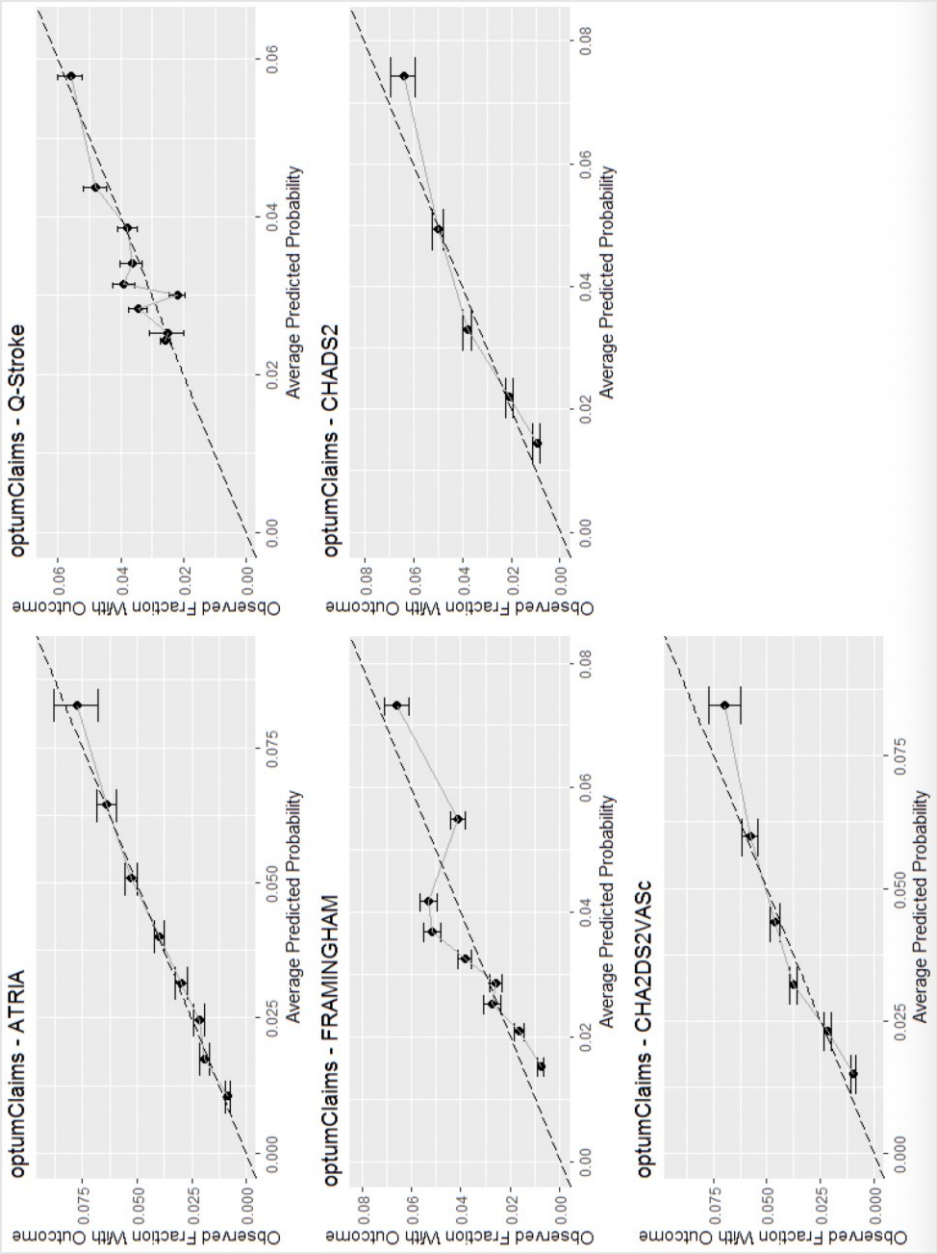


Reference	
Reps, 2020[110]	

