National Institute for Health and Care Excellence

Final

Atrial fibrillation: diagnosis and management

Evidence reviews E&F: Risk stratification tools for predicting bleeding events in people with atrial fibrillation

NICE guideline NG196

Evidence review

April 2021

Final

Developed by the National Guideline Centre, Royal College of Physicians

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 patientand, where appropriate,their carer or guardian.

Local commissioners andproviders have a responsibility to enable the guideline to be applied when individual health professionals and theirpatients 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 All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN978-1-4731-4043-1

Contents

1 Effectiveness of risk stratification tools for predicting bleeding in people with atrial fibrillation 6

1.1 Review question: What is the most clinically and cost-effective risk stratification tool for predicting bleeding in people with atrial fibrillation? 6

1.2 Introduction 6

1.3 PICO table 6

1.4 Methods and process 7

1.5 Clinical evidence 7

1.5.1 Included studies 7

1.5.2 Excluded studies 7

1.5.3 Summary of clinical studies included in the evidence review 7

1.5.4 Quality assessment of clinical studies included in the evidence review 7

1.6 Economic evidence 8

1.7 Included studies 8

1.8 Excluded studies 8

1.8.1 Unit costs 8

2 Accuracy of risk stratification tools for predicting bleeding events in people with atrial fibrillation 10

2.1 Introduction 10

2.2 Review question: What is the most accurate risk stratification tool for predicting bleeding events in people with atrial fibrillation? 10

2.3 Clinical evidence 10

2.3.1 Discrimination for MAJOR BLEEDING 27

2.3.2 Calibration for MAJOR BLEEDING 47

2.3.3 Net Reclassification improvement for MAJOR BLEEDING 54

2.3.4 Discrimination for CLINICALLY RELEVANT BLEEDING 61

2.3.5 Calibration for CLINICALLY RELEVANT BLEEDING 71

2.3.6 Net Reclassification improvement for CLINICALLY RELEVANT BLEEDING 71

2.3.7 Discrimination for INTRACRANIAL HEMORRHAGE 73

2.3.8 Calibration for INTRACRANIAL HEMORRHAGE 79

2.3.9 Net Reclassification improvement for INTRACRANIAL HEMORRHAGE 79

2.4 Economic evidence 81

2.4.1 Included studies 81

2.4.2 Excluded studies 81

2.4.3 Unit costs 81

2.5 The committee's discussion of the evidence 82

2.5.1 Interpreting the evidence 82

2.5.2 Other factors the committee took into account 84

Appendices 99

Appendix A: Review protocols 99

Appendix B: Literature search strategies 110

Appendix C: Clinical article selection 118

Appendix D: Economic article selection 120

Appendix E: Full GRADE tables (Including individual study data) 121

Appendix F: Forest plots 172

Appendix G: Clinical evidence tables 208

Appendix H: Risk of bias (PROBAST) 288

Appendix I: Economic evidence tables 294

Appendix J: Excluded clinical studies 295

Appendix K: Excluded economic studies 298

1 Effectiveness of risk stratification tools for predicting bleeding in people with atrial fibrillation

1.1 Review question: What is the most clinically and cost-effective risk stratification tool for predicting bleeding in people with atrial fibrillation?

1.2 Introduction

Anticoagulation is the therapy with the greatest influence on prognostic outcomes for patients with atrial fibrillation. Anticoagulation,however, is associated with significant risk for major haemorrhage, from one to seven per cent per annum in clinical trials. For the majority of patients with AF the benefits of anticoagulation outweigh this risk.

The risk of major haemorrhage varies among populations with AF and there is a potential to reduce harm further by identifying patients at high risk for whom to proceed with caution, particularly as many risk factors for haemorrhage on anticoagulation are modifiable. There are over twenty schemes & methods (including modifications), published, that attempt to quantify the risk of major haemorrhage on anticoagulation.The predicted risk of haemorrhage for an individual is not precise. It needs to be interpreted in context as many of the factors that increase risk of bleeding also increase the risk of embolic stroke.

The intention of this chapter is to evaluate which is the most clinical and cost effective method and to develop guidance as to how this informs clinical practice.

1.3 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.
Interventions	Any bleeding risk tool (for example, ATRIA, HEMORRHAGES, ORBIT) [Note: **treat each test using a different threshold as a separate intervention**].
Comparison	HAS-BLED (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

1.4 Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual.[89]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 how 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.

Declarations of interest were recorded according to NICE's 2018[89]conflicts of interest policy.

1.5 Clinical evidence

1.5.1 Included studies

No relevant comparative clinical studies comparing bleeding risk tools with HAS-BLED were identified.

See also the study selection flow chart in appendix C, study evidence tables in appendix D, forest plots in appendix E and GRADE tables in appendix H.

1.5.2 Excluded studies

See the excluded studies list in appendix I.

1.5.3 Summary of clinical studies included in the evidence review

No studies were included

1.5.4 Quality assessment of clinical studies included in the evidence review

Not applicable.

See appendix F for full GRADE tables.

1.6 Economic evidence

1.7 Included studies

No relevant health economic studies were identified.

1.8 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.

1.8.1 Unit costs

Outlined in **Table 2**is a description of each risk tool and any additional healthcare resources required. As demonstrated in the table most risk tools require a review of the person's medical history and in some cases computer access to complete algorithms. Only the ABC bleeding risk score required additional tests (biomarker assays), which would be an additional cost to the NHS.

Table 2: Bleeding risk tools

Risk tool	Description	Additional tests required to complete risk tool
ABC bleeding score	**- A**ge **- B**iomarkers (hematocrit, high sensitivity troponin T (hsTnT), GDF-15) **- C**linical history (prior bleeding)	Biomarkers.
Orbit bleeding score	- older age (75+ years) - reduced haemoglobin/haematocrit/history of anaemia - bleeding history - insufficient kidney function - treatment with antiplatelet	None
ATRIA	- anaemia - severe renal disease - age ≥75 years - any prior haemorrhage diagnosis - hypertension history	None
HEMORR2HAGES	- hepatic or renal disease - ethanol (alcohol) abuse - malignancy history - age >75 years - platelet count or function - rebleeding risk - hypertension (uncontrolled) - anaemia - genetic factors (CYP2C9 single nucleotide polymorphisms) - excessive fall risk	Genetic testing

Risk tool	Description	Additional tests required to complete risk tool
	- stroke history	
HAS-BLED	- uncontrolled hypertension - renal disease - liver disease - stroke history - prior major bleeding or predisposition to bleeding - labile INR - age >65 - concomtant antiplatelets or NSAIDs - alcohol excess/abuse	None

2 Accuracy of risk stratification tools for predicting bleeding events in people with atrial fibrillation

2.1 Introduction

See evidence review E.

2.2 Review question: What is the most accurate risk stratification tool for predicting bleedingevents in people with atrial fibrillation?

For full details see review protocol in Appendix A.

Table 3: PICO characteristics of review question

Question	
Population	People aged >18 with a diagnosis of atrial fibrillation, who are on anticoagulants
Risk tool	Any bleedingrisk tool (e.g HAS-BLED, ORBIT, HEMORRHAGES, ATRIA, etc) Any other version of HAS-BLEDwith modifications
Target condition or Reference standard	Later major bleeding, or other bleeding
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

2.3 Clinical evidence

We searched for cohort studies covering the validation of risk assessment tools for bleeding in people with AF. 54studies evaluating the accuracy of bleedingrisk tools for people with atrial fibrillation were included in the review[3, 5, 8, 11, 14, 19-21, 23, 25, 30-33, 36-39, 41, 52, 54, 56-58, 63, 65, 71, 74, 77, 88, 90, 91, 95, 103, 110, 113-117, 119, 120, 125, 126, 128, 135-138, 142, 146, 147, 154, 158]whichare summarised in Table 4below.The different risk schemes are outlined in Table 3.Evidence from these studies is summarised in the GRADE clinical evidence profilesbelow (Tables 4 -13). See also the study selection flow chart in Appendix B, study evidence tables in Appendix E, forest plots in Appendix D, and excluded studies list in Appendix H.

This review evaluates the accuracy of the risk tools to predict bleeding, with reference to their discriminatory capabilities (sensitivity, specificity, and C statistics), calibration statistics and

the Net Reclassification Index. The reference standard was the incidence (or not) of major bleeding (or other bleeding categories) at follow up.Only studies where all patients were anticoagulated (or where an anticoagulated sub-group were a separately analysed) were included; this was because the aim of the review is to establish which tool can best predict bleeding in those people who are taking anticoagulation.

Analyses were by cohort rather than study; that is, where a study included separate analyses for different OACs, these were analysed as separate cohorts (as if they were separate studies). This approach facilitated sub-grouping for different OACs if heterogeneity was detected.

For sub-grouping by OAC, cohorts were categorised into 1) VKA cohorts, 2) Mixed VKA/DOAC/unclear category cohorts and 3) DOACcohorts. For sub-grouping by antiplatelets use, cohorts were categorised into 1) cohorts with <33% on antiplatelets/NSAIDs/aspirin, 2)cohorts with >33%on antiplatelets, and 3) cohorts where the number on antiplatelets were not reported.

Separate analyses were performed for 1) major bleeding, 2) clinically relevant bleeding and 3) intracranial bleeding. Data concerning other forms of bleeding were not analysed in this review as they were deemed to overlap with these 3 categories, though available dataare outlined in the clinical evidence tables.

Summary of included studies

Table 4: Summary of studies included in the review

Study	Risk tool(s)	OAC	Concomitant antiplatelets or NSAIDS	Population	Number and type of outcome events	Follow up duration
Apostolakis 2012[4]	HAS-BLED HEMORRHAGES ATRIA	Warfarin	18%	2,293 patients with AF on VKAs, from AMADEUS RCT trial in UK. Age 70, 65% male, 77% hypertension, 20% DM, 13.5% previous stroke, 31% CAD, 18% antiplatelet treatment, TTR 0.57. Drops outs NR. No blinding reported.	39 MB 251 CRB	429 days
Apostolakis 2013[3]	HAS-BLED CHADS2 CHADSVASC	Warfarin	18%	As above	As above	As above
Barnes 2014[8]	CHADS2 CHADSVASC HEMORRHAGES HAS-BLED ATRIA	Warfarin	NR	2600 patients with NVAF and on warfarin were recruited. USA study. Age 70, 41.7% female, hypertension 75%, DM 25%, CAD 33%, CHF 24.2%, current smoking 6%, renal disease 12%, stroke 11.5%, bleeding diasthesis 31%, HAS-BLED score 2.6, CHADS2 score 3.4. TTR 59.3. Antiplatelets/NSAIDs not reported. No blinding. No data loss reported.	100 MB	1 year
Berg 2019[11]	HAS-BLED ABC	Warfarin Edoxaban	NR	Patients enrolled on the ENGAGE AF-TIMI 48 trial, who were therefore taking VKAs or edoxaban. Participation in this sub-study was offered to all enrolled patients until recruitment reached 9000 participants	Unclear	3 years
Beshir 2018[14]	mOBRI CBRM HEMORRHAGES HAS-BLED ATRIA ORBIT	Warfarin, rivaroxaban, dabigatran	35%	1017 patients with NVAF and on Warfarin (INR 2-3), dabigatran or rivaroxaban between 2010 and 2015. Malaysia. Age >75: 27%, 52% male, hypertension 82%, IHD 33%, renal impairment 36%, DM 40%, prior stroke/TIA: 22%, CHF: 20%. CHADS2: 2. 35% on antiplatelets. No blinding. 291 lost to follow up from original sample of 1308 patients.	23 MB 76 CRNMB	1 year
Chang 2016[19]	HTI APTT Prothrombin time	dabigatran	12.50%	208 patients (213 enrolled and 5 lost to FU) with NVAF on dabigatran (either 100mg or 150mg/day). Taiwan. Age 74.7, 67.9% male, 36% history of stroke, 24.5% DM, 79.3% hypertension, 18.8% CAD, 16.3% HF, antiplatelets/NSAIDs 12.5%, renal disease 0.5%, history of GI bleeding 23.6%, HAS-BLED 1.8. 5 lost to follow up from original cohort of 213. No blinding.	17 MB	1 year
Chao 2018a[21]	Modifiable Bleeding Risk	Warfarin	22.70%	40,450 AF patients (defined as cases where there had been at least 2 confirmed outpatient diagnoses of AF) receiving warfarin	6889 MB	4.6 years

Study	Risk tool(s)	OAC	Concomitant antiplatelets or NSAIDS	Population	Number and type of outcome events	Follow up duration
	factors score (MBR) HEMORRHAGES HAS-BLED ATRIA ORBIT			between 1998 and 2011 in Taiwan. Age 67.3, male 55.7%, hypertension 67.4%, abnormal renal function 13.2%, stroke 43%, history of bleeding 18%, use of antiplatelets 22.7%, NSAIDs 7.2%, HAS-BLED 2.51. No loss to FU. No blinding reported.	1581 ICH	
Chao 2018b[20]	HAS-BLED baseline HAS-BLED change from baseline (Delta HAS-BLED) HAS-BLED follow up	Warfarin	2.30%	19,566 AF patients on Warfarin and a HAS_BLED score of <2 identified from the NHIRD of Taiwan (1998-2011). Age 63.8, male 57.4%, hypertension 52.6%, abnormal renal function 3.4%, stroke 22.6%, bleeding 6.9%, antiplatelet / NSAID drugs 2.3%. No loss to FU reported. No blinding reported.	3032 MB 671 ICH	4.8 years
Claxton 2018[23]	Anticoagulation-Specific Bleeding Score (ABS) HAS-BLED ATRIA HEMORRHAGES ORBIT	Warfarin, dabigatran, rivaroxaban and apixaban	NR	81,285 NVAF patients on Warfarin or DOACs (initiated at baseline). Netherlands. This was an external validation cohort from the Optum Clinformatics database from 2009-2015. For warfarin group (largest) the demographics were: age 73.9, 44% woman, HAS-BLED 2.8, HF 45.5%, CHD: 47.3%, hypertension 89%, DM 39.9%, stroke 33.4%, PAD 25.7%, kidney disease 25.9%, prior GI bleed 16%, prior IC bleed: 2.1%, prior other bleed 16%. No blinding reported. No loss to follow up (as retrospective). No data on antiplatelets/NSAIDS	3238 MB	1 year
Dalgaard 2019[25]	GARFIELD-AF HAS-BLED	Unclear	Unclear	51,180 Danish patients on OACs from the Danish Nationwide registries. Aged 18 or older with NVAF. Excluded patients with rheumatic valve disease or valve surgery.	1492 MB (but unclear if some had ICH)	1 year
Elvira-Ruiz, 2020[30]	HAS-BLED ORBIT ATRIA HAS-BLEDwith existence of aortic stenosis (AS) ORBITwith AS ATRIAwith AS	Mixed VKA and DOACS (results not sub-grouped)	17.7%	2,880 NVAF patients initiating oral anticoagulants; age 77; 51.1% women; 49.3% permanent AF; hypertension 85.5%; DM 33.9%; CHADSVASC 4; HASBLED 2; ATRIA 3; ORBIT 1.	185 MB	18 months
Esteve Pastor 2016[31]	HAS-BLED ORBIT	VKA and DOACS	10.90%	1276 patients with chronic NVAF on VKA or DOAC for at least 6 months before enrolment (FANTASIIA population). SPAIN. There was another cohort of 406 patients in this paper that	46 MB	1 year

Study	Risk tool(s)	OAC	Concomitant antiplatelets or NSAIDS	Population	Number and type of outcome events	Follow up duration
				underwent electrical cardioversion, and they are not included in this extraction. Age 74, 44% male, 80.6% hypertensive, 30% HF, 29.3% DM, 6.6% VD, 12.9% previous embolism, 3.8% previous bleeding, 10% renal impairment, 1.3% liver impairment, 77.4% VKA, 22.6% DOACs, 10.9% on NSAIDS / antiplatelets. HAS-BLED score: 2. TTR 60.9. No blinding. No loss to FU reported.		
Esteve-Pastor 2017a[5]	ABC-bleedingCrC HAS-BLED	VKAs	NR	1,120 patients with paroxysmal, persistent or permanent AF, stable on VKAs (INR 2-3). Spain. Age 76, 49.5% male, 82% hypertension, 27%DM, 33% dyslipidaemia, 15.5% current smoker, 31.2% HF, 19.6% CAD, 19% previous stroke, 8.4% previous bleeding. TTR at 6 months 80, CHADSVASC 4, HAS-BLED 2, ABC 16.5. Number on antiplatelets – not reported. No loss to FU reported. No blinding.	207 MB 65 ICH 85 GIB	6.5 years
Esteve-Pastor 2017b[32]	HAS-BLED Modifiable bleeding risk factors score	VKAs	21.40%	4576 patients with paroxysmal, persistent or permanent AF. 2283 on warfarin and 2293 on Idraparinux. Taken from the multinational AMADEUS database. Spain. Age 71, 66.5% male, 21.4% on anti-platelets or NSAID, 77% hypertensive, 20%DM, 23% HF, 31% CAD, 13% previous stroke, TTR 58, CHADSVASC 3, HAS-BLED 2, Modifiable bleeding risks score 1. No loss to FU reported. Assessors BLINDED.	113 MB 597 CRB	347 days
Fang 2011[33]	ATRIA Outpatient Bleeding Index Kuijer et al. Kearon et al. HEMORRHAGES Shireman Riete risk scheme	Warfarin	NR	3063 patients in the validation cohort, taken from 9,186 patients with NVAF on warfarin (median exposure 3.5 years), taken from the ATRIA study (USA). AF defined as any ICD-9 codes. Demographic data not given for validation cohort. No blinding or loss to FU reported.	154 MB	3 years
Fox 2017[36]	GARFIELD AF Risk HAS-BLED	VKA and DOAC	NR	25,285 patients with AF that were on OACs. 8804 on DOACs and 16,491 on VKAs. Details of the characteristics of these patients are not reported. No blinding reported.	625 MB	3 years
Friberg 2012[37]	HAS-BLED HEMORRHAGES	Warfarin	NR	48, 599 patients with AF (defined by ICD-10 code 1489 with or without subscales A-F) using Warfarin at baseline identified from the Swedish National Discharge Registry. Demographic data stated to be in supplementary file but not available in that file who were on warfarin. This subset was taken from an overall cohort of 170 291 which included those not on anticoagulants. No blinding reported.	1.9 MB per 100 patient years	1.5 years

Study	Risk tool(s)	OAC	Concomitant antiplatelets or NSAIDS	Population	Number and type of outcome events	Follow up duration
Gage 2006[38]	Landefeld and Goldman and Beyth et al. Kuijer et al. Kearon et al. HEMORRHAGES	Warfarin	7.40%	1604 medicare beneficiaries on NRAF (USA) with chart-confirmed AF on warfarin. 69.2% aged > 75 years, 7.9% hepatic or renal disease, 4.8% malignancy, 37.2% previous stroke, 0.4% uncontrolled hypertension. Also on Aspirin: 7.04%. No blinding or loss to FU reported.	4.9 MB per 100 patient years	Unclear but approx. 1 year
Gallego 2012[39]	HAS-BLED	Acenocoumarol	16.60%	965 consecutive anticoagulated people with permanent or paroxysmal AF, with at least 6 months of anticoagulation with acenocoumarol (INR 2-3). 50% male, mean age 76, hypertension 57%, DM 25.5%, HF 36.5%, prev. stroke/TIA 19%, renal impairment 10%, CAD 4%, hypercholesterolemia 31%, current smoking 14%, previous bleeding 8.5%, median HAS-BLED 2, CHADS2 score 2. Antiplatelet therapy 16.6%. 95 died during FU. No blinding reported.	75MB	861 days
Garcia-Fernandez 2017[41]	vWF HAS-BLED HAS-BLED + vWF	VKA	17.80%	1215 patients with NVAF on VKA at INR 2-3. Age 76, male 49.3%, hypertension 82.5%, DM 26.4%, HF 31.1%, IHD 19%, previous stroke 18.4%, previous bleeding 8.4%, renal disease 10.3%, antiplatelet drugs 17.8%, HAS-BLED score 2. No loss to FU or blinding reported.	222MB	2373 days
Hijazi 2014[56]	CHADSVASC CHADSVASC with TnT	apixaban and warfarin	28-34%	14,897 patients with AF on apixaban or warfarin, from the ARISTOTLE trial. Likely to be a multinational multi-centre trail but not reported. Ranges of baseline data given as data given for different categories of TnT. Age 64-74, male 53.8-74.6%, CHF 28-47%, hypertension 87%, DM 18-32%, Prior stroke/TIA 16-21%, MI 6-19%. Aspirin 28-34%. Warfarin 53.2-55.7%. BLINDED ASSESORS of BLEEDING. No loss to FU reported.	674 MB	1.9 years
Hijazi 2014[56]	HAS-BLED HAS-BLED with TnI	apixaban and warfarin	29-34%	14,821 patients with AF on apixaban or warfarin, from the ARISTOTLE trial. Overlap with Hijazi, 2014[57]in terms of sample, but this study used a different risk tool. Likely to be a multinational multi-centre trial but not reported. Ranges of baseline data given as data given for different categories of TnI. Age 66-72, male 6--70%, CHF 24-51%, hypertension 87%, DM 21-28%, Prior stroke/TIA 16-21%, MI 6-19%. Aspirin 29-34%. Warfarin 49.9-56.5%. BLINDED assessors. No loss to FU reported.	674 MB	1.9 years
Hijazi 2016[54]	HAS-BLED ORBIT ABC-bleeding ABC-bleeding (cTnI-hs)	warfarin and dabigatran (SEP ANALYSES)	44%	External validation in 8468 patients with AF (67% permanent or persistent) randomised to dabigatran and warfarin in the multinational RE-LY trial. Age 72, 26% women, 44% on antiplatelets or NSAISs, 8% current smokers, 22% DM, 79% hypertension, 29% CHF, 13% previous clinically relevant	463 MB	1.9 years