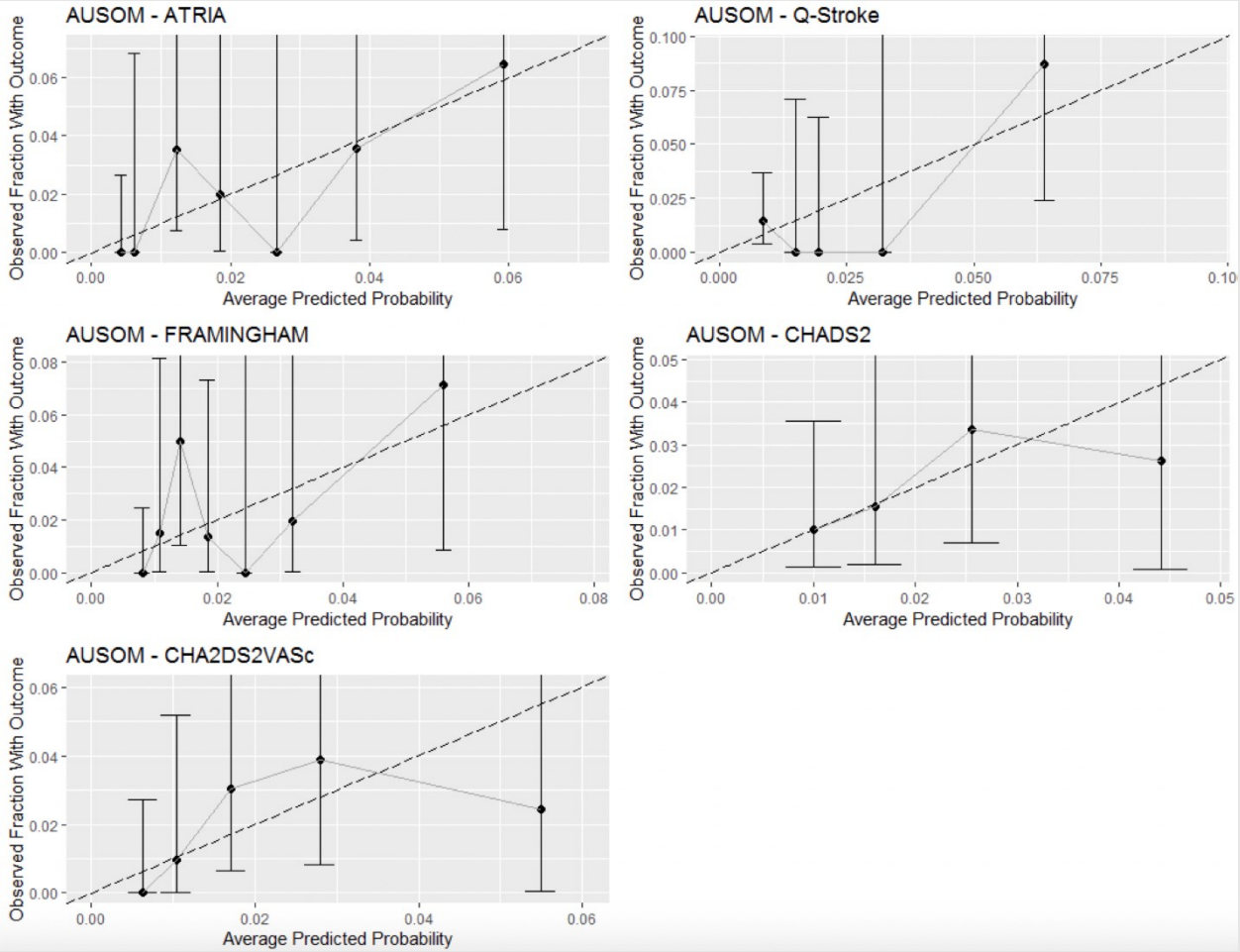
Reference	Reps, 2020[110]