Study	Risk tool(s)	OAC	Concomitant antiplatelets or NSAIDS	Population	Number and type of outcome events	Follow up duration
	ABC-bleeding (cystatin C) ABC-bleeding (CKD-EPI)			bleeding, 19% previous stroke/TIA, 17% previous MI, 4% previous PAD, 19% vascular disease, Renal function CKD-EPI 68.2. ASSESSOR BLINDING. No loss to FU reported.		
Hijazi 2017[52]	HAS-BLED ORBIT (with or without GDF-15)	warfarin and dabigatran	36-41%	8,474 AF patients (with at least 1 additional risk factor for stroke) taken from the RE-LY study, on dabigatran or warfarin. Baseline characteristics given as ranges as sub-grouped by GDF-15. Age 69-75, male 61-67%, sbp 130, DM 11-35%, HF 25-34%, hypertension 78-80%, previous stroke/TIA 20-22%, prior MI 12-21%, prev PAD/MI/CAD 23-38%, aspirin 36-41%. CHADS2 >3 22-43%. No blinding/loss to FU reported.	458 MB	1.9 years
Hilkens 2017[58]	HEMORRHAGERS Shireman HAS_BLED ATRIA ORBIT (score) ORBIT (equation)	warfarin and dabigatran (SEP ANALYSES)	NR	3623 patients with AF on warfarin or dabigatran, from the RE-LY trial in Holland. No baseline data available. No report of blinding/loss to FU.	266 MB	2 years
Jaspers Focks 2016[63]	HAS-BLED ATRIA HEMORRHAGES	VKA	4.10%	1157 AF patients aged >80 years, using a VKA from 2011-2014 in the Netherlands. Median age 84, 42.6% male, 37 months on VKA, 65.8% hypertension, 22% previous stroke/TIA, 9.8% LVEF<40%, 26.6% CAD, 25.7% DM, 21.8% previous bleeding, 5.3% recent or active malignancy, 4.1% on antiplatelets and 2.1% on NSAIDS. HAS-BLED score 2.23. No blinding reported. 735 completed 3 year follow up (367 patients died and 55 patients moved out of the area or discontinued VKA treatment	77 MB	30 months
Jover 2012[65]	CHADSVASC	acenocoumarol	17%	933 patients with permanent or paroxysmal NVAF on acenocoumarol OAC (INR 2-3) for at least 6 months. Age 76, 46% male, 85% hypertension, 27% DM, 32% hypercholesterolemia, 14% current smokers, 39% CHF, 20% prior stroke/TIA, 20% CAD, 9% PAD, 17% on antiplatelets. CHADS2 score 2, CHADSVASC score 4. No blinding reported. No loss to FU reported.	80 MB	2.5 years
Lip 2011[71]	HAS-BLED Shireman HEMORRHAGE Beyth et al. Kuijer et al.	warfarin	NR	7,329 people with NVAF on warfarin or ximelagatran. Taken from the SPORTIF III and V cohorts (Multinational cohort). Following data are for those who developed a major bleed/no major bleed: age 73.9/70.9, female 31/31%, paroxysmal AF 11/12%, hypertension 77/77%, DM 29/23%, CAD 50/45%, LV dysfunction 44/36%, stroke/TIA 26/21%, CHADS 2.6/2.2.Blinded assessors.	136 MB	499 days

Study	Risk tool(s)	OAC	Concomitant antiplatelets or NSAIDS	Population	Number and type of outcome events	Follow up duration
Lip 2014[74]	SAME-TT2R2	VKAs	17%	4,637 patients with AF (n=572 had valvular AF) who were receiving OACs. FRANCE. Mean age 71, 35% female, 60% HF, 28% CAD, 12% previous MI, 6% previous CABG, 44% hypertensive, 9% previous stroke, 9% renal insufficiency. 17% on antiplatelets, 15% on Aspirin, 6% clopidogrel, 4% DAT. Mean CHADSVASC score 3.2, Mean HAS-BLED score 1.6. Not blinded.	144 MB	1016 days
Lip 2018[77]	HAS-BLED ATRIA ORBIT	DOACS	39.10%	57,930 patients with NVAF on DOACs. Taken from 3 Danish nationwide databases. Age 73.5, female 44.6%, HF 22.5%, DM 15.2%, Vascular diseases 16.2%, hypertension 59%, CPD 13.3%, prior bleeding 14.2%, kidney diseases 3.4%, Aspirin use 39.1%, NSAIDs 22.4%. Not blinded. Loss to FU not reported.	2.41 /100 person-years	1 year
Mori, 2019[88]	ORBIT HAS-BLED	DOACS	21.5%	2216 patients with NVAF using DOACs; 63.6% male; median age 73 years; median CHADS2 2; hypertension 73.5%; DM 27.9%; Dyslipidaemia 65.2%; eGFR 64.9; CAD 19.8%; PAD 7.1%; HF 23.7%; prior stroke 20.2%; prior bleeding 27.1%; antiplatelets 21.5%	93 MB	315 days
Nielsen 2016[90]	HAS-BLED Recalibrated HAS-BLED (2 points for previous haemorrhagic stroke instead of 1 point)	unclear	NR	Unknown number of OAC-treated patients from a cohort of 210,299 patients with AF taken from 3 Danish patient registries from 1999 to 2013. Demographic data for the sub-group having OACs is not reported	4.73 MB per 100 person years	Unclear
O'Brien 2015[91]	ORBIT HAS-BLED ATRIA-bleeding	rivaroxaban and warfarin	NR	14,264 patients with AF on either rivaroxaban (20mg daily) or Warfarin. This was the external validation cohort, comprising patients from the ROCKET-AF. Demographics of this external validation sample not reported.	772 MB	1.9 years
Olesen 2011[95]	HAS-BLED HEMORRHAGES	VKA	33%	44, 771 patients with AF receiving OACs in Denmark during 1997-2006. Demographic data given as two values as separate data for those with major bleeding / those without. Age 74.6 / 71.2, male 66.8 / 61.2 %, HASBLED score 2.5-2, HF 24.4/19.8%, hypertension 51.6/49.5%, DM 11.4/9.5%, Stroke 22.3/17.4, Renal disease 8.2/4.6%, Vascular disease 18.6/14.8%, Bleeding history 22.6/8.2%, antiplatelet drugs 33% / 25.5%, NSAIDs 22.8/19.1%.	2051 MB	1 year
Pisters 2010[103]	HAS-BLED HEMORRHAGES	Unspecified OACs	NR	1956 patients on OACs only with NVAF (validation cohort). Data not given for this validation cohort subset.	1.75 MB/100 patients years	1 year

Study	Risk tool(s)	OAC	Concomitant antiplatelets or NSAIDS	Population	Number and type of outcome events	Follow up duration
Poli 2017[110]	HAS-BLED HAS-BED (HAS-BLED but without labile INR score) CHADS2 CHADSVASC	warfarin and DOACs	16.50%	4579 patients with AF on DOACS (n=1048) or VKAs (n=3531) on START register in Italy. Age 76, 55% men, 15% HF, 80% hypertensive, 20% DM, 18% CAD, 6% PAD, 43% moderate renal impairment (eGFR 30-60 ml/min), 15% previous stroke/TIA, 3.4% history of major bleeding, TTR 67, concomitant antiplatelet drugs 16.5%, dual antiplatelet therapy 1.3%.	115 MB	1.4 years
Prochaska 2018[113]	HAS-BLED HAS-BLED with a point for sustained AF Simplified HAS-BLED	VKA - phenprocoumon	18.30%	1089 patients with medical and electrophysiological evidence of AF, and on VKAs, as part of the thrombEVAL cohort. Denmark. The following baseline data is separated into paroxysmal (n=398) and sustained (n=691) sub-groups by the paper: male 63/63%, age 72/75, DM 30/33%, Family history of MI/stroke 44.5/42%, hypertension 83/81.6%, CKD 24/27%, CAD 43.6/46.7%, HF 43.5/55.2%, history of major bleeding 6.8/6.2%, history of stroke/TIA 16.7/18.7%, MI 21.8/20.8%, PAD 16.1/17.5%, aspirin 18.3/15.1	150 CRB (includes MB and CRNMB)	3 years
Proietti 2016[116]	HAS-BLED ORBIT ATRIA HEMORRAGES ORBIT with TTR <65% (adding one point to score if <65%) ATRIA with TTR <65% (adding one point to score if <65%) HEMORRAGES with TTR <65% (adding one point to score if <65%)	warfarin	19.90%	3551 patients receiving warfarin in the pooled population dataset from the SPORTIF III and V studies with AF. De-identified datasets with patient-level information for the SPORTIF trials were obtained directly from Astra Zeneca, and all the analyses were performed independent of the company. All patients assigned to the warfarin treatment arms and with available data for the clinical variables used to calculate the four bleeding prediction scores were included in the present analysis. The majority of patients were male (69.5%) and the median [IQR] age was 72 [66–77] years. HAS-BLED score >3: 71%. 706/3551 (19.9%) treated concomitantly with aspirin. 20.1% VKA naïve at baseline prior to VKA initiation.	162 MB	1.6 years
Proietti 2018a[114]	HAS-BLED ORBIT ATRIA HEMORRHAGES	dabigatran 110mg, 150mg and warfarin (SEP ANALYSESfor C statistics but mixed for sensitivity/spe	40%	18,113 patients with AF on dabigatran (110 or 150 mg) or warfarin in the RE-LY trial. Multinational cohort. Age 72, 36% female, 79% hypertension, DM 23%, CAD 28%, prev stroke 22%, symptomatic HF 27%, VKA naïve 50%, anti-platelets 40%, CHADS2 2. BLINDED ASSESSORS.	1182 MB	2 years

Study	Risk tool(s)	OAC	Concomitant antiplatelets or NSAIDS	Population	Number and type of outcome events	Follow up duration
		cificity)				
Proietti 2018b[115]	HAS-BLED GARFIELD	warfarin	19.90%	3550 AF patients enrolled on the SPORTIF III trial who were on Warfarin. Age 72, 30.5% female, 76.7% hypertension, 23.5% DM, 44.3% CAD, 20.6% stroke/TIA, 37.3% HF, 5.6% previous bleeding, 25.9% CKD, 19.9% aspirin use. TTR 68.1. HAS-BLED: 3. 804 patients interrupted Warfarin during the follow up period. BLINDED ASSESSORS.	127 MB 168 major/CRNMB	1.56 years
Quinn 2016[117]	CHADS2 CHADSVASC ATRIA HAS-BLED	warfarin	NR	13,559 patients with AF who were on and off warfarin. No demographic data provided.	unclear	unclear
Rivera-Caravaca 2017[120]	HEMORRHAGES HAS-BLED ATRIA ORBIT	VKAs	18%	1361 patients – same patients as Roldan 2017[128]- with AF who were taking VKA OACs (acenocoumarol), in Spain. Age 76, 49% male, 82% hypertensive, 27% DM, 19% previous stroke/TIA, 19% CAD, 31% HF, 7% PAD, 10% renal impairment, 33% hypercholesterolemia, 8% previous bleeding episode, 4% alcohol abuse, 1% hepatic disease, 8% cancer. Median HAS-BLED score of 2	250 MB	6.5 years
Rivera-Caravaca, 2019[119]	HAS-BLED HAS-BLED with 1 to 6 added biomarkers	VKAs	18.4%	940 patients who were taking VKA OACs (IRR 2-3), in Spain. Age 76, 50.6% male, 82% hypertensive, 26.2% DM, 18.8% previous stroke/TIA, 19.8% CAD, 30.4% HF, 10.6% renal impairment, 33.3% hypercholesterolemia, Median HAS-BLED score of 2	172MB	6.5 years
Roldan 2013a[125]	HAS-BLED ATRIA	acenocoumarol	17%	937 consecutive patients with AF receiving anticoagulant therapy with INR from 2-3. 49% male, mean age 76, 82% hypertension, 25% DM, 37% HF, 19% stroke, 10% renal impairment, 19% CAD, 9% previous bleeding, 17% antiplatelet therapy. Median HAS-BLED score of 2, median CHADS2 score of 2.	79 MB	952 days
Roldan 2013b[126]	HAS-BLED CHADS CHADSVASC	acenocoumarol	18%	1370 consecutive patients with AF receiving anticoagulant therapy with INR from 2-3. 49% male, mean age 76, 19% stroke, 10% renal impairment, 18% CAD, 9% previous bleeding, 18% antiplatelet therapy. Median HAS-BLED score of 2, median CHADS2 score of 2.	114 MB	996 days
Roldan 2017[128]	HAS-BLED Modified HAS-BLED (including vWF, high sensitivity troponin T, N-	VKAs	18%	1361 consecutive patients with AF who were taking VKA OACs (acenocoumarol), in Spain. Age 76, 49% male, 82% hypertensive, 27% DM, 19% previous stroke/TIA, 19% CAD, 31% HF, 7% PAD, 10% renal impairment, 33% hypercholesterolemia, 8% previous bleeding episode, 4% alcohol abuse, 1% hepatic disease, 8% cancer. 18% antiplatelet	250 MB	7.49 years

Study	Risk tool(s)	OAC	Concomitant antiplatelets or NSAIDS	Population	Number and type of outcome events	Follow up duration
	terminal fragment B-type natriuretic peptide, high sensitivity IL-6, time in therapeutic range and modification of diet in renal disease CHADS-VASC Modified CHADSVASC (as above)			therapy. Median HAS-BLED score of 2		
Schwartz, 2019[135]	Modified HAS-BLED	VKAs and DOACS	NR	Data from 9819 patients with AF who were on DOACs or VKAs were retrieved from the Northwestern Healthcare system's Enterprise Database Warehouse. The data allowed identification of bleeding outcomes, and calculation of prior HAS-BLED 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	604 MB	971 days
Senoo 2016a[136]	HAS-BLED ORBIT	Idraparinux	NR	2283 patients with AF on non-warfarin OAC. UK. Age 70. No other details of demographics reported.	74 MB 346 CRB	311 days
Senoo 2016b[137]	HAS-BLED ORBIT ATRIA Also with TTR for NRI analysis of ORBIT and ATRIAS only	warfarin	16.50%	2293 patients with AF warfarin OAC. UK. Age 71, 65.5% male, paroxysmal AF 35.5%, persistent AF 9.3%, permanent AF 54.9%, hypertension 77%, HF 24%, DM 20%, CAD 31%, Stroke/TIA 25%, TTR 58%, **Aspirin 16.5%;NSAIDS 5.4%.**CHASVASC of 0-2: 28.8%, HAS-BLED 2.	39 MB 251 CRB	Unclear but probably < 1 year
Serna 2018[138]	HAS-BLED GEN /HAS-BLED (added point if patient carrying VKORC1 allele and CYP2C9*3 polymorphisms)	acenocoumarol (VKA)	NR	652 consecutive ASF patients stable on VKAs (INR 2-3) for 6 months. Spain. Age 76, 48.6% male, 82.8% hypertension, 24.2% DM, 18.7% history of stroke/TIA, 18.4% CAD, 31.9% hypercholesterolemia, 34.5% HF, 9.2% renal impairment, 1.5% hepatic impairment, 8.3% previous bleeding. HAS-BLED score 2. No data on antiplatelets.	106 MB	7.6 years

Study	Risk tool(s)	OAC	Concomitant antiplatelets or NSAIDS	Population	Number and type of outcome events	Follow up duration
Siu 2014[142]	HAS-BLED	warfarin	NR	1912 patients with NVAF (not defined) who received OACs (Warfarin). Mean age 73, 47% female, 55.8% hypertensive, 24% DM, 1.8% renal failure on dialysis, 24% HF, 24% CAD, 6.3% PAD, 29.6% prior stroke/TIA, prior IC haemorrhage 2.1%. Mean CHADSVASC 3.3. No data on antiplatelets	30 ICH	3.19 years
Steinberg 2016[146]	ATRIA HAS-BLED	warfarin and dabigatran	NR	7420 AF patients on OACs, out of an original cohort of 9715 from the ORBIT-AF trial. USA. Ranges for baseline data given as different data given for people in low, intermediate and high risk categories. Age 73-77, female 40-46%, hypertension 83-87%, diabetes 28-38%, previous GI bleed 5.7-16%, CAD 32-48%, Prior stroke/TIA 14-26%, CHF 30-46%, HAS-Bled 1.61-2.17, CHADS2 2.17-2.81. No data on antiplatelets.	632 MB	Unclear
Suzuki 2014[147]	HAS-BLED Modified HAS_BLED (renal dysfunction defined by eGFR <60, with exclusion of the 'elderly' factor because eGFR is calculated based on patient age)	warfarin	36.9-50%	231 NVAF patients on warfarin for at least 1 year. Demographics given as ranges as only reported for sub-groups of eGFR: age 68-74, 63.1-80% male, hypertension 53.2 to 64.4%, CAD 14.4 to 16.7%, CHF: 20 to 25.2%, dyslipidaemia 28.8 to 36.7%, eGFR 12.7 to 74.3 mL/min/1.73m2) antiplatelet drugs 36.9 to 50%. TTR 56.9 to 65.1%.	44 MB	7.1 years
Wang 2016[154]	HAS-BLED	dabigatran and warfarin (SEP ANALYSES)	NR	21,934 adults with AF who were starting dabigatran (30%) or Warfarin. Patients were on a healthcare claims database in USA. Demographic data given for those on Warfarin (n=15418): Age 65, female 34%, 27% CHF, 31% DM, 93% hypertensive, 20% prior stroke, 22% PVD. 43% with HAS-BLED score of 3 or more. 32% with CHADS2 score of 3 or more.	4.6 MB per 100 patient years	5 months
Yao 2017[158]	CHADSVASC CHADS HAS-BLED ORBIT ATRIA	DOACS (results not sub-grouped)	7%	39, 539 patients with NVAF from USA insurance database (OptumsLabs Data Warehouse) who had started DOACs between 2010 and 2015. Age 71, 42% female, 20% non-white, 28% HF, 86% hypertension, 34% DM, 14% previous strokes/TIA, 48% vascular disease, 7% stage II or IV CKD, 4% abnormal liver function, 9% previous major bleeding, 7% using antiplatelets, 5% using NSAIDs, 28% had had previous warfarin exposure. HAS-BLED: 2	115 MB	0.6 years

MB=major bleeding, CRB= clinically relevant bleeding, CRNMB= clinically relevant non-major bleeding, ICH= Intracranial hemorrhage

Table 5: Summary of risk tools and their constituent variables

Risk tool	Variables and scoring	Bleeding risk interpretation (where applicable)
ABC-bleeding	Prior bleeding, age, hs-troponin, GDF-15 and Hb. Continuous values inputted (where appropriate) and a probability score derived by algorithm.	Score is the 1 year risk of major bleeding
ABC bleeding CrC	ABC-bleeding with creatinine clearance replacing GDF-15	
ABC-bleeding CKD-EPI	ABC-bleeding with CKD-EPI biomarker added to the scheme	
ABC-bleeding cTnI-hs	ABC-bleeding with cTnI-hs biomarker added to the scheme	
ABC-bleeding cystatin C	ABC-bleeding with cystatin C biomarker added to the scheme	
Anticoagulation-specific Bleeding Score (ABS)	The 1-year risk of bleeding can be calculated as 1 - (0.98101) Exp[0.02306(Age - 70.1736) + 0.29958(Kidney Disease -0.13244) + 0.19215(COPD -0.31286)+ 0.23529(Prior Bleed -0.21338) +0.32257(Anemia -0.24892) + 0.21811(Heart Failure-0.33899)+ 0.22599(Antiplatelet-0.16341) + 0.15944 (Diuretics-0.4518) + 0.2111(Diabetes Mellitus-0.31686) + 0.16806 (Cancer-0.16955) - 0.28572 (Antiarrhythmic -0.11919) + 0.13743(Ischemic stroke - 0.26681) + 0.10269(Coronary Artery Disease -0.40768) - 0.04775(Male Sex-0.59637) -0.30127 (Dabigatran) + 0.01299(Rivaroxaban) - 0.52426(Apixaban)]	1 year risk of bleeding yielded
APTT	Biomarker: activated partial thromboplastin time	No pre-set thresholds provided in paper
ATRIA	Anaemia (3 points), severe renal disease (eGFR <30) (3 points), age >75 years (2 points), any prior bleeding (1 point), hypertension history (1 point)	Low: 0-3 Moderate: 4 High: 5 or more
ATRIA with AS	ATRIA with existence of aortic stenosis added inas a risk factor to the scheme	
ATRIA with TTR (<65% TTR)	ATRIA with time in therapeutic range of <65% added in as a risk factor to the scheme	
Beyth	See mOBRI	
CBRM	See Shireman	
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
CHADSVASC	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

Risk tool	Variables and scoring	Bleeding risk interpretation (where applicable)
CHADSVASC with TnT	CHADSVASC with TnT levels added in to the scheme	
GARFIELD/ GARFIELD AF	Age, pulse, systolic blood pressure, history of vascular disease, history of bleeding, heart failure, renal disease and use of OACs.	Score is a measure of bleeding risk
GDF-15	Biomarker: levels of Growth Differentiation Factor 15	
GEN/HAS-BLED	HAS-BLED with added point if patient carrying VKORC1 allele and CYP2C9*3 polymorphisms	
HAS-BED	HAD-BLED with elimination of labile INR factor.	
HAS-BLED	Hypertension, abnormal renal/liver function (1 point each), stroke, bleeding history or predisposition, labile INR, elderly drugs/alcohol concomitantly (1 point each). Maximum 9 points	Low: 0 Moderate: 1-2 High: 3 or more
HAS-BLED with AS	HAS-BLED with existence of aortic stenosis added inas a risk factor to the scheme	
HAS-BLED with GDF-15	HAS-BLED with GDF biomarker added to the scheme	
HAS-BLED with point for sustained AF	HAS-BLED with additional factor of 'sustained AF in the presence of HF'.	
HAS-BLED with TnI	HAS-BLED with TnT levels added in to the scheme	
HAS-BLED with VWF	HAS-BLED with Van Willebrandlevels added into the scheme	
HAS-BLED with no labile INR and no stroke/TIA component	HAS-BLED with no labile INR and no stroke/TIA component	
HAS-BLED + VWF + NT-proBNP	HAS-BLED with Van Willebrand levels and N-terminal pro-B-type natriuretic peptide added into the scheme	
HAS-BLED + VWF + NT-proBNP + IL-6	HAS-BLED with Van Willebrand levels and N-terminal pro-B-type natriuretic peptide and Interleukin-6 added into the scheme	
HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T	HAS-BLED with Van Willebrand levels and N-terminal pro-B-type natriuretic peptide and Interleukin-6 and Troponin T added into the scheme	
HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T + BTP	HAS-BLED with Van Willebrand levels and N-terminal pro-B-type natriuretic peptide and Interleukin-6 and Troponin T and Beta trace protein added into the scheme	
HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T + BTP + soluble fibrin monomer complex	HAS-BLED with Van Willebrand levels and N-terminal pro-B-type natriuretic peptide and Interleukin-6 and Troponin T and Beta trace protein and soluble fibrin monomer complex added into the scheme	

Risk tool	Variables and scoring	Bleeding risk interpretation (where applicable)
HEMORRHAGES	Hepatic or renal disease (1 point) Ethanol abuse (1 point)* Malignancy (1 point) Older age >75 yrs (1 point) Reduced platelet count or function (1 point) Re-bleeding risk (2 points) Hypertension (1 point) Anaemia (1 point) Genetic factors (1 point) Excessive fall risk or neuropsychiatric disease (1 point) Stroke (1 point)	Low: 0-1 Intermediate: 2-3 High: 4 and above
HEMORRHAGES with TTR (<65% TTR)	HEMORRHAGES with time in therapeutic range of <65% added in as a risk factor to the scheme	
HTI	Biomarker: Hemoclot thrombin inhibitor levels	No pre-set thresholds provided in paper
Kearon 2003	Age >65yrs (1 point) Prior stroke (1 point) Prior peptic ulcer disease (1 point) Prior GI bleeding (1 point) Creatinine >1.5 mg/dl (1 point) Anemia or thrombocytopenia (1 point) Liver disease (1 point) Diabetes mellitus (1 point) Antiplatelet therapy (1 point)	Low: 0-1 Intermediate:2 High 3 or more
Kuijer 1999	Age >60 yrs (1.6 points) Female (1.3 points) Malignancy (2.2 points)	Low: 0 Intermediate 1-2 High 3 or more
Landefield and Goldman and Beyth	See mOBRI	

Risk tool	Variables and scoring	Bleeding risk interpretation (where applicable)
MBRFS	See MBR	
mOBRI (also known as Landefield and Goldman and Beyth, or simply Beyth)	Age > 65 years, GI bleed in past 2 weeks, previous stroke, comorbidities (recent MI, Hct <30%, diabetes, creatinine >1.5 ml/l) with 1 point for presence of each risk factor	Low: 0 Moderate; 1-2 High: 3 or more
MBR (Modifiable Bleeding Risk factors score)	Defined as the cumulative number of modifiable bleeding risk factors of each patient according to the 2016 ESC guideline, including hypertension, medication predisposing to bleeding, and excess alcohol. 1 point for each.	Score ranges from 0-3.
Modified CHADSVASC	CHADSVASC with vWF, high sensitivity troponin T, N-terminal fragment B-type natriuretic peptide, high sensitivity IL-6, time in therapeutic range and modification of diet in renal disease	
Modified HAS-BLED (multiple additions using biomarkers)	HAS-BLED with addition ofvWF, high sensitivity troponin T, N-terminal fragment B-type natriuretic peptide, high sensitivity IL-6, time in therapeutic range and modification of diet in renal disease	
Modified HAS-BLED (single change of renal dysfunction threshold)	HAS-BLED with modification of the renal impairment factor (from eGFR <30 to eGFR <60)	
ORBIT	Older age (75 years and above) (1point), reduced hemoglobin, hematocrit, or history of anemia (2 points), bleeding history: (2 points), insufficient kidney function (eGFR below 60 mL/min/1.73 m2)(1 point), treatment with an antiplatelet agent (1 point).	Low: 0-2 Moderate:3 High: 4 or more
ORBIT with AS	ORBIT with existence of aortic stenosis added inas a risk factor to the scheme	
ORBIT with GDF-15	ORBIT with GDF-15 levels added into the scheme	
ORBIT with TTR (<65% TTR)	ORBIT with time in therapeutic range of <65% added in as a risk factor to the scheme	
Outpatient bleeding Index (OBI)	Age >65 yrs (1 point) Prior stroke (1 point) Prior GI bleeding (1 point) Recent MI, diabetes mellitus, hematocrit <30%, creatinine >1.5 mg/dl (1 point if any of the above)	Low: 0 Intermediate 1-2 High 3 or more
Prothrombin time	Biomarker: Prothrombin time	No pre-set thresholds provided

Risk tool	Variables and scoring	Bleeding risk interpretation (where applicable)
		in paper
Riete	Recent major bleeding (□15 days before thrombotic event) (2 points) Creatinine >1.2 mg/dl (1.5 points) Anemia (1.5 points) Malignancy (1 point) Clinically overt pulmonary embolism (1 point) Age >75 yrs (1 point)	Low: 0 Intermediate: 1-4 High: >4
Same TTR	Sum of points after addition of one point for female sex, age <60 years, medical history of >2 comorbidities (amongst hypertension, DM, CAD/MI, PAD, CHF, previous CVA, pulmonary disease and hepatic/renal disease, treatment and 2 points each for smoking and non-white race.	Low:0-1 Moderate: 2 High >2
Shireman 2006 (also known as CBRM)	Age >70 yrs Female Remote bleeding event Recent bleeding event Alcohol or drug abuse Diabetes mellitus Anemia (Hct <30% during index hospitalization) Antiplatelet drugs (aspirin, clopidogrel, or ticlodipine at discharge) Risk score = 0.49 (age >70) + 0.32 (female) + 0.58 (remote bleed) + 0.62 (recent bleed) + 0.71 (alcohol/drug abuse) + 0.27 (diabetes) + 0.86 (anemia) + 0.32 (antiplatelet use)	Low <1.07 Intermediate >1.07, <2.19 High >2.19
Simplified HAS-BLED	HAS-BLED, containing only the factors of age >65 years, history of major bleeding, and sustained AF in the presence of heart failure	
TnI	Biomarker: Troponin I levels	
TnT	Biomarker: Troponin T levels	
vWF	Biomarker: levels of plasma glycoprotein von Willebrand factor	

2.3.1 Discriminationfor MAJOR BLEEDING

Table 6: Clinical evidence profile: accuracy of prediction of Major Bleedingin all risk tools featured in the studies (see table 3). Outcomes split across subgroups are only shown if sub-grouping was able to reduce $I^2$to <50% in all sub-groups.

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
HAS-BLED	47	532,442	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**POOLED RESULT: Random effect: 0.62 (0.61-0.64) [I^2=94%]**	VERY LOW
Modified HASBLED[135]	1	9819	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.60(0.55-0.66)('Non-white' participants) 0.57(0.55-0.60) ('white' participants)	VERY LOW
HAS-BLED with AS	1	2880	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.68(0.66-0.70)	MODERATE
HAS-BLED with GDF-15	1	8474	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.69(0.67-0.72)	VERY LOW
HAS-BLED with vWF	2	1215	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	**POOLED RESULT: Fixed effect: 0.62 (0.60-0.64) [I^2=6%]**	MOD
HAS-BLED + VWF + NT-proBNP	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.64(0.61-0.67)	MOD
HAS-BLED +	1	940	Serious risk	No serious	No	No serious	0.64(0.61-0.67)	MOD

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
VWF + NT-proBNP + IL-6			of bias[a]	inconsistency	serious indirectness	imprecision		
HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.64(0.61-0.67)	MOD
HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T + BTP	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.64(0.60-0.67)	MOD
HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T + BTP + soluble fibrin monomer complex	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.64(0.60-0.67)	MOD
GEN/HAS-BLED	1	652	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.65(0.61-0.68)	MOD
Modified HAS-BLED (multiple additions using biomarkers)	1	1361	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.60(0.56-0.64)	MOD
Modified HAS-	1	231	Very serious risk	No serious inconsisten	No serious	Serious imprecision[c]	0.67(0.57-0.75)	VERY LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
BLED (single change of renal dysfunction threshold)			of bias[a]	cy	indirectness			
HAS-BED	1	4579	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.58(0.53-0.64)	LOW
HAS-BLED with TnI	1	14,821	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.63	LOW
HEMORRHAGES	19	240,995	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**POOLED RESULT: Random effect: 0.63 (0.60-0.66) [I^2=97%]**	VERY LOW
HEMORRHAGES with TTR (<65% TTR)	2	4912	Serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**Median: 0.65**	VERY LOW
ATRIA	23	286,664	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**POOLED RESULT: Random effect: 0.64 (0.61-0.66) [I^2=97%]**	VERY LOW
ATRIA with AS	1	2880	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.67(0.66-0.69)	MODERATE
ATRIA with	2	4912	Serious risk	Very	No	No serious	**Median: 0.68**	VERY LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
TTR (<65% TTR)			of bias[a]	serious risk of incon-sistency[b]	serious indirectness	imprecision		
ORBIT	21	270,606	Very serious risk of bias[a]	Very serious riskof incon-sistency[b]	No serious indirectness	No serious imprecision	**POOLED RESULT: Random effect: 0.64 (0.61-0.67) [I^2=97%]**	VERY LOW
ORBIT with AS	1	2880	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.68(0.67-0.70)	MODERATE
ORBIT with TTR (<65% TTR)	2	4912	Serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**Median: 0.67**	VERY LOW
ORBIT with GDF-15	1	8474	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.71(0.68-0.73)	VERY LOW
CHADS2	5	61,647	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**POOLED RESULT: Random effect: 0.61 (0.57-0.64) [I^2=85%]**	VERY LOW
CHADSVASC	8	24,402	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**POOLED RESULT: Random effect: 0.59 (0.54-0.64) [I^2=92%]**	VERY LOW
Modified CHADSVASC	1	1361	Serious risk of bias[a]	No serious inconsisten	No serious	No serious imprecision	0.56(0.53-0.60)	MOD

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
				cy	indirectness			
CHADSVASC with TnT	1	14,897	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.63(0.61-0.65)	LOW
GARFIELD	3	62,172	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**Pooled effect: Random effects 0.60 (0.56-0.65); I2=96%**	VERY LOW
GARFIELD subgrouped by OAC - VKA	1	3550	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.56(0.54-0.58)	LOW
GARFIELD subgrouped by OAC – Mixed VKA/DOACs	1	7442	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.61(0.59-0.63)	LOW
GARFIELD subgrouped by antiplatelets - <33% with antiplatelets	1	3550	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.56(0.54-0.58)	LOW
GARFIELD subgrouped by antiplatelets – unknown % with	1	7442	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.61(0.59-0.63)	LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
antiplatelets								
ABC-bleeding	3	16869	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	Serious imprecision[c]	**POOLED RESULT: Random effect: 0.69(0.65-0.74) [I^2=85%]**	VERY LOW
ABC-bleeding Subgrouped by OAC - VKA	1	2814	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.65(0.61-0.70)	VERY LOW
ABC-bleeding Subgrouped by OAC - Mixed	1	8705	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.69(0.66-0.71)**[Mixed]**	VERY LOW
ABC-bleeding Subgrouped by OAC - NOACs	1	5350	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.74(0.71-0.76)**[DOAC]**	VERY LOW
ABC-bleeding CrC	1	1120	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.52(0.49-0.55)	LOW
ABC-bleeding cTnI-hs	2	8164	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	Serious imprecision[c]	**POOLED RESULT: Random effect: 0.70 (0.61-0.78) [I2=92%]**	VERY LOW
ABC-bleeding cTnI-hs subgrouped by OAC - VKA	1	2814	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.65(0.61-0.70	VERY LOW
ABC-bleeding	1	5350	Very	No serious	No	No serious	0.74(0.71-0.76)	LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
cTnl-hs subgrouped by OAC - DOAC			serious risk of bias[a]	inconsistency	serious indirectness	imprecision		
ABC-bleeding cystatin C	2	8164	Very serious risk of bias[a]	Very serious risk of inconsistency[b]	No serious indirectness	Serious imprecision[c]	**POOLED RESULT: Random effect: 0.68 (0.65-0.72) [I2=90.6%]**	VERY LOW
ABC-bleeding cystatin C subgrouped by OAC - VKA	1	2814	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.60(0.54-0.66)	LOW
ABC-bleeding cystatin C subgrouped by OAC - DOAC	1	5350	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.72(0.68-0.75)	VERY LOW
ABC-bleeding CKD-EPI	2	8164	Very serious risk of bias[a]	Very serious risk of inconsistency[b]	No serious indirectness	Serious imprecision[c]	**POOLED RESULT: Random effect: 0.70 (0.68-0.72) [I2=79%]**	VERY LOW
ABC-bleeding CKD-EPI subgrouped by OAC - VKA	1	2814	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.65(0.60-0.69)	LOW
ABC-bleeding CKD-EPI subgrouped by OAC - DOAC	1	5350	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.71(0.69-0.74)	VERY LOW
vWF	1	1215	Serious risk	No serious	No	No serious	0.61(0.57-0.65)	MOD

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
			of bias[a]	inconsistency	serious indirectness	imprecision		
ABS	1	81285	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.67(0.65-0.68)[warfarin], 0.72(0.69-0.76)[dabigatran] 0.70(0.68-0.73)[rivaroxaban] 0.72(0.67-0.77) [apixaban]	VERY LOW
OBI	1	3063	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.59(0.58-0.611	LOW
Kuijer	3	8332	Very serious risk of bias[a]	Serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**POOLED EFFECT: Random effects: 0.54 (0.51-0.58) [I^2=72%]**	VERY LOW
Kearon	2	4667	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	**Median: 0.675**	LOW
Riete	1	3063	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.68(0.65-0.70)	LOW
Shireman / CBRM	5	12385	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**POOLED EFFECT: Random effect: 0.64(0.59-0.69) [I^2=80%]**	VERY LOW
mOBRI/Landefield and Goldman and Beyth / Beyth	3	8762	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	**POOLED EFFECT: Fixed effect: 0.56(0.51-0.60) [I^2=0%].**	LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
TnT	1	14,897	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.62(0.60-0.64)	LOW
TnI	1	14,821	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.60	LOW
GDF-15	1	8474	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.67(0.65-0.69)	LOW
MBR	1	40,450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.53(0.52-0.53)	LOW
HTI	1	208	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.65	LOW
Prothrombin time	1	208	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.54(0.47-0.62)	VERY LOW
Same TTR	1	4637	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.55 (0.54-0.57)	LOW
APTT	1	208	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.58(0.50-0.69)	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 fewof the studiesreported 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 manystudieswith 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 studiesmay 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 around two clinical thresholds: C statisticsof 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 7: Clinical evidence profile: sensitivity and specificityof prediction of Major Bleeding in all risk tools featured in the studies (see table 3). 95% CIs are given for non-pooled results; for meta-analysed results the 95% credible intervals are given for the pooled effect only.