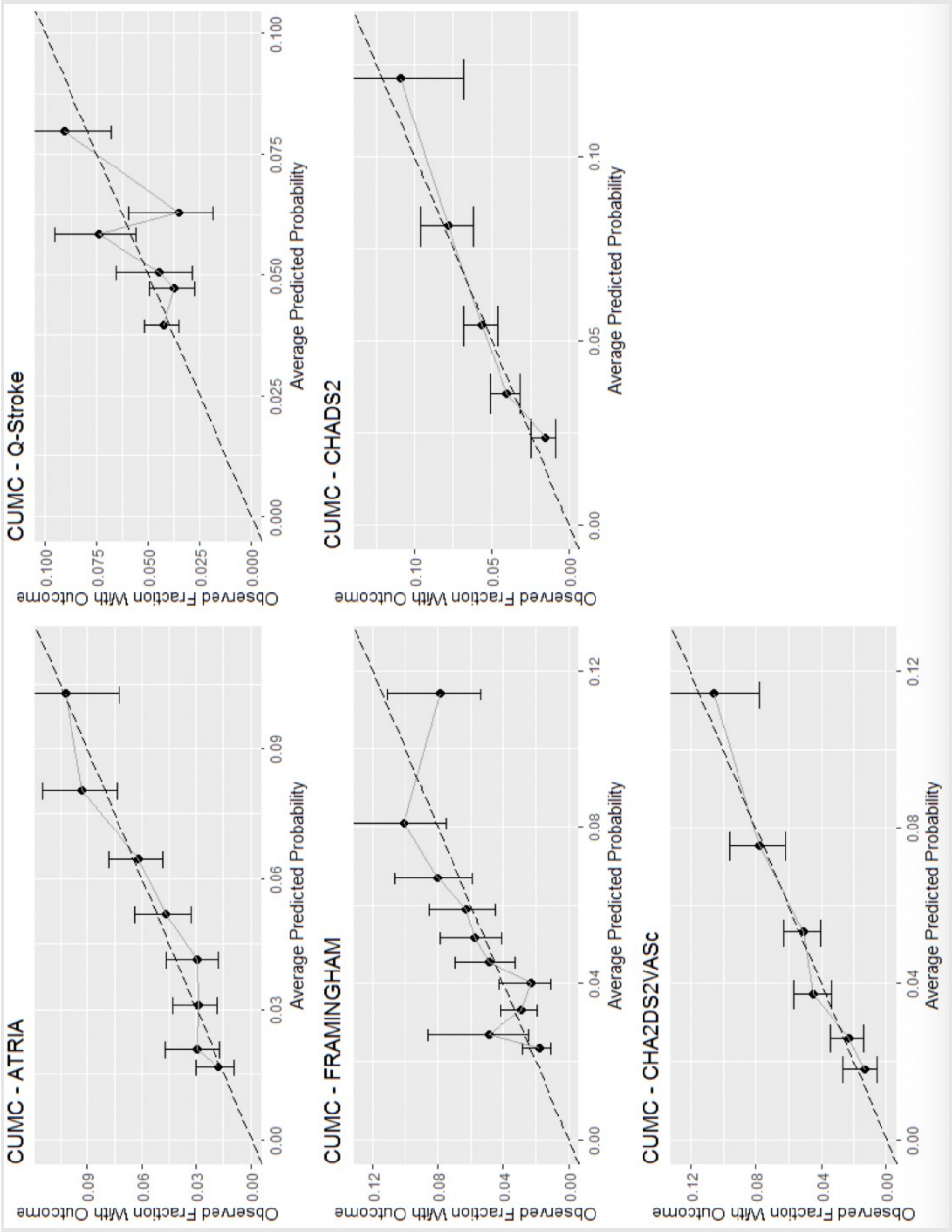
Reference	Reps, 2020[110]

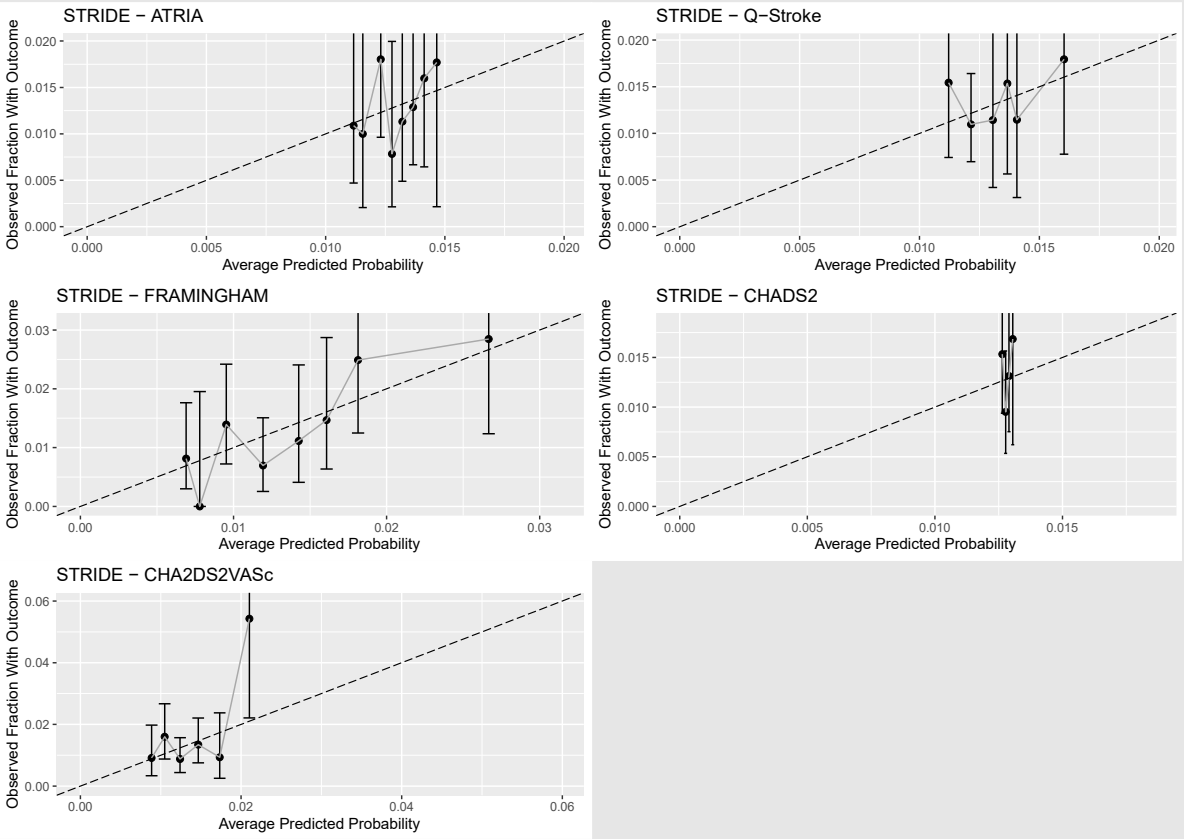


Reference	Reps, 2020[110]



Reference	Reps, 2020[110]



Reference	Reps, 2020[110]
	
Why the group were not anticoagulated	The data from the databases were selected for people who were not anticoagulated.

Appendix G: Risk of bias (PROBAST)

Study	Appropriate data sources?	Appropriate inc and exc?	Similar health across participants?	Predictors defined/ass' d same for all?	Predictor assessments made without knowledge of outcome data?	Predictors all available at time model meant to be used?	All relevant predictors analysed?	Pre-specified outcome used?	Predictors excluded from outcome definition?	Outcome defined in same way for all?	Outcome determined without knowledge of predictor information?	Reasonable number of outcome events? (100)	Time interval between baseline and outcome appropriate? (5 years)	All enrolled included in analysis?	Missing data handled appropriately?	Non:binary predictors handled appropriately?	Complexities in data accounted for?	Relevant performance measures?	Model recalibrated or likely that calibration not needed?	Overall rating
Abraham 2013[2]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y: 457 events	Y: up to 17 years	2.3% lost to FU	Y	Y	Y	Y	Y	Serious
Abumaileq 2015a[3]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N: 9	N: 11 months	only 4/15 4 lost	Y	Y	Y	Y	Y	Very serious
Aspberg 2016[10]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y: 11052	N: up to 5 years	yes	N A	Y	Y	Y	Y	Very serious
Chao 2016[21]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y:21008	Y: 10 years	U	U	Y	Y	Y	Y	Serious
Fang 2008[30]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y: 685	Y: 6 years	U	U	Y	Y	Y	Y	Serious
Fox 2017[33]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N: 51	N: 3 years	U	U	Y	Y	Y	Y	Very Serious
Friberg 2012b[36]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y:5359	N: 1.4 years	U	U	Y	Y	Y	Y	Very serious
Gage 2001 [39]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N: 94	N: 1.2 years	U	U	Y	Y	Y	Y	Very serious
Gage 2004[38]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y:207	N: 1.9 years	U	U	Y	Y	Y	Y	Very serious