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
HAS-BLED at threshold of ≥1	7	128791	**Pooled sensitivity: 0.979(0.941-0.993)**	**Pooled specificity: 0.070(0.027-0.174)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency [b]	No serious indirectness	No serious imprecision	VERY LOW
					Specificity				

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					Very serious risk of bias[a]	Serious inconsistency [b]	No serious indirectness	Serious imprecision[c]	VERY LOW
HAS-BLED at threshold of ≥2	10	177728	**Pooled sensitivity: 0.793(0.570-0.919)**	**Pooled specificity: 0.396(0.207-0.624)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency [b]	No serious indirectness	Very serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency [b]	No serious indirectness	Serious imprecision[c]	VERY LOW
HAS-BLED at threshold of ≥3	13	170197	**Pooled sensitivity: 0.512(0.385-0.637)**	**Pooled specificity: 0.679(0.554-0.782)**	**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

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
HAS-BLED at threshold of ≥4	1	3525	0.543(0.453-0.632)	0.591(0.575-0.608)	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
Modified HASBLED[135] at threshold of ≥1	1	9819	0.925 (0.902-0.945)	0.1504(0.143-0.158)	**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
Modified HASBLED[135] at threshold of ≥2	1	9819	0.644(0.604-0.682)	0.4937(0.483-0.5040	**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
Modified	1	9819	0.311(0.275-0.349)	0.826(0.819-0.834)	**Sensitivity**				

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
HASBLED[135]at threshold of ≥3					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
HEMORRHAGES at threshold of ≥1	3	7406	**Pooled sensitivity: 0.919(0.658-0.985)**	**Pooled specificity: 0.167(0.037-0.5207)**	**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[a]	No serious indirectness	Serious imprecision[c]	VERY LOW
HEMORRHAGES at threshold of ≥2	6	60023	**Pooled sensitivity: 0.631(0.417-0.798)**	**Pooled specificity: 0.549(0.349-0.734))**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					Very serious risk of bias[a]	Serious inconsistency [a]	No serious indirectness	Serious imprecision[c]	VERY LOW
HEMORRHAGES at threshold of ≥3	2	5138	0.478(0.354-0.603) 0.171 (0.112-0.250)	0.739(0.716-0.761) 0.886(0.874-0.896)	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency [a]	No serious indirectness	No serious imprecision	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency [a]	No serious indirectness	No serious imprecision	VERY LOW
ATRIA at threshold of ≥1	4	103289	**Pooled sensitivity: 0.955(0.864-0.986)**	**Pooled specificity: 0.132(0.061-0.259)**	**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	Serious imprecision[c]	VERY LOW
ATRIA at	5	103289	**Pooled sensitivity:**	**Pooled specificity:**	**Sensitivity**				

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
threshold of >2			**0.685(0.450-0.848)**	**0.539(0.354-0.716)**	Very serious risk of bias[a]	Serious inconsistency [a]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	VERY LOW
ATRIA at threshold of ≥3	3	101023	**Pooled sensitivity: 0.571(0.212-0.856)**	**Pooled specificity: 0.638(0.35446-0.861)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency [a]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency [a]	No serious indirectness	No serious imprecision	VERY LOW
ATRIA at threshold of ≥4	6	111338	**Pooled sensitivity: 0.259(0.096-0.513)**	**Pooled specificity: 0.874(0.714-0.941)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency [a]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency [a]	No serious indirectness	No serious imprecision	VERY LOW

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
ORBIT at threshold of ≥1	4	103302	**Pooled sensitivity: 0.804(0.610-0.916)**	**Pooled specificity: 0.381(0.217-0.574)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Very serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Serious imprecision[c]	VERY LOW
ORBIT at threshold of ≥2	4	103302	**Pooled sensitivity: 0.460(0.233-0.692)**	**Pooled specificity: 0.716(0.528-0.849)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Serious imprecision[c]	VERY LOW
ORBIT at threshold of ≥3	8	114895	**Pooled sensitivity: 0.340(0.213-0.493)**	**Pooled specificity: 0.845(0.766-0.900)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	No serious imprecision	VERY LOW
					Specificity				

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					Very serious risk of bias[a]	Serious inconsistency [a]	No serious indirectness	No serious imprecision	VERY LOW
CHADS2 at threshold of ≥1	1	39539	0.991(0.981-0.998)	0.084(0.081-0.086)	**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
CHADS2 at threshold of ≥2	1	39539	0.865(0.836-0.889)[]	0.341(0.336-0.346)	**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
CHADS2 at threshold of ≥3	1	39539	0.552(0.513-0.590)	0.776(0.775-0.779)	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very	NA	No serious	No	LOW

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					serious risk of bias[a]		indirectness	serious imprecision	
CHADSVASC at threshold of ≥1	1	39539	0.998(0.992-1.00)	0.385(0.366-0.404)	**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
CHADSVASC at threshold of ≥2	1	39539	0.984(0.970-0.992)	0.129(0.125-0.132)	**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
CHADSVASC at threshold of ≥3	1	39539	0.929(0.907-0.948)	0.271(0.267-0.276)	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
					Specificity				
					Very serious	NA	No serious indirectness	No serious	LOW

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					risk of bias[a]		s	imprecis on	
ABC-bleedingCrCat threshold of ≥2%	1	1120	0.835(0.778-0.884)	0.194(0.169-0.221)	**Sensitivity**				
					Serious risk of bias[a]	NA	No serious indirectnes s	No serious imprecis on	LOW
					Specificity				
					Serious risk of bias[a]	NA	No serious indirectnes s	No serious imprecis on	LOW
HTI**at** threshold >117 ng/ml	1	208	0.59[no raw data or 95% Cis reported in paper]	0.71[no raw data or 95% Cis reported in paper]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectnes s	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NAS	No serious indirectnes s	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.Subgrouping to attempt to resolve heterogeneity was not carried out because there would always be <3 studies in any of the constituent sub-group categories, making it not possible to do a further meta-analysis within each sub-group.

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.

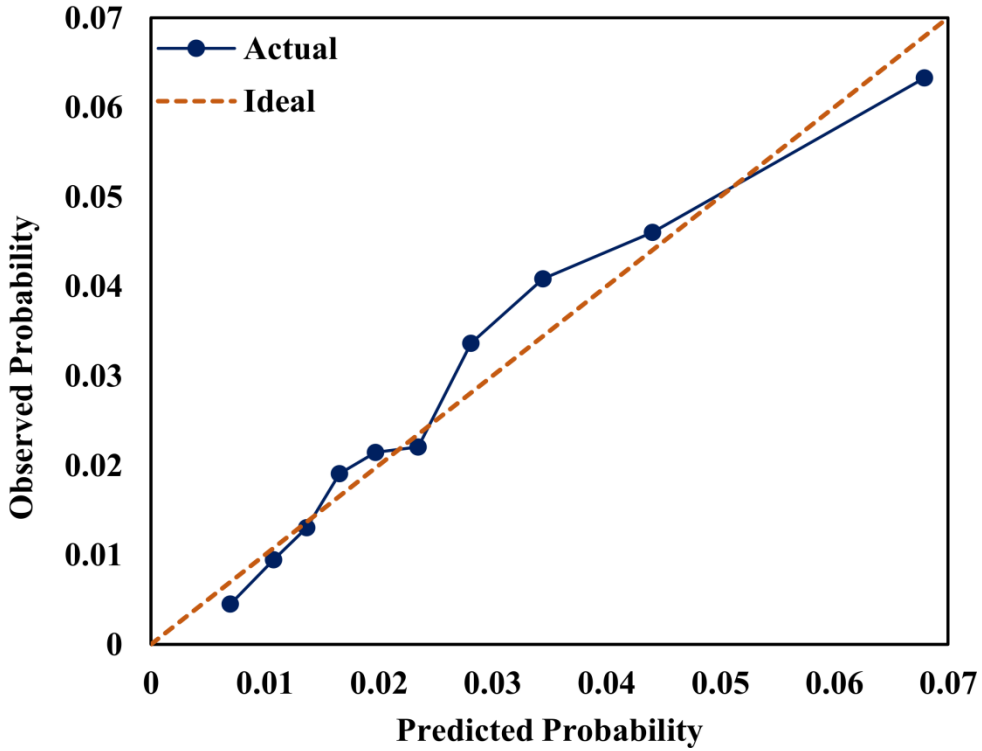
2.3.2 Calibrationfor MAJOR BLEEDING

Calibration waspredominantlyreportedwith graphical rather than numerical data. Hence this section has been dealt with narratively.

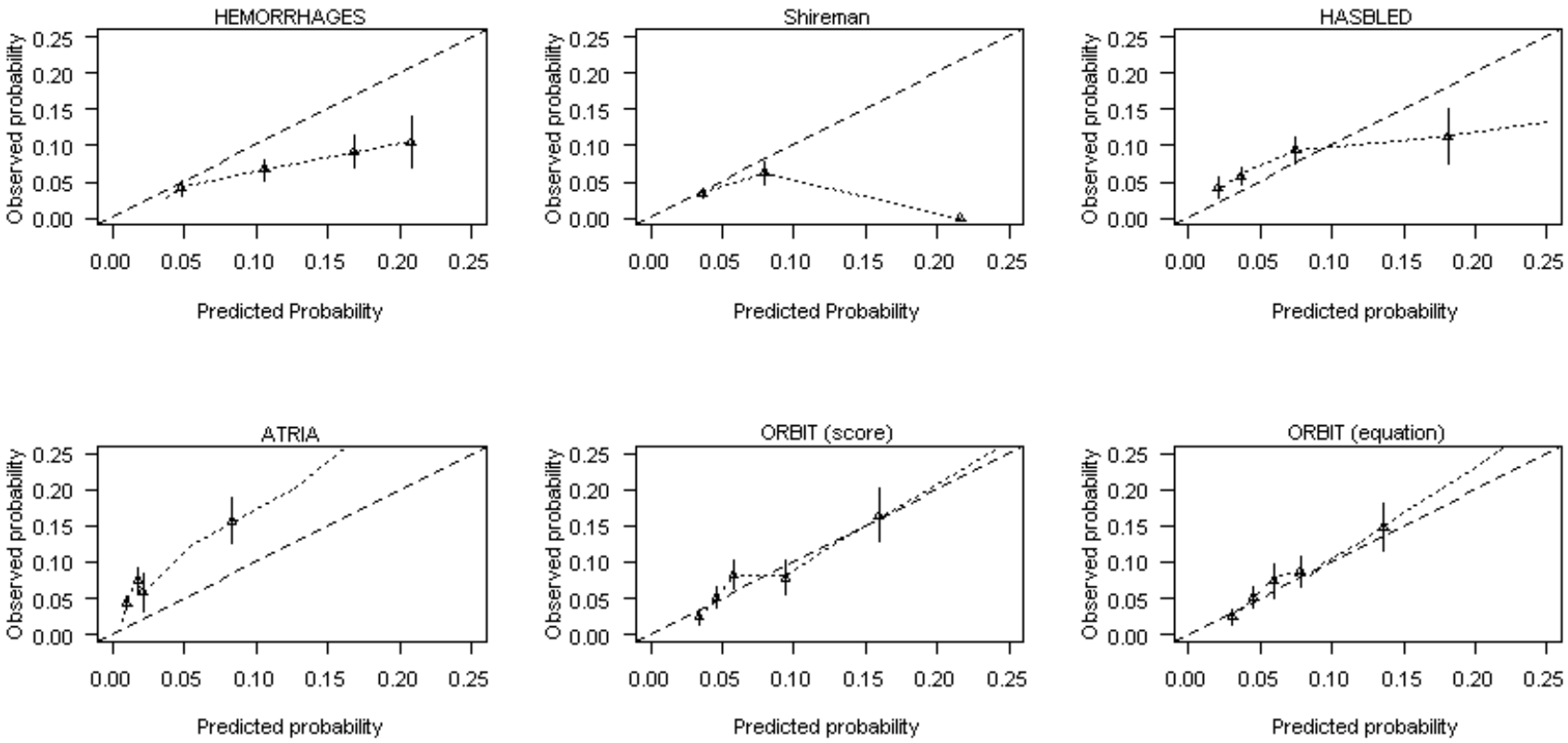
Several studies merely reported a non-comparative'adequate'calibration, usuallybased on a Hosmer-Lemeshow p value >0.05. 'Adequate' goodness of fit was thus described for ATRIA[4, 14, 63], HAS-BLED[4, 14, 63, 71], HEMORRHAGES[4, 14, 63, 71], ORBIT[14], Shireman[71], mOBRI/Beyth[71], Kuijer[71]and ABC[11, 23, 54]. It was not possible, based on these data, to compare thelevels of calibration acrossthese tools.

However, some studies performed a relative, albeit qualitatively described,evaluation, which was based on inspection of calibration plots. Hilkens, 2017[58]stated that ORBIT had a better calibration at 2 years than HEMORRHAGES, ATRIA, Shireman and HAS-BLED. ORBIT was also regarded as better calibrated than HAS-BLED and ATRIA by fourfurther studies,[77, 91, 114, 158]although Mori, 2019[88]did not note a difference.ATRIA was identified as the least well-calibrated by twoof the studies[91, 158]but better than HAS-BLED by one[114]. Proietti 2018[114]noted that whilst ORBIT had the best calibration over all risk strata, HEMORRHAGES tended to underestimate risk, particularly in patients with a higher predicted risk, whereas ATRIA and HAS-BLED tended to over-estimate bleeding risk. Similarly, O'Brien[91]noted that whilst ORBIT was good at predicting risk in all risk strata, HAS-BLED tended to have worse calibration in low-risk strata, and ATRIA performed badly at mostrisk strata. Claxton, 2018[23]evaluated the calibration of the Anticoagulation-specific bleeding score (ASBS) alone, demonstrating good calibration. Calibration plots are shown below.

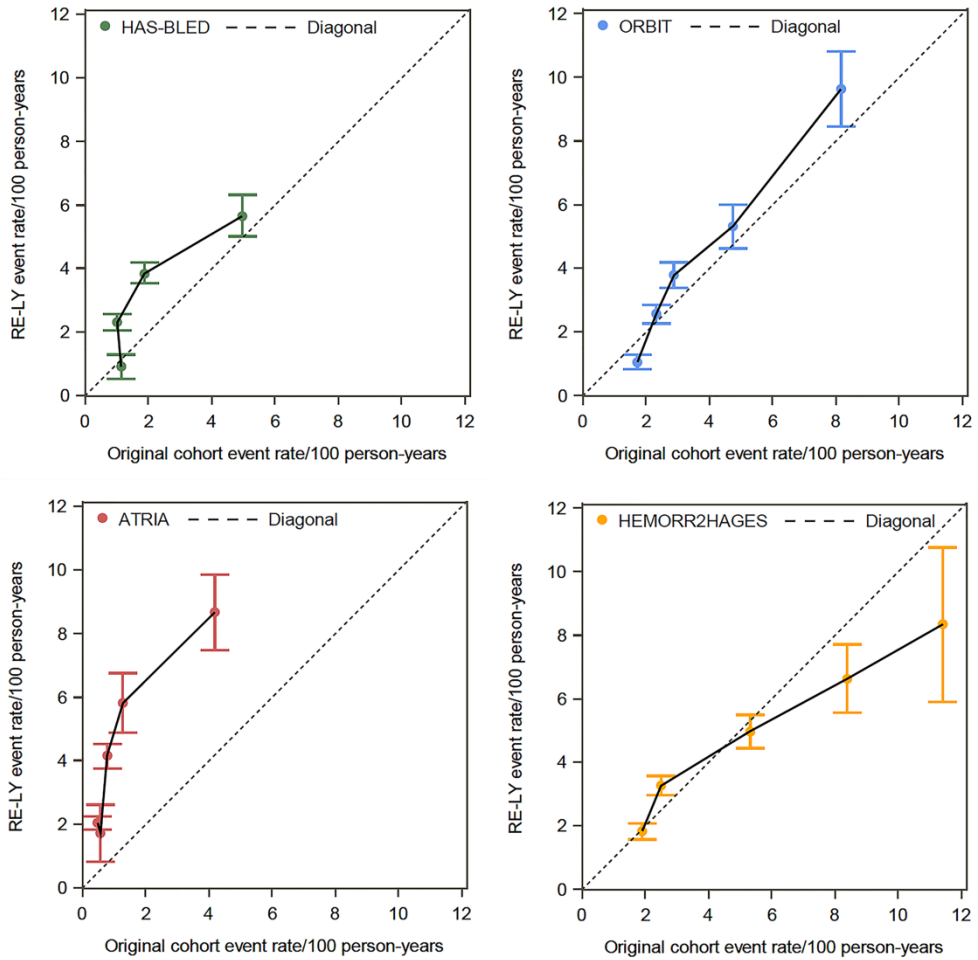
Note that Lip, 2018[77], Mori, 2019[88]and Yao, 2017[158]only used DOACcohorts, but O'Brien, 2015[91]and Claxton, 2018[23]used a mixed cohort. Both Hilkens, 2017[58]and Proietti, 2018[114]contained separate cohorts of patients taking dabigatran and warfarin, but it appears that the plots reproduced below were from their total, mixed, cohort. It should also be noted that Proietti 2018[114]failed to specify if calibration data referredto major bleeding, although major bleedingis assumedto be the most likely bleeding



Source: Calibration plot in Claxton, 2018[23]. This was for the Anticoagulation-specific bleeding score and was based on a mixed (VKA and DOAC) cohort.

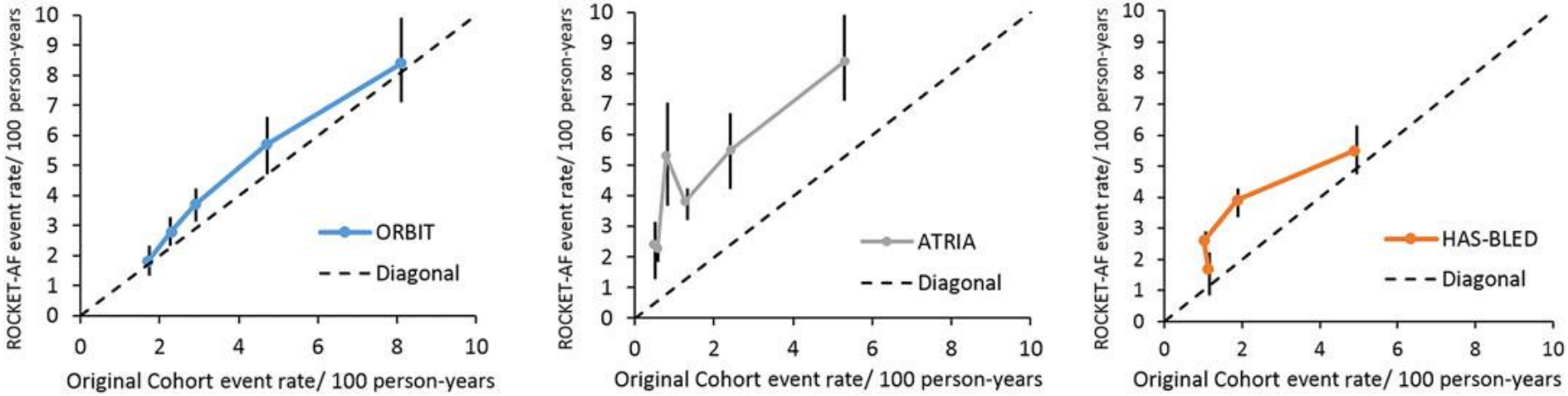


Source: Calibration plot in Hilkens, 2017[58]. This was based on a mixed (VKA and DOAC) cohort.

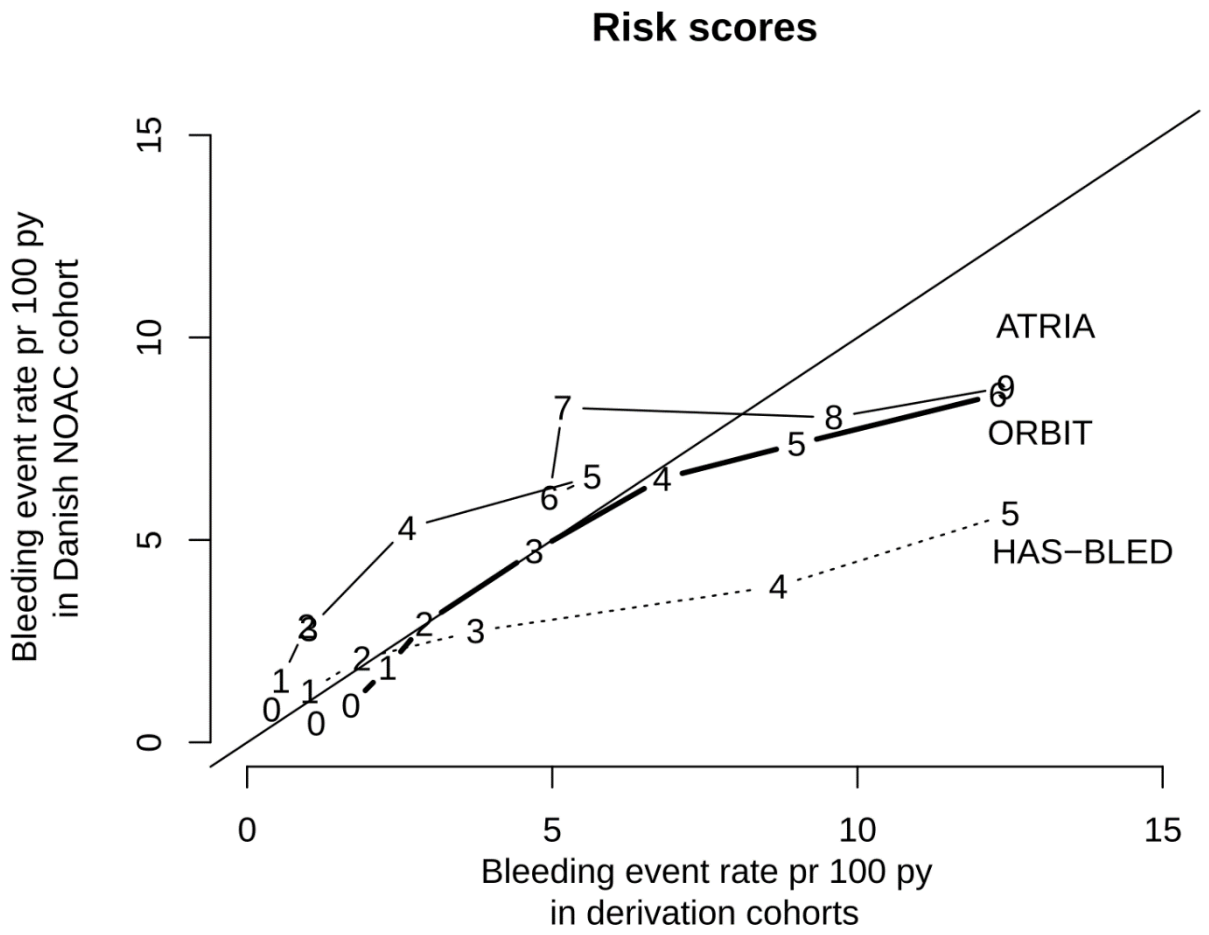


Source: Calibration plot in Proietti et al. 2018[114](bleeding risk scores calibration between derivation cohorts and RE-LY cohort events rates). This probably relates to their total, mixed, cohort.