Study	Appropriate data sources?	Appropriate inc and exc?	Similar health across participants?	Predictors defined/ass' d same for all?	Predictor assessments made without knowledge of outcome data?	Predictors all available at time model meant to be used?	All relevant predictors analysed?	Pre-specified outcome used?	Predictors excluded from outcome definition?	Outcome defined in same way for all?	Outcome determined without knowledge of predictor information?	Reasonable number of outcome events? (100)	Time interval between baseline and outcome appropriate? (5 years)	All enrolled included in analysis?	Missing data handled appropriately?	Non:binary predictors handled appropriately?	Complexities in data accounted for?	Relevant performance measures?	Model recalibrated or likely that calibration not needed?	Overall rating
Guo 2013[41]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N:85	N: 1.9 years	U	U	Y	Y	Y	Y	Very serious
Hippisley Cox 2013[49]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y: 890	Y: 10 years	U	U	Y	Y	Y	Y	Serious
Kang 2017[60]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y: 888	N: 1.2 years	U	U	Y	Y	Y	Y	Very serious
Kim 2017[66]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y: 819	N: 4.2 years	U	U	Y	Y	Y	Y	Very serious
Larsen 2012[71]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N: U	Y: 5.4 years	U	U	Y	Y	Y	Y	Very serious
Lip 2006[74]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N: U	N: 1.6 years	U	U	Y	Y	Y	Y	Very serious
Lip 2010[75]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N: U	N: 1 year	U	U	Y	Y	Y	Y	Very serious
Lip 2014[76]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y:273	Y: 10 years	U	U	Y	Y	Y	Y	Serious
McAlister, 2017[83]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y:7,364	N – 2.5 years	U	U	Y	Y	Y	Y	Very serious
McAlister, 2018[84]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y:10,827	N – 1 years	U	U	Y	Y	Y	Y	Very serious

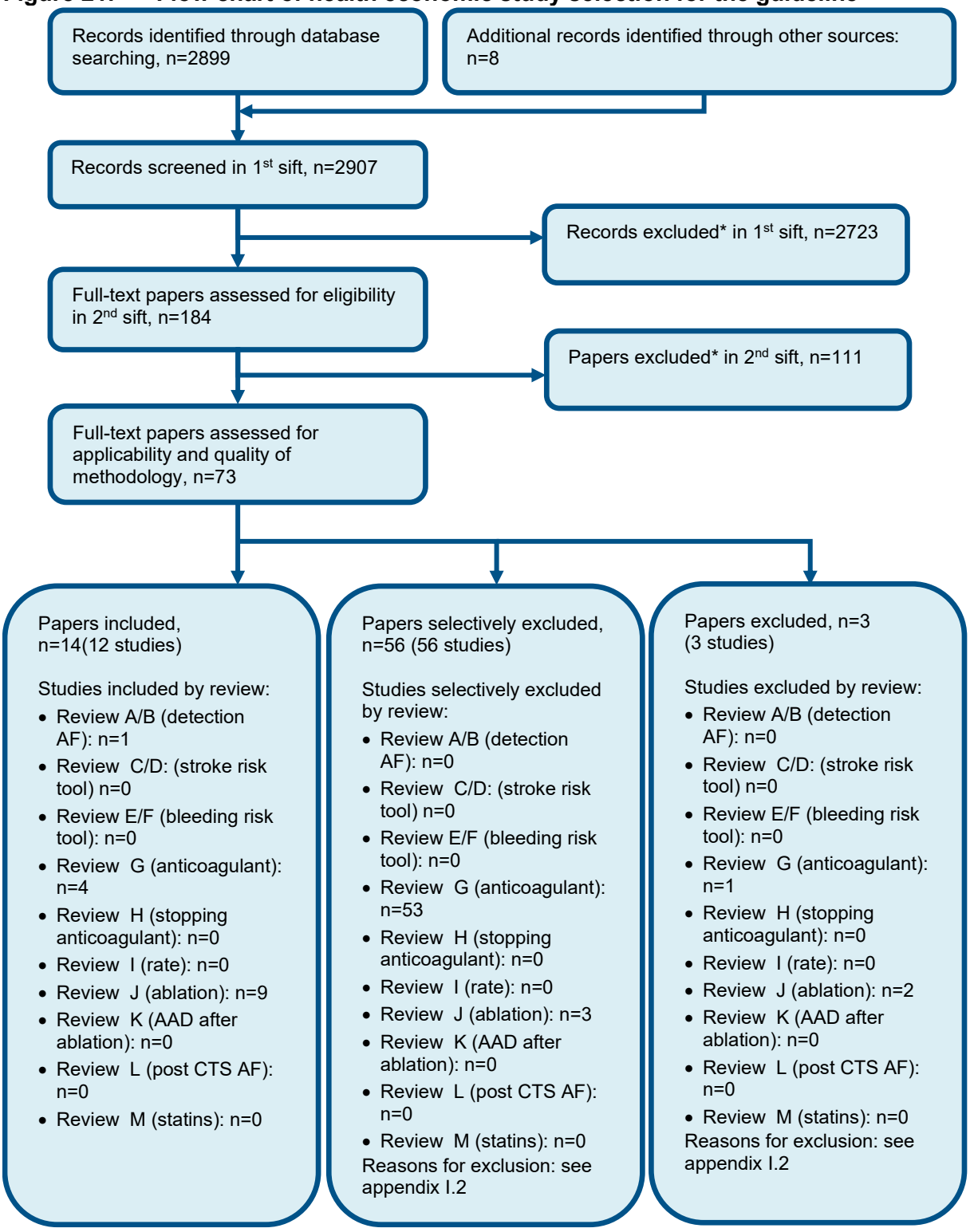
Study	Appropriate data sources?	Appropriate inc and exc?	Similar health across participants?	Predictors defined/ass' d same for all?	Predictor assessments made without knowledge of outcome data?	Predictors all available at time model meant to be used?	All relevant predictors analysed?	Pre-specified outcome used?	Predictors excluded from outcome definition?	Outcome defined in same way for all?	Outcome determined without knowledge of predictor information?	Reasonable number of outcome events? (100)	Time interval between baseline and outcome appropriate? (5 years)	All enrolled included in analysis?	Missing data handled appropriately?	Non:binary predictors handled appropriately?	Complexities in data accounted for?	Relevant performance measures?	Model recalibrated or likely that calibration not needed?	Overall rating
Olesen 2011[96]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	U but likely to be >100	N: 1 year	U	U	Y	Y	Y	Y	Very serious
Olesen 2012 [95]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N:14	Y: 10 years	U	U	Y	Y	Y	Y	Very serious
Olesen 2012b[97]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y: 4599	Y: 12 years	U	U	Y	Y	Y	Y	Serious
Singer 2013 [123]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y: 496	N: 1 year	U	U	Y	Y	Y	Y	Very serious
Siu 2014[124]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y: 847	N: 3.2 years	U	U	Y	Y	Y	Y	Very serious
Suzuki 2015[127]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N: 69	N: 1.4 years	U	U	Y	Y	Y	Y	Very serious
Tomita 2015[131]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N: 30	N: 2 years	N	Y	Y	Y	Y	Y	Very serious
Van dem Ham 2015[133]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y: 3751	N: 2.1 y	U	U	Y	Y	Y	Y	Very serious
Van Staa 2011[136]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y: 1233	N: 4 years	U	U	Y	Y	Y	Y	Very serious
Wang 2003[137]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N: 83	N: 4 years	U	U	Y	Y	Y	Y	Very serious

Study	Appropriate data sources?	Appropriate inc and exc?	Similar health across participants?	Predictors defined/ass' d same for all?	Predictor assessments made without knowledge of outcome data?	Predictors all available at time model meant to be used?	All relevant predictors analysed?	Pre-specified outcome used?	Predictors excluded from outcome definition?	Outcome defined in same way for all?	Outcome determined without knowledge of predictor information?	Reasonable number of outcome events? (100)	Time interval between baseline and outcome appropriate? (5 years)	All enrolled included in analysis?	Missing data handled appropriately?	Non:binary predictors handled appropriately?	Complexities in data accounted for?	Relevant performance measures?	Model recalibrated or likely that calibration not needed?	Overall rating
Xing 2016[139]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N:59	N: 2years	U	U	Y	Y	Y	Y	Very serious
Xing 2018[140]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N:49	N: 2.6 years	U	U	Y	Y	Y	Y	Very serious
Yoshizawa 2017[142] and Komatzu, 2014[68]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N: 2.1% of 332 per year	N: 53 months	U	U	Y	Y	Y	Y	Very serious
Schwartz, 2019[122]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y	N	U	U	Y	Y	Y	Y	Very serious
Maheshwari, 2019[81]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N	N	U	U	Y	Y	Y	Y	Very serious
Piccini, 2013 [100]	Y	U	U	U	U	Y	Y	Y	NA	Y	U	U	N	U	U	Y	Y	Y	Y	Very serious
Wicke, 2019 [138]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y	N	U	U	Y	Y	Y	Y	Serious
Tomasdottir, 2019[130]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y	N	U	U	Y	Y	Y	Y	Serious
Reps, 2020[110]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	Y	Y	N	U	Y	Y	Y	Y	Y	Serious