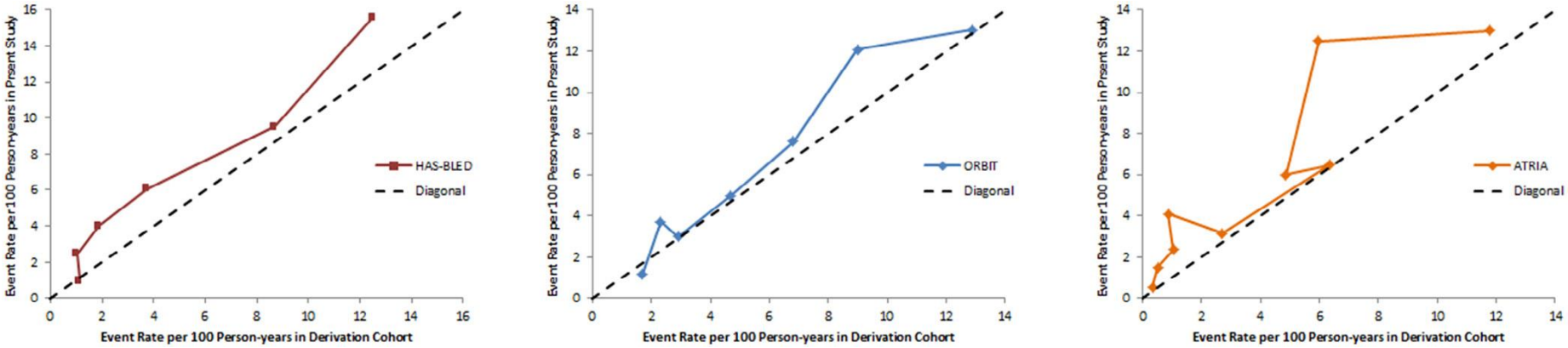
Figure 1: <Insert graphic title here>



Source: Calibration plot in O'Brien 2015[91]. This was a mixed cohort.



Source: Calibration plot in Lip, 2018[77]. This was based on an exclusively DOAC-using cohort.



Source: Calibration plot in Yao, 2017[158]. This was based on an exclusively DOAC-using cohort.

2.3.3 Net Reclassification improvementfor MAJOR BLEEDING

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 bleeding) 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 bleeding) 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.

Table 8: NRI for major bleeding – HAS-BLED versus other tools.

Prediction tool comparison	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	NRI(95% CI)	Quality
HAS-BLED v HEMORRHAGES	5	50,051	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	**Pooled: Random effects NRI: + 0.080(-0.030to +0.190); I^2= 69%**	VERY LOW
HAS-BLED v ATRIA	6	50,988	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	**Pooled: Random effects NRI: + 0.070(-0.020to +0.160); I^2= 52%**	VERY LOW
HAS-BLED v MBR	1	40450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.056 (0.043 to 0.068)	LOW
HAS-BLED v CHADS2	3	17529	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	**Pooled fixed effect NRI: +0.440(+0.250to +0.630); I^2=0%**	LOW
HAS-BLED v ORBIT	3	46284	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	**Pooled fixed effect NRI: +0.050(+0.040to +0.070); I^2=0%**	LOW

HAS-BLED v CHADSVASC	3	5518	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	**Pooled fixed effect NRI: +0.37 (+0.21 to +0.52); I^2=0%**	LOW
HAS-BLED v ABC	1	8705	Serious risk of bias[a]	Noserious inconsistency	No serious indirectness	Serious imprecision[c]	-0.138(-0.080to 0.228)	LOW
HAS-BLED v ABC CrC	1	1120	Serious risk of bias[a]	Noserious inconsistency	No serious indirectness	Serious imprecision[c]	+0.137(-0.010to 0.290)	LOW
HAS-BLED v GARFIELD	1	3550	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.042(-0.087 to 0.189)	VERY LOW
HAS-BLED v HAS-BLED with vWF	2	2155	Serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	**Pooled random effect NRI: -0.12 (-0.33 to +0.09); I^2=92%**	VERY LOW
HAS-BLED v HAS-BLED + VWF + NT-proBNP	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.201(-0.329 to -0.002)	MOD
HAS-BLED v HAS-BLED + VWF + NT-proBNP + IL-6	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.192(-0.325to -0.001)	MOD
HAS-BLED v HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.194(-0.337 to -0.003)	MOD
HAS-BLED v HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T + BTP	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.196(-0.327 to -0.005)	MOD
HAS-BLED v HAS-BLED + VWF + NT-proBNP + IL-6	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.203(-0.342 to -0.004)	MOD

+ Troponin T + BTP + soluble fibrin monomer complex								
HAS-BLED v Recalibrated HAS-BLED	1	Unknown	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.090(-0.123 to -0.0480)	LOW
HAS-BLED v modified HAS-BLED (including multiple biomarkers)	1	1361	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.062 (-0.020to 0.140)	LOW
HAS-BLED v modified HAS-BLED (including new renal dysfunction definition)	1	231	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.500(-0.820to -0.180)	LOW
HAS-BLED v GEN/HAS_BLES	1	652	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.044(0.010to 0.080)	MOD
HAS-BLED vs HAS-BLED with AS	1	2880	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.0481(p=0.034)	MOD

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) Inconsistency was serious if I2 was 50-74% and very serious if 75% of higher
c) Imprecision serious if the 95% CIs crossed zero.

Table 9: NRI for major bleeding – ATRIA versus other tools

Prediction tool comparison	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	NRI(95% CI)	Quality
ATRIA v CHADS2	2	16159	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	**MEDIAN: +0.43**	LOW
ATRIA v ORBIT	1	3551	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	+0.0355	LOW
ATRIA v CHADSVASC	2	42139	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	**MEDIAN:+0.32**	LOW
ATRIA v HEMORRHAGES	5	12664	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	**Pooled random effect NRI: +0.090(-0.080to +0.207); I2=83%**	VERY LOW
ATRIA v OBI	1	3063	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	+0.505	LOW
ATRIA v Kuijer	1	3063	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	+0.566	LOW
ATRIA v Kearon	1	3063	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	+0.277	LOW
ATRIA v Shireman	1	3063	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	+0.344	LOW

ATRIA v Riete	1	3063	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	+0.448	LOW
ATRIA v ATRIA with TTR<65%	3	4005	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	No serious imprecision	**Pooled random effect NRI: -0.230(-0.410to -0.040); I^2=64%**	VERY LOW
ATRIA v MBR	1	40450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	+0.007 (-0.014 to 0.027)	LOW
ATRIA vs ATRIA with AS	1	2880	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.0645(p=0.025)	MOD

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) Inconsistency was serious if I2 was 50-74% and very serious if 75% of higher
c) Imprecision serious if the 95% CIs crossed zero.

Table 10: NRI for major bleeding – HEMORRHAGES versus other tools

Prediction tool comparison	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	NRI(95% CI)	Quality
HEMORRHAGES v CHADS2	1	2600	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.540(0.220to 0.860)	LOW
HEMORRHAGES v CHADSVASC	1	2600	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.590(0.240to 0.940)	LOW

HEMORRHAGES v ORBIT	1	3551	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	-0.216	LOW
HEMORRHAGES v HEMORRHAGES with TTR<65%	2	1712	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	**MEDIAN: -0.161**	MOD
HEMORRHAGES v MBR	1	40450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.012 (-0.007 to 0.032)	VERY 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) Inconsistency was serious if I2 was 50-74% and very serious if 75% of higher
c) Imprecision serious if the 95% CIs crossed zero.

Table 11: NRI for major bleeding – ORBIT versus other tools

Prediction tool comparison	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	NRI(95% CI)	Quality
ORBIT v ORBIT with TTR<65%	3	4009	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	**Pooled random effect NRI: -0.21(-0.44 to 0.02); I^2=77%**	VERY LOW
ORBIT v CHADSVASC	1	39539	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	+0.010	LOW
ORBIT v MBR	1	40450	Very serious risk of	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.000 (-0.021 to 0.021)	VERY LOW

			bias[a]					
ORBIT vs ORBIT with AS	1	2880	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	-0.014(p=0.170)	VERY 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) Inconsistency was serious if I2 was 50-74% and very serious if 75% of higher
c) Imprecision serious if the 95% CIs crossed zero.

Table 12: NRI for major bleeding – CHADSVASC versus other tools

Prediction tool comparison	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	NRI(95% CI)	Quality
CHADSVASCv CHADS2	3	55698	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	**MEDIAN: +0.040**	VERY LOW
CHADSVASC v modified CHADSVASC (including multiple biomarkers)	1	1361	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.0026 (-0.020to 0.030)	VERY 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) Inconsistency was serious if I2 was 50-74% and very serious if 75% of higher
c) Imprecision serious if the 95% CIs crossed zero.

2.3.4 Discrimination for CLINICALLY RELEVANT BLEEDING

Table 13: Clinical evidence profile: accuracy of prediction of CRBin all risk tools featured in the studies (see table 3). Outcomes split across subgroups are only shown if sub-grouping was able to reduce I2 to <50% in all sub-groups.

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
HAS-BLED	8	18258	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**Pooled result: Random effect: 0.56(0.54-0.59). I^2=83%**	VERY LOW
HEMORRHAGES	3	4467	Very serious risk of bias[a]	Serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**Pooled effect: Random effects 0.56 (0.52-0.60); I2=64%**	VERY LOW
HEMORRHAGES subgrouped by OAC - VKA	2	3450	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	**Pooled effect: fixed effect 0.54(0.51-0.56); I2=0%**	LOW
HEMORRHAGES subgrouped by OAC – Mixed VKA/DOAC	1	1157	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.61(0.55-0.68)	LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
HEMORRHAGES subgrouped by antiplatelets - <33%	2	3450	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	No serious imprecision	**Pooled effect: fixed effects 0.54(0.51-0.56); I2=0%**	LOW
HEMORRHAGES subgrouped by antiplatelets - >33%	1	1157	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	No serious imprecision	0.61(0.55-0.68)	LOW
ATRIA	4	6760	Very serious risk of bias[a]	Serious risk of inconsistency[b]	No serious indirectness	Serious imprecision	**Pooled effect: Random Effects 0.52 (0.49-0.56); I^2=63%**	VERY LOW
ATRIA subgrouped by OAC - VKA	3	5743	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	Serious imprecision[c]	**Pooled effect: Fixed effects 0.51(0.49-0.53); I^2=0%**	VERY LOW
ATRIA subgrouped by OAC – Mixed	1	1017	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	No serious imprecision	0.61(0.54-0.67)	LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
VKA/DOACs								
ATRIA subgrouped by antiplatelets – <33%	3	5743	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	Serious imprecision[c]	**Pooled effect: Fixed effects 0.51(0.49-0.53); I^2=0%**	VERY LOW
ATRIA subgrouped by antiplatelets – >33%	1	1017	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	No serious imprecision	0.61(0.54-0.67)	LOW
ORBIT	3	5593	Very serious risk of bias[a]	Very serious risk of inconsistency[b]	No serious indirectness	No serious imprecision	**Pooled effect: Random Effects 0.57(0.52-0.61); I^2=73%**	VERY LOW
ORBIT subgrouped by antiplatelets - <33%	1	2293	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	Serious imprecision[c]	0.52(0.48-0.56)	VERY LOW
ORBIT subgrouped by antiplatelets - >33%	1	1017	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	No serious imprecision	0.61(0.54-0.68)	LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
ORBIT subgrouped by antiplatelets – not reported	1	2283	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	No serious imprecision	0.58(0.55-0.61)	LOW
CHADS2	1	2293	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	Serious imprecision[c]	0.51(0.47-0.55)	VERY LOW
CHADSVASC	1	2293	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	Serious imprecision[c]	0.53(0.49-0.57)	VERY LOW
GARFIELD	1	3550	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	No serious imprecision	0.57(0.55-0.58)	LOW
MBRFS	1	4576	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	No serious imprecision	0.53(0.52-0.54)	LOW
mOBRI	1	1017	Very	No	No serious	No serious	0.56(0.50-0.62)	LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
			serious risk of bias[a]	serious risk of incon-sistency	indirectness	imprecision		
CBRM /Shireman	1	1017	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.58(0.54-0.62)	LOW
Simplified HAS-BLED	1	1089	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.642(0.60-0.68)	LOW
HAS-BLED with point for sustained AF	1	1089	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.61(0.57-0.65)	LOW

GRADE was conducted with emphasis on C statistics as this was the primary measure discussed in decision making.
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 few 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 around two clinical thresholds: C statisticsof 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 14: Clinical evidence profile: sensitivity and specificityof prediction of clinically relevant bleedingin all risk tools featured in the studies (see table 3). 95% CIs are given for non-pooled results.

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
HAS-BLED at threshold ≥1	2	4566	**Median[d]: 0.913(0.880-0.940)**	**Median[d]: 0.171(0.160-0.190**	**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[b]	No serious indirectness	No serious imprecision	VERY LOW
HAS-BLED at threshold ≥2	2	4566	**Median[d]: 0.496(0.440-0.550)**	**Median[d]: 0.686(0.670-0.710)**	**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	No serious imprecision	VERY LOW
HAS-BLED	2	4566	**Median[d]: 0.110(0.080-0.150)**	**Median[d]: 0.950(0.940-0.960)**	**Sensitivity**				

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
at threshold ≥3					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	No serious imprecision	VERY LOW
ATRIA at threshold ≥1	1	2268	0.879(0.832-0.917)	0.113(0.099-0.128)	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	Serious imprecision[c]	VERY LOW
ATRIA at threshold ≥2	1	2268	0.411(0.349-0.475)	0.583(0.561-0.605)	**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
Hemmorhages at	1	2268	0.742(0.683-0.795)	0.353(0.332-0.374)	**Sensitivity**				
					Very	NA	No serious	No serious	LOW

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
threshold ≥1					serious risk of bias[a]		indirectness	imprecision	
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
Hemmorhages at threshold ≥2	1	2268	0.266(0.212-0.326)	0.779(0.770-0.788)	**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
ORBIT at threshold ≥1	1	2283	0.734(0.684-0.779)	0.388(0.367-0.411)	**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
ORBIT at threshold ≥2	1	2283	0.283(0.236-0.334	0.812(0.793-0.829)	**Sensitivity**				
					Very serious	NA	No serious indirectness	No serious imprecision	LOW

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					risk of bias[a]				
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
CHADS2 at threshold ≥1	1	2293	0.972(0.943-0.988)[3]	0.0230(0.170-0.305)[3]	**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
CHADS2 at threshold ≥2	1	2293	0.637(0.575-0.697)	0.385(0.364-0.406)	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	Serious imprecision [c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
CHADSVASC at threshold ≥2	1	2293	0.936(0.899-0.963)	0.079(0.069-0.093)	**Sensitivity**				
					Very serious risk of	NA	No serious indirectness	Serious imprecision [c]	VERY LOW

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					bias[a]				
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
CHADSVASC at threshold ≥3	1	2293	0.753(0.695-0.805)	0.292(0.273-0.313)	**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

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.

2.3.5 Calibration for CLINICALLY RELEVANT BLEEDING

Calibration was poorly reported in most papers, with all papers merely reporting the p value for Hosmer-Lemeshow statistics and proving a qualitative assessment of the relative calibration between tools. All studies simply reported a non-comparative 'adequate' calibration, usually based on a Hosmer-Lemeshow p value >0.05. 'Adequate' goodness of fit was thus described for ATRIA,[4, 14, 63]HAS-BLED,[4, 14, 63, 71]HEMORRHAGES[4, 14, 63]and ORBIT[14]. It was not possible, based on these data, to compare thelevels of calibration between these tools.

2.3.6 Net Reclassification improvement for CLINICALLY RELEVANT BLEEDING

Table 15: NRI for clinically relevant bleeding

Prediction tool comparison	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	NRI(95% CI)	Quality
HAS-BLED v HEMORRHAGES	2	3450	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	**Pooled: Random effects NRI: + 0.030(-0.130to +0.180); I^2= 89%**	VERY LOW
HAS-BLED v ATRIA	2	3450	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	**Pooled: Random effects NRI: + 0.040(-0.150to +0.220); I^2= 92%**	VERY LOW
ATRIA v HEMORRHAGES	2	3450	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	**Pooled: Random effects NRI: + 0.060(-0.060to +0.190); I2 = 81%**	VERY LOW
HAS-BLED v CHADS2	1	2293	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.130(0.050to 0.210)	LOW
HAS-BLED v GARFIELD	1	3550	Very serious risk of	No serious inconsistency	No serious indirectness	Serious imprecision[c]	-0.033(-0.129 to 0.094)	VERY LOW

			bias[a]					
HAS-BLED v CHADSVASC	1	2293	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.130(0.050to 0.210)	LOW
HAS-BLED v ORBIT	1	2283	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.156(0.043 to 0.27)	MOD
ATRIA v ATRIA +TTR	1	2293	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.260 (-0.480to -0.040)	LOW
ORBIT v ORBIT + TTR	1	2293	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.260 (-0.480to -0.040)	MOD

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) Inconsistency was serious if I2 was 50-74% and very serious if 75% of higher
c) Imprecision serious if the 95% CIs crossed zero.

2.3.7 Discrimination for INTRACRANIAL HEMORRHAGE

Table 16: Clinical evidence profile: accuracy of prediction of ICHin all risk tools featured in the studies (see table 3). Outcomes split across subgroups are only shown if sub-grouping was able to reduce I2 to <50% in all sub-groups.

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
HAS-BLED	7	110,194	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**Pooled effect: Random effects 0.56(0.53-0.60); I^2=83%**	VERY LOW
HAS-BLED subgrouped by antiplatelets - <33%	1	40,450	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.53(0.51-0.54)	LOW
HAS-BLED subgrouped by antiplatelets - >33%	3	18.113	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	**Pooled effect: Fixed effects 0.56(0.52-0.60); I2=0%**	LOW
HAS-BLED subgrouped by antiplatelets – not reported	3	51631	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	**Pooled effect: Fixed effects 0.59(0.58-0.61); I2=0%**	LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
HEMORRHAGES	5	107,162	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**Pooled effect: Random effects: 0.58(0.52-0.64); I2=93%**	VERY LOW
HEMORRHAGES subgrouped by antiplatelets – <33%	1	40,450	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.53(0.51-0.54)	LOW
HEMORRHAGES subgrouped by antiplatelets – >33%	3	18,113	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	**Pooled effect: Fixed effects 0.59(0.55-0.63); I2=0%**	LOW
HEMORRHAGES subgrouped by antiplatelets – not reported	1	48,599	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.62(0.60-0.64)	LOW
ATRIA	4	58,563	Very serious risk	Very serious	No serious indirectnes	No serious imprecision	**Pooled effect: Random effects 0.56(0.50-0.61); I2=75%**	VERY LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
			of bias[a]	risk of incon-sistency[b]	s			
ATRIA subgrouped for antiplatelets - <33%	1	40,450	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	Serious imprecision[c]	0.50(0.49-0.52)	VERY LOW
ATRIA subgrouped for antiplatelets - >33%	3	18.113	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	**Pooled effect: Fixed effects 0.58(0.54-0.63); I2=0%**	LOW
ORBIT	4	58,563	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	**Pooled effectRandom effects 0.58(0.50-0.67); I2=91%**	VERY LOW
ORBIT subgrouped for antiplatelets - <33%	1	40,450	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	serious imprecision[c]	0.50(0.48-0.51)	VERY LOW
ORBIT subgrouped for antiplatelets - >33%	3	18,113	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	**Pooled effect: Fixed effects 0.62(0.58-0.66); I2=0%**	LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median	Quality
ABCBleeding CrC	1	1120	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	Serious imprecision[c]	0.47(0.40-0.53)	VERY LOW
MBR	1	40450	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	No serious imprecision	0.52(0.50-0.53)	LOW

GRADE was conducted with emphasis on C statisticsas this was the primary measure discussed in decision making.
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 few 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 around 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 specificityof prediction of intracranial haemmorhagein all risk tools featured in the studies (see table 3). 95% CIs are given for non-pooled results.