N=no, Y=yes, U=unclear

Appendix H: Health economic evidence selection

Figure 21: Flow chart of health economic study selection for the guideline



* Non-relevant population, intervention, comparison, design or setting; non-English language

Appendix I: Economic evidence tables

None for both reviews.

Appendix J: Excluded clinical studies

Table 59: Studies excluded from the clinical review on the effectiveness of tools to predict stroke or thromboembolic events

Reference	Reason for exclusion
Guo 2017[42]	Incorrect comparison: decision tool versus usual care
Karlsson 2017[63]	Incorrect study design: study protocol
Karlsson 2018[62]	Incorrect comparison: decision tool versus usual care
Pandya 2018[98]	Incorrect study design: prospective cohort study

Table 60: Studies excluded from the clinical review on the accuracy of tools to predict stroke or thromboembolic events

Study	Exclusion reason
Aakre, 2014[1]	Included anticoagulated participants
Abumaileq, 2015b[4]	Included anticoagulated participants
Al-Radeef, 2019[5]	Descriptive study – no predictive risk analysis
Al-Turaiki, 2016[6]	Included anticoagulated participants
Alraies, 2017[7]	conference abstract
Andersson, 2017[8]	no accuracy outcomes
Asberg, 2010[9]	non AF population
Atzema, 2015[11]	No stroke/TE outcomes
Banerjee, 2013[13]	conference abstract
Banerjee, 2013[14]	Included anticoagulated participants
Banerjee, 2014[12]	Included anticoagulated participants
Baruch, 2007[15]	Included anticoagulated participants
Basili, 2017[16]	Included anticoagulated participants
Berg,2019[17]	All patients on OACs
Borre, 2018[18]	SR - papers checked
Camelo-Castilo, 2020[19]	no specific predictive accuracy outcomes for stroke
Chan, 2016[20]	No accuracy outcomes
Chao, 2012[22]	no accuracy outcomes
Chao, 2015a[24]	Derivation study
Chao, 2015b[23]	no accuracy outcomes
Charfeddine, 2020[25]	abstract
Chen, 2019[26]	abstract
Dalgaard, 2019[27]	Overall analysis used a mixture of people on and not on OACs. A sub-group analysis was performed for low risk patients (men CHADSVASC 0-1, women 0-2) who were not on OACs but this group is regarded as a special group and not representative of the overall population
Di Toro, 2013[28]	Non English
Dzeshka, 2014[29]	Review -papers checked
Fauchier, 2016[31]	no accuracy outcomes
Forslund, 2014[32]	no accuracy outcomes
Friberg, 2012[35]	Contained large proportion taking warfarin and analysis not stratified for warfarin/no warfarin.
Friberg, 2012a[34]	no accuracy outcomes
Friberg, 2015[37]	no accuracy outcomes

Study	Exclusion reason
Gazova, 2019[40]	86% received OACs; no relevant analyses
Gupta, 2012[43]	Unclear if included anticoagulated participants
Hijazi, 2015[46]	conference abstract
Hijazi, 2016[47]	Included anticoagulated participants
Hijazi, 2016[48]	conference abstract
Hijazi, 2016a[45]	Included anticoagulated participants
Hijazi, 2017[44]	Included anticoagulated participants
Hippisley-Cox, 2014[50]	Not in AF population
Holt,2018[51]	Not a predictive risk analysis study
Horne, 2019[52]	11% on warfarin with no sub-grouping
Hu, 2018[53]	Unclear if anticoagulated
Huang, 2017[54]	prediction of left atrial thrombus
Inohara, 2017[55]	Included anticoagulated participants
Inoue, 2006[56]	no accuracy outcomes
Jaakkola, 2018[57]	no accuracy outcomes
Jaakola, 2018[57]	No relevant outcomes; mixture of people on and off OACs
Joundi, 2016[58]	SR
Kabra, 2016[59]	Included anticoagulated participants
Kang, 2017[61]	Most of the sample without AF
Kearon,2019[64]	Commentary on Berg, 2019
Kim, 2015[65]	Included anticoagulated participants
Kim, 2017b[67]	no accuracy outcomes
Komatsu, 2012[69]	no accuracy outcomes
Laguna, 2005[70]	no accuracy outcomes
Larsen, 2011[72]	conference abstract
Lin,2018[73]	No predictive analyses undertaken; Insufficient data to calculate predictive measures (numbers of people at each CHADSVASC score given, but not proportion of these with stroke. Incidence density (strokes per 100 person-years) given for stroke but cannot use this to extrapolate numbers with stroke as the incidence density may be confounded by a person having > 1 stroke.
Lip, 2013[78]	Included anticoagulated participants
Lip, 2014[77]	Valvular AF
Lowres,2019[79]	SR and meta-analysis
Maeda, 2020[80]	no predictive accuracy data
Masaki, 2009[82]	Included anticoagulated participants
Naccarelli, 2012[85]	Included anticoagulated patients
Nagahara, 2020[86]	anticoagulated cohort
Nakagawa, 2011[87]	No accuracy outcomes
Nielsen, 2019[90]	anticoagulated cohort in AF sub-group
Nielsen, 2020[89]	narrative review
Ntaios, 2019[91]	Prediction of mortality, not stroke
O'Brien, 2015[92]	conference abstract
Oldgren, 2016a[93]	conference abstract
Oldgren, 2016b[94]	Included anticoagulated participants
Parsons,2018[99]	Non AF population

Study	Exclusion reason
Piyaskulkaew, 2014[101]	No stroke/TE outcome; population limited to CHADS 0-1
Poli, 2011[105]	Included anticoagulated participants
Poli, 2014[104]	No accuracy outcomes
Poli, 2017[103]	Included anticoagulated participants
Poli,2009a[102]	Included anticoagulated participants
Potpara, 2012[107]	Included anticoagulated participants
Potpara, 2012[106]	Included anticoagulated participants
Proietti, 2020[108]	SR
Puurunen, 2014[109]	Population undergoing percutaneous coronary intervention
Rietbrock, 2008[111]	Included anticoagulated participants
Rivera-Caravaca, 2017[114]	conference abstract
Rivera-Caravaca, 2017b[113]	bleeding risk study
Rivera-Caravaca, 2018[115]	Included anticoagulated participants
Rivera-Caravaca, 2018[115]	All patients on VKAs
Rivera-Caravaca,2017a[112]	Included anticoagulated participants
Roldan, 2018[116]	Included anticoagulated participants
Ruff, 2016[117]	Included anticoagulated participants
Ruiz-Ortiz, 2010[118]	Included anticoagulated participants
Saito, 2020[119]	anticoagulated cohort
Sander Van Doorn, 2018[120]	SR - references checked
Somme, 2010[125]	Included anticoagulated participants and no accuracy outcomes
Sun, 2019[126]	No predictive outcomes
Tanaka, 2015[128]	Outcomes were severity of stroke in a cohort who all had stroke
Tanaka, 2018[129]	Included anticoagulated participants
Tsai, 2014[132]	Included anticoagulated participants
Van Den Ham, 2014[134]	conference abstract
Van Mieghem, 2017[135]	Review; All patients on VKAs
Yang, 2018[141]	Included anticoagulated participants
Zhu, 2017[143]	SR - references checked
Zhu, 2020[144]	not an AF population

Appendix K: Excluded economic studies

Studies that meet the review protocol population and interventions, and the economic study inclusion criteria but have not been included in the review based on applicability and/or methodological quality are summarised below with reasons for exclusion.

Table 61: Studies excluded from the health economic review

Reference	Reason for exclusion
None for both reviews	