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
HAS-BLEDat threshold ≥3	1		0.538(0.410-0.660)	0.572(0.540-0.600)	**Sensitivity**				
					Serious risk of bias[a]	NA	No serious indirectnes	Serious imprecision[c]	LOW
					Specificity				
					Serious risk of bias[a]	NA	No serious indirectnes	No serious imprecision	MOD
ABCCrC at threshold ≥2%	1		0.785(0.670-0.880)	0.186(0.160-0.210)	**Sensitivity**				
					Serious risk of bias[a]	NA	No serious indirectnes	No serious imprecision	MOD
					Specificity				
					Serious risk of bias[a]	NA	No serious indirectnes	No serious imprecision	MOD

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.

2.3.8 Calibration for INTRACRANIAL HEMORRHAGE

Proietti et al 2018[114]reported that the ORBIT score had best agreement between predicted and observed risks, that ATRIA had worst agreement and thatATRIA and HAS-BLED tended to overestimate the risk of bleeding. Meanwhile, HEMORRHAGES tended to underestimate bleeding risk. However it was unclear if this related specifically to intracranial bleeding.

2.3.9 Net Reclassification improvement for INTRACRANIAL HEMORRHAGE

Table 18: NRI for intracranial bleeding

Prediction tool comparison	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	NRI(95% CI)	Quality
HAS-BLED v HEMORRHAGES	1	40,450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.030(-0.001 to 0.060)	VERY LOW
HAS-BLED v ATRIA	1	40,450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.060(0.026 to 0.093)	LOW
HAS-BLED V ORBIT	1	40,450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.048(0.013 to 0.082)	LOW
HAS-BLED v MBR	1	40,450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.007(-0.018 to 0.033)	VERY LOW
HAS-BLED v ABCCrC	1	1120	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.139(-0.010to 0.290)	LOW
MBR v HEMORRHAGES	1	40,450	Very serious	No serious inconsistency	No serious indirectness	Serious imprecision[c]	-0.022(-0.062 to 0.017)	VERY

			risk of bias[a]					LOW
MBR v ATRIA	1	40,450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.052(-0.094 to -0.011)	LOW
MBR v ORBIT	1	40,450	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	-0.040(-0.083 to 0.002)	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) Inconsistency was serious if I2 was 50-74% and very serious if 75% of higher
c) Imprecision serious if the 95% CIs crossed zero.

2.4 Economic evidence

2.4.1 Included studies

No relevant health economic studies were identified.

2.4.2 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 D.

2.4.3 Unit costs

See 1.8.1.

2.5 The committee's discussion of the evidence

2.5.1 Interpreting the evidence

2.5.1.1 The outcomes that matter most

No clinical evidence was generated by thereviewon the effectiveness of risk stratification tool for predicting bleeding. The committee discussed the predictive accuracy evidence only, as this was felt to be sufficient to inform recommendations relevant to the most appropriate methods to predict bleeding in people with AF, without the need for any consensus recommendations or research recommendations pertaining to the effectivenessreview.

The committee agreed that the most critical predictive accuracy outcome measures for decision-making were calibration data. This was because the committee agreed that the best use of bleeding risk tools was as a means to guide a shared patient/clinician plan for alleviating reversible risk factors for bleeding; such a plan would require an accurate measure of absolute risk, the accuracy of which is best measured by calibration outcome data. Accurate binary decision-thresholds, such as those measured by discrimination outcome data (C statistics or sensitivity/specificity) were regarded as less critical, given that bleeding risk tools were not regarded as a decision aid for anticoagulant use (see second paragraph in section2.5.1.3). Net reclassification improvement (NRI) data, although also less critical than calibration data, was regarded as slightly more important than C statistics or sensitivity/specificity because of its propensity to sensitively differentiate the accuracy of different tools.

2.5.1.2 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.

2.5.1.3 Benefits and harms

The benefit of an accurate estimation of bleeding risk is that this may prompt appropriate and directed alleviation of any reversible causes of bleeding, as well as allowing appropriate levels of vigilance during anticoagulation. One possible disadvantage (harm) of using bleeding risk tools is underestimating bleeding risk, which may lead to insufficient attention to preventable risk factors and insufficient monitoring. Another potential harm is over-estimating bleeding risk, which can lead to unnecessary over-vigilance and possibly reluctance on the part of the patient (and maybe clinician) to commence anticoagulation. Thus using accurate bleeding risk prediction tools was seen by the committee as vital to maximise benefits and minimise harms.

The committee discussed the commonly observed clinical practice of using the bleeding risk score as a counterbalance to the stroke risk score, which tends to be done in order to facilitate binary decisions about initiating anticoagulation. The drawbacks of this were discussed. Comparisons of the actual bleeding and stroke risk tool scores were regarded by the committee as largely meaningless, given the varying significance of scores across different tools. In addition, comparison of absolute stroke and bleeding risks (derived from the scores) was also regarded as potentially misleading in the context of a decision to anti-coagulate, because bleeding risk includes the risk of bleeding events of lower severity than a stroke. Thus, for example, the committee noted that an equal absolute risk of stroke and bleeding would not necessarily represent equipoise, as the two competing events might not be of comparable severity. Any assessment of risk must also weigh up the probability of an

event occurring and consider the consequences of the event occurring. The committee reiterated the importance of using a bleeding risk tool to inform plans to reduce reversible causes of bleeding and to maintain appropriate levels of vigilanceduring anticoagulation, and that it should not be used as a threshold-based tool to determine if anticoagulation should take place.

The committee noted the importance of respecting any decision by an individual not to take anticoagulants. The committee were aware of the recommendations on tailoring healthcare services to the individual in the NICE guideline on patient experience of adult services (CG138).

Committee discussion focussed on tools where the weight of evidence was sufficient to warrant a recommendation. Therefore for tools that had been investigated in only one or two smaller studies, relatively little consideration was given to their possible useeven if predictive accuracy was encouraging. In addition, for those tools with larger amountsof evidence, the clearly less effective tools such as HEMORRHAGES(which had poorer calibration than ORBIT, HASBLED and ATRIA, as well as inferior discriminationand NRI)were given less consideration. Discussion focussed on three main tools: ORBIT, HAS-BLED and ATRIA, with the emphasis, as previously justified, on calibration data.

The calibration evidence suggested that ORBIT was better than HASBLED and ATRIA inaccurately predictingrisk of major bleeding. This was found in both mixed cohorts and DOAC-only cohorts. Importantly, ORBIT was better calibrated at all, and particularly higher, levels of risk. Given the relevance of calibration outcomes to the intended use of the tools - allowing an informed discussion about reversing modifiable risk factors and having an appropriate level of monitoring as a result of an accurate assessment of absolute risk - this finding was an important factor in the recommendation decision. Discrimination data were also discussed, and the committee agreed that the C statistics data supported the calibration data's indication that ORBIT was the most appropriate tool. Although the C-statisticsevidence suggested little to choose between HAS-BLED, ATRIA and ORBIT for people on VKAs, the C statisticsevidence suggested that ORBIT was the most accurate tool to use for patients on DOACs. The committee noted that around 90% of patients were currently on DOACS, and that this proportion would continue to increase with time. Hence this supported ORBIT beingregarded as the most appropriate bleeding risk tool for current and future patients.The sensitivity and specificity data at the established thresholds suggested that HAS-BLED and other tools might be more sensitive than ORBIT in predicting who will bleed whilst on anticoagulants, but this was counterbalanced bythe greater specificity of ORBIT. In contrast to the situation when predicting strokes, reduced sensitivity of bleeding risk prediction was not regarded as a serious problem because failure to detect high bleeding risk would not necessarily change decisions. This was because prediction of bleeding would not be used to withhold anticoagulants; instead, the risk prediction would be used as an objective aid to discussion with the patient about the need to modify bleeding risks and to be vigilant about possible bleeding. Meanwhile, the NRI evidence was fairly equivocal, suggesting similarities between ORBIT and HAS-BLED, and the committee felt that it did not negate the calibration evidence that ORBIT was the most appropriate tool.

There was some discussion about a two-tier recommendation – recommending ORBIT for people on DOACs and continuing with HAS-BLED for those patients restricted to VKAs (given that HAS-BLED appears to be as accurate, based on discrimination data, as ORBIT and ATRIA in VKA populations). This idea was rejected, partly because it was believed that the people who would currently be given VKAs would tend to be different from the VKA populations in the included studies. The VKA study populations tended to be fairly typical samples of people with NVAF, because VKAs were the principal anticoagulant therapy available at the time of these studies. In contrast, patients currently being given VKAs would tend to be atypical (for example, people with serious renal dysfunction). The committee therefore believed that the evidence suggesting HAS-BLED might be appropriate for people on VKAs was not relevant to current users of VKAs. In addition, ORBIT was superior when

measured by calibration outcomes in mixed cohorts. Given the greater relevance of calibration outcomes to the purported usage of bleeding risk tools, this strongly supported the decision to recommend ORBIT for all patients.

In addition to recommending ORBIT as a bleeding prediction tool, the committee also made recommendations on addressing the modifiable bleeding risk factors inherent in ORBIT, as well as the modifiable bleeding risk factors listed in the 2014 recommendations. Although the 2014 bleeding risk factors were related to the HAS-BLED, all were still thought to be relevant to a shared clinical decision on alleviating bleeding risk factors. Reversible causes of anaemia were listed as an additional modifiable risk factor as anaemia is a component of the ORBIT tool.

The committee were of the opinion that the decision to withhold anticoagulation because of concerns over bleeding risk meant depriving a patient of a treatment which, were it not for the bleeding risk, might have been of benefit in stroke prevention. As a number of factors contributing to bleeding risk are dynamic and also potentially correctable, the committee considered that the decision to withhold anticoagulation should not be made in perpetuity but should be subject to regular review and reconsideration as appropriate. They also thought it important that both the review and the outcome of the review should be documented.The committee expressed concern that anticoagulation was often erroneously not initiated due to a perceived high risk of falls, even though a very large number of falls (in excess of 300 per year) are known to be necessary to significantly increase the risk of bleeding. In addition, the committee noted that old age is often used as a reason to not anti-coagulate, even though age is already a factor in the bleeding risk tools used (and therefore would already be accounted for). Therefore the 2014 recommendation that anticoagulation should not be withheld because of the risk of falling was maintained, with an additional note that age should also not be a factor encouraging non-anticoagulation. The committee discussed referring to frailty in the recommendation but given it is so difficult to define they decided against this.

2.5.1.4 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 it was noted that ORBIT required knowledge of whether a patient had reduced haemoglobin or haematocrit. This was not part of the HAS-BLED score, the previously recommended bleeding risk tool, and so would be a change from current practice. The committee noted however that this should be available from patient history and so is unlikely to require additional NHS resource.

The committee also discussed the importance of using the most accurately calibratedbleeding tool as this would help to accurately identify individuals at higher risk of bleeding and therefore prompt the physicians to modify any bleeding risk factors and ensure adequate monitoring is provided. A more accurate tool, as demonstrated with the calibration data presented for ORBIT, would ensure the correct patients are being monitored and so NHS resources would be used more efficiently. That is only those who are truly at higher risk of bleeding are being monitored.

The committee agreed that there was sufficient clinical evidence of superiority for ORBIT to warrant an inevitablechange in practice.It involves measuring some parameters, such as haemoglobin and haematocrit, that are not included in the HAS-BLED tool used in current practice. However, the committee agreed that these factors would be measured routinely for people starting anticoagulation, regardless of the risk tool used, so extra resources are unlikely to be needed.

2.5.2 Other factors the committee took into account

Thecommitteenoted that people from black and ethnic minoritygroups do have a greater risk of stroke but the relationship with atrial fibrillation is unclear. For example, it is not clearif it

isthe presence of comorbidities or ethnic group, or an interaction beween these, that increases the risk of stroke. The committee also noted that a greater proportion of people from black and ethnic minority groups are undiagnosed compared to the general population. This is in part related to who is targeted for screening which is outside of the remit of this guideline.

The use of the ORBIT score is a change in practice, and may lead to some implementation hurdles. One potential problem is that ORBIT does not measure all of the modifiable risk factors previously included in HAS-BLED. At first sight this appears to imply additional testing is needed to ensure that all modifiable risk factors are measured. We would argue that whether ORBIT or HAS-BLED are used does not actually change the amount of modifiable risk factor investigations that need to be carried out by the investigating clinician. For example, full blood count, labile INR, blood pressure, liver function tests and renal function tests will need to be carried out in either case to evaluate whether current bleeding, increased blood pressure or treatable liver or renal disorders are present, each of which can be treated if needed to reduce bleeding risk. The only difference is that the results of labile INR, blood pressure, liver function tests and renal function tests will feed into informing the HAS-BLED score whereas haemoglobin and renal function results (GFR) will feed into the ORBIT score. This does not make ORBIT any more costly in terms of clinician time and resources, as other variables in ORBIT do not require invasive investigations. It could be argued that if the modifiable risk factors are not part of the tool then clinicians will not be prompted to discuss their modification. This is unlikely provided good practice is observed, as knowledge of the modifiable risk factors of bleeding is a basic clinical skill for any clinician dealing with AF patients, and such prompting should not be necessary. Another potential problem is that recommended bleeding risk evaluation for other conditions (such as venous thromboembolism) does not use ORBIT. This means that if ORBIT is used for AF, another tool (such as HAS-BLED) has to be used for other conditions. We would argue that if other tools need to be used for other conditions this does not constitute a major hurdle for clinicians, as the use of these tools is not difficult, and access to the online versions is straightforward. Nevertheless, to avoid clinician confusion with the unfamiliar tool, there will be a need for an initial transition period when new practices are being learned. This may require re-education in both primary and secondary care, which will have a resource impact, although this will be a time-limited impact, as each clinician will require limited training. Finally, unlike HAS-BLED, ORBIT is not embedded in the GP system. This will initially lead to the need to work outside this system, causing some practical difficulties. It is hoped, however, that ORBIT will eventually become embedded in the GP system. Again, this will have a resource impact, but given that centralised software changes are unlikely to be too difficult, the impact is not believed to be too large. Whilst implementation of ORBIT will provide some challenges, these should be overcome by the advantages of a tool that can provide a more accurate measure of bleeding risk.

References

1. Abumuaileq RRY, Abu-Assi E, Raposeiras-Roubin S, Lopez-Lopez A, Redondo-Dieguez A, Alvarez-Iglesias D et al. Comparative evaluation of HAS-BLED and ATRIA scores by investigating the full potential of their bleeding prediction schemes in non-valvular atrial fibrillation patients on vitamin-K antagonists. International Journal of Cardiology. 2014; 176(3):1259-1261

2. 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

3. Apostolakis S, Lane DA, Buller H, Lip GYH. Comparison of the CHADS2, CHA2DS2 - VASc and HAS-BLED scores for the prediction of clinically relevant bleeding in anticoagulated patients with atrial fibrillation: The AMADEUS trial. Thrombosis and Haemostasis. 2013; 110(5):1074-1079

4. Apostolakis S, Lane DA, Guo Y, Buller H, Lip GY. Performance of the HEMORR(2)HAGES, ATRIA, and HAS-BLED bleeding risk-prediction scores in patients with atrial fibrillation undergoing anticoagulation: the AMADEUS (evaluating the use of SR34006 compared to warfarin or acenocoumarol in patients with atrial fibrillation) study. Journal of the American College of Cardiology. 2012; 60(9):861-867

5. Asuncion Esteve-Pastor M, Miguel Rivera-Caravaca J, Roldan V, Vicente V, Valdes M, Marin F et al. Long-term bleeding risk prediction in 'real world' patients with atrial fibrillation: comparison of the HAS-BLED and ABC-Bleeding risk scores. Thrombosis and Haemostasis. 2017; 117(10):1848-1858

6. Atzema CL, Dorian P, Fang J, Tu JV, Lee DS, Chong AS et al. A clinical decision instrument to predict 30-day death and cardiovascular hospitalizations after an emergency department visit for atrial fibrillation: the atrial fibrillation in the emergency room, part 2 (AFTER2) study. American Heart Journal. 2018; 203:85-92

7. 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

8. Barnes GD, Gu X, Haymart B, Kline-Rogers E, Almany S, Kozlowski J et al. The predictive ability of the CHADS2 and CHA2DS2-VASc scores for bleeding risk in atrial fibrillation: the MAQI(2) experience. Thrombosis Research. 2014; 134(2):294-299

9. Benezet-Mazuecos J, Cinza S, Marin F, Ruiz Diaz MA, Chaves J, Fernandez De Cabo S. Validation of a screening tool (ICUSI questionnaire) for AF patients receiving vitaminin K antagonist therapy. Value in Health. 2017; 20(9):A625

10. Benito-Gonzalez T, Estevez-Loureiro R, de Prado AP, Minguito-Carazo C, Del Castillo Garcia S, Garrote-Coloma C et al. Incidence and prognostic implications of late bleeding events after percutaneous mitral valve repair. International Journal of Cardiology Heart and Vasculature. 2018; 21:16-21

11. 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

12. Bernaitis N, Ching CK, Badrick T, Anoopkumar-Dukie S. Identifying warfarin control with stroke and bleed risk scores. Heart, Lung and Circulation. 2018; 27(6):756-759

13. Bernaitis N, Ching CK, Chen L, Hon JS, Teo SC, Badrick T et al. A high HASBLED score identifies poor warfarin control in patients treated for non-valvular atrial fibrillation in Australia and Singapore. Basic and Clinical Pharmacology and Toxicology. 2017; 121(6):499-504

14. Beshir SA, Aziz Z, Yap LB, Chee KH, Lo YL. Evaluation of the predictive performance of bleeding risk scores in patients with non-valvular atrial fibrillation on oral anticoagulants. Journal of Clinical Pharmacy and Therapeutics. 2018; 43(2):209-219

15. Burgess S, Crown N, Louzada ML, Dresser G, Kim RB, Lazo-Langner A. Clinical performance of bleeding risk scores for predicting major and clinically relevant non-major bleeding events in patients receiving warfarin. Journal of Thrombosis and Haemostasis. 2013; 11(9):1647-1654

16. Caldeira D, Costa J, Fernandes RM, Pinto FJ, Ferreira JJ. Performance of the HAS-BLED high bleeding-risk category, compared to ATRIA and HEMORR2HAGES in patients with atrial fibrillation: a systematic review and meta-analysis. Journal of Interventional Cardiac Electrophysiology. 2014; 40(3):277-284

17. 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

18. Candeias Faria D, Ferreira J, Beringuilho M, Augusto J, Roque D, Santos M et al. Atrial fibrillation stroke and bleeding risk scores as predictors of mortality at 3 years follow-up. European Heart Journal. 2018; 39(Suppl 1):195

19. Chang YT, Hu YF, Liao JN, Chern CM, Lin YJ, Chang SL et al. The assessment of anticoagulant activity to predict bleeding outcome in atrial fibrillation patients receiving dabigatran etexilate. Blood Coagulation and Fibrinolysis. 2016; 27(4):389-395

20. Chao TF, Lip GYH, Lin YJ, Chang SL, Lo LW, Hu YF et al. Incident risk factors and major bleeding in patients with atrial fibrillation treated with oral anticoagulants: A comparison of baseline, follow-up and delta HAS-BLED scores with an approach focused on modifiable bleeding risk factors. Thrombosis and Haemostasis. 2018; 118(4):768-777

21. Chao TF, Lip GYH, Lin YJ, Chang SL, Lo LW, Hu YF et al. Major bleeding and intracranial hemorrhage risk prediction in patients with atrial fibrillation: attention to modifiable bleeding risk factors or use of a bleeding risk stratification score? A nationwide cohort study. International Journal of Cardiology. 2018; 254:157-161

22. Chia PL, Teoh X, Hua CM, Ching ME, Foo D. Anticoagulation use and predictors of stroke, bleeding and mortality in multi-ethnic Asian patients with atrial fibrillation: A single centre experience. Medical Journal of Malaysia. 2016; 71(5):256-258

23. Claxton JS, MacLehose RF, Lutsey PL, Norby FL, Chen LY, O'Neal WT et al. A new model to predict major bleeding in patients with atrial fibrillation using warfarin or direct oral anticoagulants. PloS One. 2018; 13(9):e0203599

24. Coleman CI, Vaitsiakhovich T, Nguyen E, Weeda ER, Sood NA, Bunz TJ et al. Agreement between coding schemas used to identify bleeding-related

hospitalizations in claims analyses of nonvalvular atrial fibrillation patients. Clinical Cardiology. 2018; 41(1):119-125

25. 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

26. Deitelzweig SB, Jing Y, Swindle JP, Makenbaeva D. Reviewing a clinical decision aid for the selection of anticoagulation treatment in patients with nonvalvular atrial fibrillation: applications in a US managed care health plan database. Clinical Therapeutics. 2014; 36(11):1566-1573.e1563

27. Diemberger I, Fantecchi E, Reggiani MLB, Martignani C, Angeletti A, Massaro G et al. Atrial fibrillation and prediction of mortality by conventional clinical score systems according to the setting of care. International Journal of Cardiology. 2018; 261:73-77

28. Donze J, Rodondi N, Waeber G, Monney P, Cornuz J, Aujesky D. Scores to predict major bleeding risk during oral anticoagulation therapy: a prospective validation study. American Journal of Medicine. 2012; 125(11):1095-1102

29. Dukanovic A, Staerk L, Fosbol EL, Gadsboll K, Gislason GH, Olesen JB. Predicted risk of stroke and bleeding and use of oral anticoagulants in atrial fibrillation: Danish nationwide temporal trends 2011-2016. Thrombosis Research. 2017; 160:19-26

30. Elvira-Ruiz G, Caro-Martinez C, Flores-Blanco PJ, Cerezo-Manchado JJ, Albendin-Iglesias H, Lova-Navarro A et al. Aortic valve stenosis provides complementary information to bleeding risk scores in non-valvular atrial fibrillation patients initiating anticoagulation. Journal of Geriatric Cardiology. 2020; 17(3):141-148

31. Esteve-Pastor MA, Garcia-Fernandez A, Macias M, Sogorb F, Valdes M, Roldan V et al. Is the ORBIT bleeding risk score superior to the HAS-BLED Score in anticoagulated atrial fibrillation patients? Circulation Journal. 2016; 80(10):2102-2108

32. Esteve-Pastor MA, Rivera-Caravaca JM, Shantsila A, Roldan V, Lip GYH, Marin F. Assessing bleeding risk in atrial fibrillation patients: comparing a bleeding risk score based only on modifiable bleeding risk factors against the has-bled score. The AMADEUS Trial. Thrombosis and Haemostasis. 2017; 117(12):2261-2266

33. Fang MC, Go AS, Chang Y, Borowsky LH, Pomernacki NK, Udaltsova N et al. A new risk scheme to predict warfarin-associated hemorrhage: The ATRIA (Anticoagulation and Risk Factors in Atrial Fibrillation) Study. Journal of the American College of Cardiology. 2011; 58(4):395-401

34. Fanola CL, Giugliano RP, Ruff CT, Trevisan M, Nordio F, Mercuri MF et al. A novel risk prediction score in atrial fibrillation for a net clinical outcome from the ENGAGE AF-TIMI 48 randomized clinical trial. European Heart Journal. 2017; 38(12):888-896

35. Fauchier L, Chaize G, Gaudin AF, Vainchtock A, Rushton-Smith SK, Cotte FE. Predictive ability of HAS-BLED, HEMORR2HAGES, and ATRIA bleeding risk scores in patients with atrial fibrillation. A French nationwide cross-sectional study. International Journal of Cardiology. 2016; 217:85-91

36. 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

37. 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

38. Gage BF, Yan Y, Milligan PE, Waterman AD, Culverhouse R, Rich MW et al. Clinical classification schemes for predicting hemorrhage: results from the National Registry of Atrial Fibrillation (NRAF). American Heart Journal. 2006; 151(3):713-719

39. Gallego P, Roldan V, Torregrosa JM, Galvez J, Valdes M, Vicente V et al. Relation of the HAS-BLED bleeding risk score to major bleeding, cardiovascular events, and mortality in anticoagulated patients with atrial fibrillation. Circulation: Arrhythmia and Electrophysiology. 2012; 5(2):312-318

40. Garcia-Fernandez A, Marin F, Roldan V, Galcera-Jornet E, Martinez-Martinez JG, Valdes M et al. The HAS-BLED score predicts long-term major bleeding and death in anticoagulated non-valvular atrial fibrillation patients undergoing electrical cardioversion. International Journal of Cardiology. 2016; 217:42-48

41. Garcia-Fernandez A, Roldan V, Rivera-Caravaca JM, Hernandez-Romero D, Valdes M, Vicente V et al. Does von Willebrand factor improve the predictive ability of current risk stratification scores in patients with atrial fibrillation? Scientific Reports. 2017; 7:41565

42. Geersing GJ. Balancing stroke and bleeding risk using CHA2DS2-VASc in treatment of primary care patients with Atrial Fibrillation. 2012. Available from: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-00985604/fullLast accessed: 21/01/2020.

43. Giustozzi MG, Vedovati MC, Verso M, Conti S, Verdecchia P, Bogliari G et al. Predictors of major bleeding in patients aged 90 years or over with atrial fibrillation on anticoagulant treatment. European Heart Journal. 2018; 39(Suppl 1):1309

44. Gorman EW, Perkel D, Dennis D, Yates J, Heidel RE, Wortham D. Validation of the HAS-BLED tool in atrial fibrillation patients receiving rivaroxaban. Journal of Atrial Fibrillation. 2016; 9(2):1461

45. 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

46. Guo Y, Lane DA, Chen Y, Lip GYH, m AFAIITi. Regular bleeding risk assessment associated with reduction in bleeding outcomes: the mAFA-II randomized trial. American Journal of Medicine. 2020; 133(10):1195-1202.e1192.

47. Guo Y, Zhu H, Chen Y, Lip GYH. Comparing bleeding risk assessment focused on modifiable risk factors only versus validated bleeding risk scores in atrial fibrillation. American Journal of Medicine. 2018; 131(2):185-192

48. Guo YT, Zhang Y, Shi XM, Shan ZL, Wang CJ, Wang YT et al. Assessing bleeding risk in 4824 Asian patients with atrial fibrillation: The Beijing PLA Hospital Atrial Fibrillation Project. Scientific Reports. 2016; 6:31755

49. 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

50. 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

51. Hijazi Z, Lindback J, Siegbahn A, Alexander JH, Held C, Hanna M et al. A new biomarker based risk score for predicting major bleeding in atrial fibrillation-the ABC (age, biomarkers, current disease) risk score. European Heart Journal. 2014; 1:357

52. Hijazi Z, Oldgren J, Andersson U, Connolly SJ, Eikelboom JW, Ezekowitz MD et al. Growth-differentiation factor 15 and risk of major bleeding in atrial fibrillation: insights from the Randomized Evaluation of Long-Term Anticoagulation Therapy (RE-LY) trial. American Heart Journal. 2017; 190:94-103

53. Hijazi Z, Oldgren J, Lindback J, Alexander J, Connolly S, Eikelboom J et al. External validation of the biomarker-based ABC-bleeding risk score for atrial fibrillation. Journal of the American College of Cardiology. 2016; 67(13 Suppl):893

54. 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

55. Hijazi Z, Oldgren J, Lindback J, Alexander JH, Connolly SJ, Eikelboom JW et al. A biomarker-based risk score to predict death in patients with atrial fibrillation: the ABC (age, biomarkers, clinical history) death risk score. European Heart Journal. 2018; 39(6):477-485

56. Hijazi Z, Siegbahn A, Andersson U, Granger CB, Alexander JH, Atar D et al. High-sensitivity troponin I for risk assessment in patients with atrial fibrillation: insights from the Apixaban for Reduction in Stroke and other Thromboembolic Events in Atrial Fibrillation (ARISTOTLE) trial. Circulation. 2014; 129(6):625-634

57. Hijazi Z, Wallentin L, Siegbahn A, Andersson U, Alexander JH, Atar D et al. High-sensitivity troponin T and risk stratification in patients with atrial fibrillation during treatment with apixaban or warfarin. Journal of the American College of Cardiology. 2014; 63(1):52-61

58. Hilkens NA, Algra A, Greving JP. Predicting major bleeding in ischemic stroke patients with atrial fibrillation. Stroke. 2017; 48(11):3142-3144

59. Hippisley-Cox J, Coupland C. Predicting risk of upper gastrointestinal bleed and intracranial bleed with anticoagulants: cohort study to derive and validate the QBleed scores. BMJ. 2014; 349:g4606

60. 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

61. Iwasaki Y. Clinical usefulness of ATRIA score to predict cardiovascular outcomes in patients with atrial fibrillation undergoing percutaneous coronary intervention with stenting. European Heart Journal. 2018; 39(Suppl 1):1296

62. 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

63. Jaspers Focks J, van Vugt SPG, Albers-Akkers MTH, Lamfers EJP, Bloem-de Vries LM, Verheugt FWA et al. Low performance of bleeding risk models in the very elderly with atrial fibrillation using vitamin K antagonists. Journal of Thrombosis and Haemostasis. 2016; 14(9):1715-1724

64. Jensen M, Skjoeth F, Nielsen PB, Larsen TB, Melgaard L, Lip GYH. Stroke and bleeding risk scores in patients with atrial fibrillation and valvular heart disease: Prospective validation of the EHRA classification in a nationwide cohort study. European Heart Journal. 2018; 39(Suppl 1):25

65. Jover E, Roldan V, Gallego P, Hernandez-Romero D, Valdes M, Vicente V et al. Predictive value of the CHA2DS2-VASc score in atrial fibrillation patients at high risk for stroke despite oral anticoagulation. Revista Española de Cardiología. 2012; 65(7):627-633

66. 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

67. Lamberts M, Staerk L, Olesen JB, Fosbol EL, Hansen ML, Harboe L et al. Major bleeding complications and persistence with oral anticoagulation in non-valvular atrial fibrillation: Contemporary findings in real-life Danish patients. Journal of the American Heart Association. 2017; 6(2):e004517

68. Lee KT, Chang SH, Yeh YH, Tu HT, Chan YH, Kuo CT et al. The CHA2DS2-VASc Score predicts major bleeding in non-valvular atrial fibrillation patients who take oral anticoagulants. Journal of Clinical Medicine. 2018; 7(10):1-9

69. Li Kam Wa ME, Khouri A, Whitfield M, Amin R, Mohamed MO, McWilliams N et al. The ARDVAARC study: real-world outcomes and the utility of bleeding risk scores in patients who require anticoagulation following percutaneous coronary intervention (PCI). European Heart Journal. 2018; 39 (Suppl 1):1333

70. Lip GY, Banerjee A, Lagrenade I, Lane DA, Taillandier S, Fauchier L. Assessing the risk of bleeding in patients with atrial fibrillation: the Loire Valley Atrial Fibrillation project. Circulation: Arrhythmia and Electrophysiology. 2012; 5(5):941-948

71. Lip GY, Frison L, Halperin JL, Lane DA. Comparative validation of a novel risk score for predicting bleeding risk in anticoagulated patients with atrial fibrillation: the HAS-BLED (Hypertension, Abnormal Renal/Liver Function, Stroke, Bleeding History or Predisposition, Labile INR, Elderly, Drugs/Alcohol Concomitantly) score. Journal of the American College of Cardiology. 2011; 57(2):173-180

72. Lip GY, Lin HJ, Hsu HC, Su TC, Chen MF, Lee YT et al. Comparative assessment of the HAS-BLED score with other published bleeding risk scoring schemes, for intracranial haemorrhage risk in a non-atrial fibrillation population: the Chin-Shan Community Cohort Study. International Journal of Cardiology. 2013; 168(3):1832-1836

73. Lip GYH. Can we predict stroke in atrial fibrillation? Clinical Cardiology. 2012; 35(SUPPL. 1):21-27

74. 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

75. 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. 2018; 06:1-8

76. 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

77. Lip GYH, Skjoth F, Nielsen PB, Kjaeldgaard JN, Larsen TB. The HAS-BLED, ATRIA, and ORBIT bleeding scores in atrial fibrillation patients using non-vitamin K antagonist oral anticoagulants. American Journal of Medicine. 2018; 131(5):574.e513-574.e527

78. Lobos-Bejarano JM, Barrios V, Polo-Garcia J, Escobar C, Vargas-Ortega D, Marin-Montanes N et al. Evaluation of SAMe-TT2R2 score and other clinical factors influencing the quality of anticoagulation therapy in non-valvular atrial fibrillation: a nationwide study in Spain. Current Medical Research and Opinion. 2016; 32(7):1201-1207

79. Loewen P, Dahri K. Risk of bleeding with oral anticoagulants: an updated systematic review and performance analysis of clinical prediction rules. Annals of Hematology. 2011; 90(10):1191-1200

80. Lv MN, Zheng XC, Zhang HQ, Xu FD, Wu TT, Chen WJ et al. Comparison of clinical performance of four gastrointestinal bleeding risk scores in Chinese patients with atrial fibrillation receiving oral anticoagulants. Journal of Thrombosis and Thrombolysis. 2020; https://doi.org/10.1007/s11239-020-02152-1

81. 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

82. Marcucci M, Lip GY, Nieuwlaat R, Pisters R, Crijns HJ, Iorio A. Stroke and bleeding risk co-distribution in real-world patients with atrial fibrillation: the Euro Heart Survey. American Journal of Medicine. 2014; 127(10):979-986.e972

83. Marcucci M, Nobili A, Tettamanti M, Iorio A, Pasina L, Djade CD et al. Joint use of cardio-embolic and bleeding risk scores in elderly patients with atrial fibrillation. European Journal of Internal Medicine. 2013; 24(8):800-806

84. 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

85. 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

86. Methavigul K. Use of SAMe-TT2R2 score to predict the quality of anticoagulation control in patients with atrial fibrillation receiving warfarin in Thailand. Journal of the Medical Association of Thailand. 2020; 103(6):548-552

87. Molnar AO, Sood MM. Predicting in a predicament: stroke and hemorrhage risk prediction in dialysis patients with atrial fibrillation. Seminars in Dialysis. 2018; 31(1):37-47

88. Mori N, Sotomi Y, Hirata A, Hirayama A, Sakata Y, Higuchi Y. External validation of the ORBIT bleeding score and the HAS-BLED score in nonvalvular atrial fibrillation patients using direct oral anticoagulants (Asian data from the DIRECT registry). American Journal of Cardiology. 2019; 124(7):1044-1048

89. 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

90. Nielsen PB, Larsen TB, Lip GYH. Recalibration of the HAS-BLED score: should hemorrhagic stroke account for one or two points? Chest. 2016; 149(2):311-314

91. O'Brien EC, Simon DN, Thomas LE, Hylek EM, Gersh BJ, Ansell JE et al. The ORBIT bleeding score: a simple bedside score to assess bleeding risk in atrial fibrillation. European Heart Journal. 2015; 36(46):3258-3264

92. O'Caoimh R, Igras E, Ramesh A, Power B, O'Connor K, Liston R. Assessing the appropriateness of oral anticoagulation for atrial fibrillation in advanced frailty: use of stroke and bleeding risk-prediction models. The Journal of Frailty and Aging. 2017; 6(1):46-52

93. Okumura K, Inoue H, Atarashi H, Yamashita T, Tomita H, Origasa H et al. Validation of CHA2DS2-VASc and HAS-BLED scores in Japanese patients with nonvalvular atrial fibrillation: an analysis of the J-RHYTHM Registry. Circulation Journal. 2014; 78(7):1593-1599

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, Lip GY, Hansen PR, Lindhardsen J, Ahlehoff O, Andersson C et al. Bleeding risk in 'real world' patients with atrial fibrillation: comparison of two established bleeding prediction schemes in a nationwide cohort. Journal of Thrombosis and Haemostasis. 2011; 9(8):1460-1467

96. Olesen JB, Lip GY, Lindhardsen J, Lane DA, Ahlehoff O, Hansen ML et al. Risks of thromboembolism and bleeding with thromboprophylaxis in patients with atrial fibrillation: a net clinical benefit analysis using a 'real world' nationwide cohort study. Thrombosis and Haemostasis. 2011; 106(4):739-749

97. Olesen JB, Lip GYH, Hansen PR, Lindhardsen J, Ahlehoff O, Andersson C et al. Predicting bleeding risk in "real world" patients with atrial fibrillation with or without anticoagulation: A comparison of two established bleeding prediction schemes in a nationwide cohort study. European Heart Journal. 2011; 1:671

98. Omran H, Bauersachs R, Rubenacker S, Goss F, Hammerstingl C. The HAS-BLED score predicts bleedings during bridging of chronic oral anticoagulation. Results from the national multicentre BNK Online bRiDging REgistRy (BORDER). Thrombosis and Haemostasis. 2012; 108(1):65-73

99. Pardo Sanz A, Rincon LM, Tamayo A, De Lara G, Contreras H, Rueda A et al. Performance of atrial fibrillation ischemic and bleeding risk scores in patients with cancer. European Heart Journal. 2018; 39(Suppl 1):305

100. Parks AL, Fang MC. Scoring systems for estimating the risk of anticoagulant-associated bleeding. Seminars in Thrombosis and Hemostasis. 2017; 43(5):514-524

101. Peacock WF, Tamayo S, Patel M, Sicignano N, Hopf KP, Yuan Z. CHA2DS2-VASc scores and major bleeding in patients with nonvalvular atrial fibrillation who are receiving rivaroxabans. Annals of Emergency Medicine. 2017; 69(5):541-550.e541

102. Perez-Copete J, Esteve-Pastor MA, Roldan V, Valdes M, Marin F. Thromboembolic and bleeding risk scores in atrial fibrillation. Revista Espanola de Cardiologia Suplementos. 2016; 16(Suppl 1):25-32

103. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. Chest. 2010; 138(5):1093-1100

104. Poli D, Antonucci E, Grifoni E, Carini U, Ciampa A, Da Col P et al. Low bleeding risk of very old atrial fibrillation women on VKA treatment: Results from a prospective

collaborative study. on behalf of the ad hoc study group of FCSA. Journal of Thrombosis and Haemostasis. 2011; 2:886

105. Poli D, Antonucci E, Grifoni E, Ciuti G, Marcucci R, Mannini L et al. Stroke risk in atrial fibrillation patients on warfarin: predictive ability of risk stratification schemes for primary and secondary prevention. Journal of Thrombosis and Haemostasis. 2009; 7(Suppl 2):433

106. Poli D, Antonucci E, Grifoni E, Di Gennaro L, Falanga A, Falco P et al. Bleeding risk in very old patients on VKA treatment: results of a prospective collaborative study. Journal of Thrombosis and Haemostasis. 2011; 2:747

107. Poli D, Antonucci E, Grifoni E, Marcucci R, Mannini L, Abbate R et al. Bleeding risk during oral anticoagulation in atrial fibrillation patients older than 80 years. Journal of Thrombosis and Haemostasis. 2009; 7(Suppl 2):433

108. Poli D, Antonucci E, Marcucci R, Fatini C, Alterini B, Mannini L et al. Risk of bleeding in very old atrial fibrillation patients on warfarin: relationship with ageing and CHADS2 score. Thrombosis Research. 2007; 121(3):347-352

109. Poli D, Antonucci E, Marongiu F, Pengo V, Testa S, Tripodi A et al. Similar performance of HAS-BLED, CHADS2 and CHA2DS2VASC scores in bleeding risk prediction of atrial fibrillation patients: the refined HAS-BED score results from the start register. Blood Transfusion. 2016; 14(Suppl 5):S777

110. 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

111. Poli D, Antonucci E, Testa S, Cosmi B, Palareti G, Ageno W et al. The predictive ability of bleeding risk stratification models in very old patients on vitamin K antagonist treatment for venous thromboembolism: results of the prospective collaborative EPICA study. Journal of Thrombosis and Haemostasis. 2013; 11(6):1053-1058

112. Poli D, Antonucci E, Testa S, Tosetto A, Ageno W, Palareti G et al. Bleeding risk in very old patients on vitamin K antagonist treatment: results of a prospective collaborative study on elderly patients followed by Italian Centres for Anticoagulation. Circulation. 2011; 124(7):824-829

113. Prochaska JH, Gobel S, Nagler M, Knopfler T, Eggebrecht L, Lamparter H et al. Sustained atrial fibrillation increases the risk of anticoagulation-related bleeding in heart failure. Clinical Research in Cardiology. 2018; 107(12):1170-1179

114. Proietti M, Hijazi Z, Andersson U, Connolly SJ, Eikelboom JW, Ezekowitz MD et al. Comparison of bleeding risk scores in patients with atrial fibrillation: insights from the RE-LY trial. Journal of Internal Medicine. 2018; 283(3):282-292

115. Proietti M, Rivera-Caravaca JM, Esteve-Pastor MA, Romiti GF, Marin F, Lip GYH. Predicting bleeding events in anticoagulated patients with atrial fibrillation: A comparison between the HAS-BLED and GARFIELD-AF bleeding scores. Journal of the American Heart Association. 2018; 7(18):e009766

116. Proietti M, Senoo K, Lane DA, Lip GY. Major bleeding in patients with non-valvular atrial fibrillation: impact of time in therapeutic range on contemporary bleeding risk scores. Scientific Reports. 2016; 6:24376

117. Quinn GR, Singer DE, Chang Y, Go AS, Borowsky LH, Fang MC. How well do stroke risk scores predict hemorrhage in patients with atrial fibrillation? American Journal of Cardiology. 2016; 118(5):697-699

118. 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

119. Rivera-Caravaca JM, Marin F, Vilchez JA, Galvez J, Esteve-Pastor MA, Vicente V et al. Refining stroke and bleeding prediction in atrial fibrillation by adding consecutive biomarkers to clinical risk scores. Stroke. 2019; 50(6):1372-1379

120. 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

121. 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

122. 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

123. Rivera Caravaca JM, Esteve-Pastor MA, Vilchez JA, Galvez J, Vicente V, Marin F et al. Refining stroke and bleeding risk prediction by adding consecutive biomarkers to CHA2DS2-VASc and HAS-BLED scores. European Heart Journal. 2018; 39(Suppl 1):821-822

124. Roldan V, Marin F, Diaz J, Gallego P, Jover E, Romera M et al. High sensitivity cardiac troponin T and interleukin-6 predict adverse cardiovascular events and mortality in anticoagulated patients with atrial fibrillation. Journal of Thrombosis and Haemostasis. 2012; 10(8):1500-1507

125. Roldan V, Marin F, Fernandez H, Manzano-Fernandez S, Gallego P, Valdes M et al. Predictive value of the HAS-BLED and ATRIA bleeding scores for the risk of serious bleeding in a "real-world" population with atrial fibrillation receiving anticoagulant therapy. Chest. 2013; 143(1):179-184

126. Roldan V, Marin F, Manzano-Fernandez S, Gallego P, Vilchez JA, Valdes M et al. The HAS-BLED score has better prediction accuracy for major bleeding than CHADS2 or CHA2DS2-VASc scores in anticoagulated patients with atrial fibrillation. Journal of the American College of Cardiology. 2013; 62(23):2199-2204

127. Roldan V, Marin F, Muina B, Torregrosa JM, Hernandez-Romero D, Valdes M et al. Plasma von Willebrand factor levels are an independent risk factor for adverse events including mortality and major bleeding in anticoagulated atrial fibrillation patients. Journal of the American College of Cardiology. 2011; 57(25):2496-2504

128. 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

129. Rutherford OCW, Jonasson C, Ghanima W, Halvorsen S. A new score for assessing bleeding risk in patients with atrial fibrillation treated with NOACs. European Heart Journal. 2018; 39(Suppl 1):1009-1010

130. Sadeghi R, Mahjoob MP, Asadollahi M, Abbasi Z. Prevalence, main determinants, and early outcome of patients with atrial fibrillation hospitalized with ischemic stroke: evaluation of the value of risk assessment scores for predicting risk of stroke or major bleeding following anticoagulation therapy. Acta Bio-Medica de l Ateneo Parmense. 2015; 86(2):162-169

131. 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

132. Salpagarova ZK, Chashkina M, Andreev DA, Bykova AA, Sychev DA, Kozlovskaya NL et al. The new warfarin dosing algorithm for patients with atrial fibrillation and severe chronic kidney disease. European Heart Journal. 2018; 39(Suppl 1):614

133. Sanders GD, Lowenstern A, Borre E, Chatterjee R, Goode A, Sharan L et al. Stroke prevention in patients with atrial fibrillation: a systematic review update. Rockville (MD). Agency for Healthcare Research and Quality, 2018. Available from: https://effectivehealthcare.ahrq.gov/products/stroke-afib-update/research-2018

134. Sani M, Ayubi E, Mansori K, Khazaei S. Predictive ability of HAS-BLED, HEMORR2HAGES, and ATRIA bleeding risk scores in patients with atrial fibrillation: methodological issues of prediction models. International Journal of Cardiology. 2016; 222:949

135. 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

136. Senoo K, Lip GY. Predictive abilities of the HAS-BLED and ORBIT bleeding risk scores in non-warfarin anticoagulated atrial fibrillation patients: an ancillary analysis from the AMADEUS trial. International Journal of Cardiology. 2016; 221:379-382

137. Senoo K, Proietti M, Lane DA, Lip GYH. Evaluation of the HAS-BLED, ATRIA, and ORBIT bleeding risk scores in patients with atrial fibrillation taking warfarin. American Journal of Medicine. 2016; 129(6):600-607

138. Serna MJ, Rivera-Caravaca JM, Gonzalez-Conejero R, Esteve-Pastor MA, Valdes M, Vicente V et al. Pharmacogenetics of vitamin K antagonists and bleeding risk prediction in atrial fibrillation. European Journal of Clinical Investigation. 2018; 48(6):e12929

139. Shah RR, Pillai A, Omar A, Zhao J, Arora V, Kapoor D et al. Utility of the HAS-BLED score in risk stratifying patients on dual antiplatelet therapy post 12 months after drug-eluting stent placement. Catheterization and Cardiovascular Interventions. 2017; 89(4):E99-E103

140. Shahid F, Lip GYH. Risk stratification models in atrial fibrillation. Seminars in Thrombosis and Hemostasis. 2017; 43(5):505-513

141. Silva R, Silva PAE, Lima MC, Sant'Anna LT, Silva TC, Moreira PHV et al. Thromboembolism and bleeding risk scores and predictors of cardiac death in a population with atrial fibrillation. Arquivos Brasileiros de Cardiologia. 2017; 109(1):5-13

142. 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

143. Sogaard M, Skjoth F, Kjaeldgaard JN, Lip GYH, Larsen TB. Bleeding complications in anticoagulated patients with atrial fibrillation and sepsis: a propensity-weighted cohort study. Journal of the American Heart Association. 2017; 6(11):09

144. 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

145. Sood MM, Larkina M, Thumma JR, Tentori F, Gillespie BW, Fukuhara S et al. Major bleeding events and risk stratification of antithrombotic agents in hemodialysis: Results from the DOPPS. Kidney International. 2013; 84(3):600-608

146. Steinberg BA, Shrader P, Kim S, Thomas L, Fonarow GC, Ansell J et al. How well does physician risk assessment predict stroke and bleeding in atrial fibrillation? Results from the Outcomes Registry for Better Informed Treatment of Atrial Fibrillation (ORBIT-AF). American Heart Journal. 2016; 181:145-152

147. Suzuki M, Matsue Y, Nakamura R, Matsumura A, Hashimoto Y. Improvement of HAS-BLED bleeding score predictive capability by changing the definition of renal dysfunction in Japanese atrial fibrillation patients on anticoagulation therapy. Journal of Cardiology. 2014; 64(6):482-487

148. Tchen S, Ryba N, Patel V, Cavanaugh J, Sullivan JB. Validation of bleeding risk prediction scores for patients with major bleeding on direct oral anticoagulants. Annals of Pharmacotherapy. 2020; 54(12):1175-1184

149. Thomas IC, Sorrentino MJ. Bleeding risk prediction models in atrial fibrillation. Current Cardiology Reports. 2014; 16(1):432

150. Toyoda K, Yasaka M, Uchiyama S, Iwade K, Koretsune Y, Nagata K et al. CHADS2 and CHA2DS2-VASc scores as bleeding risk indices for patients with atrial fibrillation: the Bleeding with Antithrombotic Therapy Study. Hypertension Research - Clinical and Experimental. 2014; 37(5):463-466

151. van Doorn S, Rutten FH, O'Flynn CM, Oudega R, Hoes AW, Moons KGM et al. Effectiveness of CHA2DS2-VASc based decision support on stroke prevention in atrial fibrillation: a cluster randomised trial in general practice. International Journal of Cardiology. 2018; 273:123-129

152. 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

153. Wang C, Yu Y, Zhu W, Yu J, Lip GYH, Hong K. Comparing the ORBIT and HAS-BLED bleeding risk scores in anticoagulated atrial fibrillation patients: a systematic review and meta-analysis. Oncotarget. 2017; 8(65):109703-109711

154. Wang SV, Franklin JM, Glynn RJ, Schneeweiss S, Eddings W, Gagne JJ. Prediction of rates of thromboembolic and major bleeding outcomes with dabigatran or warfarin among patients with atrial fibrillation: new initiator cohort study. BMJ. 2016; 353:i2607

155. Wang SV, Rogers JR, Jin Y, Bates DW, Fischer MA. Use of electronic healthcare records to identify complex patients with atrial fibrillation for targeted intervention. Journal of the American Medical Informatics Association. 2017; 24(2):339-344

156. Wang TK, Sathananthan J, Marshall M, Kerr A, Hood C. Relationships between anticoagulation, risk scores and adverse outcomes in dialysis patients with atrial fibrillation. Heart, Lung and Circulation. 2016; 25(3):243-249

157. Wang Y, Bajorek B. Pilot of a Computerised Antithrombotic Risk Assessment Tool Version 2 (CARATV2.0) for stroke prevention in atrial fibrillation. Cardiology Journal. 2017; 24(2):176-187

158. Yao X, Gersh BJ, Sangaralingham LR, Kent DM, Shah ND, Abraham NS et al. Comparison of the CHA2DS2-VASc, CHADS2, HAS-BLED, ORBIT, and ATRIA risk scores in predicting non-vitamin K antagonist oral anticoagulants-associated bleeding in patients with atrial fibrillation. American Journal of Cardiology. 2017; 120(9):1549-1556

159. Zeng J, Yu P, Cui W, Wang X, Ma J, Zeng C. Comparison of HAS-BLED with other risk models for predicting the bleeding risk in anticoagulated patients with atrial fibrillation: a PRISMA-compliant article. Medicine. 2020; 99(25):e20782

160. Zhu W, He W, Guo L, Wang X, Hong K. The HAS-BLED score for predicting major bleeding risk in anticoagulated patients with atrial fibrillation: a systematic review and meta-analysis. Clinical Cardiology. 2015; 38(9):555-561

161. Ziviello F, Pilgrim T, Kroon H, Ooms JF, van Wiechen MP, Daemen J et al. Has-Bled score and actual bleeding in elderly patients undergoing transcatheter aortic valve implantation. JACC: Cardiovascular Interventions. 2019; 12(Suppl 4):S49

162. Zulkifly H, Lip GYH, Lane DA. Bleeding risk scores in atrial fibrillation and venous thromboembolism. American Journal of Cardiology. 2017; 120(7):1139-1145

Appendices

Appendix A: Review protocols

Table 19: Review question: What is the most clinically and cost-effective risk stratification tool for predicting bleeding 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 bleeding risk in people with atrial fibrillation
2.	Review question	What is the most clinically and cost-effective tool for assessing bleeding risk in people with atrial fibrillation?
3.	Objective	To identify the most clinically and cost effective tool to measure the risk of bleeding 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 bleeding risk score (such as ABC bleeding score, Orbit bleeding score, ATRIA, HEMORR2HAGES or any version of HAS-BLED with modifications [treat each test using a different threshold as a separate intervention; for example, ABC bleeding score using the threshold of X for 'need to consider high bleeding risk' is treated as a separate intervention to ABC bleeding score using the threshold of Y for 'need to consider high bleeding risk'].

ID	Field	Content
8.	Comparator/Reference standard/Confounding factors	HAS-BLED (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 major bleeding stroke or thromboembolic complications 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 manualsection 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).
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)

ID	Field	Content	
		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: Type of anticoagulant (Vit K antagonist vs R v E v A v D). Concomitant anti-platelet agents/NSAIDs vs none	
18.	Type and method of review	☒	Intervention
		☐	Diagnostic
		☐	Prognostic
		☐	Qualitative
		☐	Epidemiologic
		☐	Service Delivery
		☒	Other (please specify): RCT of prediction tools
19.	Language	English	
20.	Country	England	
21.	Anticipated or		

ID	Field	Content		
	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.		
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		

ID	Field	Content	
		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, bleeding 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 20: 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
1.	Review title	Accuracy of risk stratification tools for predicting bleeding events in people with atrial fibrillation.
2.	Review question	What is the most accurate risk stratification tool for predicting bleeding events in people with atrial fibrillation?
3.	Objective	To identify the most accurate tool to measure the risk of bleeding 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

ID	Field	Content
		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 on oral anticoagulants.
7.	Index Test	Any risk tool designed to predict risk of bleeding (such as, ABC bleeding score, Orbit bleeding score, ATRIA, HEMORR2HAGES, HAS-BLED, and any version of HAS-BLED with modifications
8.	Comparator/Reference standard/Confounding factors	Later major bleeding Later bleeding, not specified as major These will be dealt with separately
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-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.

ID	Field	Content			
		A standardised form will be used to extract data from the included studies (see Developing NICE guidelines: the manualsection 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: Type of anticoagulant (Vit K antagonist vs R v E v A v D). Concomitant anti-platelet agents/NSAIDs vs not			
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.	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	☐	☑	

ID	Field	Content		
		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.		
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.]		

ID	Field	Content	
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 21: 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.[89] **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 tool for assessing bleeding risk in people with atrial fibrillation?**
- **What is the most accurate risk stratification tool for predicting bleeding 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.[89]

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. Search filters were applied to the search where appropriate.

Searches were constructed using the following approaches:

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

Table 22: 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 Prognostic/risk factor studies
Embase (OVID)	1974 – 10 September 2020	Exclusions Randomised controlled trials Systematic review studies Observational studies Prognostic/risk factor 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.	proportional hazards models/ or logistic models/ or risk assessment/ or risk factors/ or decision support systems, clinical/ or decision support techniques/
26.	(risk* tool* or stratification or rating scale* or scor* system* or scor* schem* or risk* schem* or risk* stratif* or risk* classif* or risk* assess*).ti,ab.
27.	Hemorrhage/
28.	25 and 26 and 27
29.	ATRIA.ti,ab.
30.	((ABC or Orbit) adj2 (bleed* or scor*)).ti,ab.
31.	HEMORR2HAGES.ti,ab.
32.	"HEMORR(2)HAGES".ti,ab.
33.	(hasbled or has-bled).ti,ab.
34.	((bleed* or hemorrhag* or haemorrhag*) adj3 scor*).ti,ab.
35.	((bleed* or hemorrhag* or haemorrhag*) adj3 (risk* tool* or stratification or rating scale* or scor* system* or scor* schem* or risk* schem* or risk* stratif* or risk* classif* or risk* assess*)).ti,ab.
36.	or/28-35
37.	24 and 36
38.	randomized controlled trial.pt.
39.	controlled clinical trial.pt.
40.	randomi#ed.ab.
41.	placebo.ab.
42.	randomly.ab.
43.	clinical trials as topic.sh.
44.	trial.ti.
45.	or/38-44
46.	Meta-Analysis/
47.	Meta-Analysis as Topic/
48.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
49.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
50.	(reference list* or bibliograph* or hand search* or manual search* or relevant

	journals).ab.
51.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
52.	(search* adj4 literature).ab.
53.	(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.
54.	cochrane.jw.
55.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
56.	or/46-55
57.	Epidemiologic studies/
58.	Observational study/
59.	exp Cohort studies/
60.	(cohort adj (study or studies or analys* or data)).ti,ab.
61.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
62.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
63.	Controlled Before-After Studies/
64.	Historically Controlled Study/
65.	Interrupted Time Series Analysis/
66.	(before adj2 after adj2 (study or studies or data)).ti,ab.
67.	exp case control study/
68.	case control*.ti,ab.
69.	Cross-sectional studies/
70.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
71.	Or/ 57-70
72.	exp prognosis/
73.	(prognos* or predict*).ti,ab.
74.	Logistic models/
75.	Disease progression/
76.	or/72-75
77.	37 and (45 or 56 or 71 or 76)

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.	proportional hazards model/ or hazard ratio/ or risk assessment/ or risk factors/ or decision support system/ or rating scale/ or scoring system/ or "named inventories, questionnaires and rating scales"/
24.	*bleeding/
25.	(risk* tool* or stratification or rating scale* or scor* system* or scor* schem* or risk* schem* or risk* stratif* or risk* classif* or risk* assess*).ti,ab.
26.	23 and 24 and 25
27.	ATRIA.ti,ab.
28.	((ABC or Orbit) adj2 (bleed* or scor*)).ti,ab.
29.	HEMORR2HAGES.ti,ab.
30.	"HEMORR(2)HAGES".ti,ab.
31.	*"HAS BLED Score"/
32.	(hasbled or has-bled).ti,ab.
33.	((bleed* or hemorrhag* or haemorrhag*) adj3 scor*).ti,ab.
34.	((bleed* or hemorrhag* or haemorrhag*) adj3 (risk* tool* or stratification or rating scale* or scor* system* or scor* schem* or risk* schem* or risk* stratif* or risk* classif* or risk* assess*)).ti,ab.
35.	or/26-34
36.	22 and 35
37.	systematic review/
38.	Meta-Analysis/
39.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
40.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
41.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
42.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
43.	(search* adj4 literature).ab.
44.	(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.
45.	cochrane.jw.
46.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
47.	or/37-46
48.	random*.ti,ab.
49.	factorial*.ti,ab.
50.	(crossover* or cross over*).ti,ab.
51.	((doubl* or singl*) adj blind*).ti,ab.
52.	(assign* or allocat* or volunteer* or placebo*).ti,ab.

53.	crossover procedure/
54.	single blind procedure/
55.	randomized controlled trial/
56.	double blind procedure/
57.	or/48-56
58.	Epidemiologic studies/
59.	Observational study/
60.	exp Cohort studies/
61.	(cohort adj (study or studies or analys* or data)).ti,ab.
62.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
63.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
64.	Controlled Before-After Studies/
65.	Historically Controlled Study/
66.	Interrupted Time Series Analysis/
67.	(before adj2 after adj2 (study or studies or data)).ti,ab.
68.	exp case control study/
69.	case control*.ti,ab.
70.	Cross-sectional studies/
71.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
72.	or/58-71
73.	(prognos* or predict*).ti,ab.
74.	prognosis/
75.	predictive value/
76.	or/73-75
77.	36 and (47 or 57 or 72 or 76)

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.	MeSH descriptor: [Proportional Hazards Models] this term only
#6.	MeSH descriptor: [Logistic Models] this term only
#7.	MeSH descriptor: [Risk Assessment] this term only
#8.	MeSH descriptor: [Risk Factors] this term only
#9.	MeSH descriptor: [Decision Support Systems, Clinical] this term only
#10.	MeSH descriptor: [Decision Support Techniques] this term only
#11.	(or #5-#10)
#12.	(risk* tool* or stratification or rating scale* or scor* system* or scor* schem* or risk* schem* or risk* stratif* or risk* classif* or risk* assess*):ti,ab
#13.	MeSH descriptor: [Hemorrhage] this term only
#14.	#11 and #12 and #13
#15.	ATRIA:ti,ab
#16.	((ABC or Orbit) near/2 (bleed* or scor*)):ti,ab

#17.	HEMORR2HAGES:ti,ab
#18.	HEMORR(2)HAGES:ti,ab
#19.	(hasbled or has-bled):ti,ab
#20.	((bleed* or hemorrhag* or haemorrhag*) near/3 scor*):ti,ab
#21.	((bleed* or hemorrhag* or haemorrhag*) near/3 (risk* tool* or stratification or rating scale* or scor* system* or scor* schem* or risk* schem* or risk* stratif* or risk* classif* or risk* assess*)):ti,ab
#22.	(or #14-#21)
#23.	#4 and #22

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 on Medline and Embase.

Table 23: 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

Figure2: **Flow chart of clinical study selection for the review of the effectiveness bleeding prediction tools**

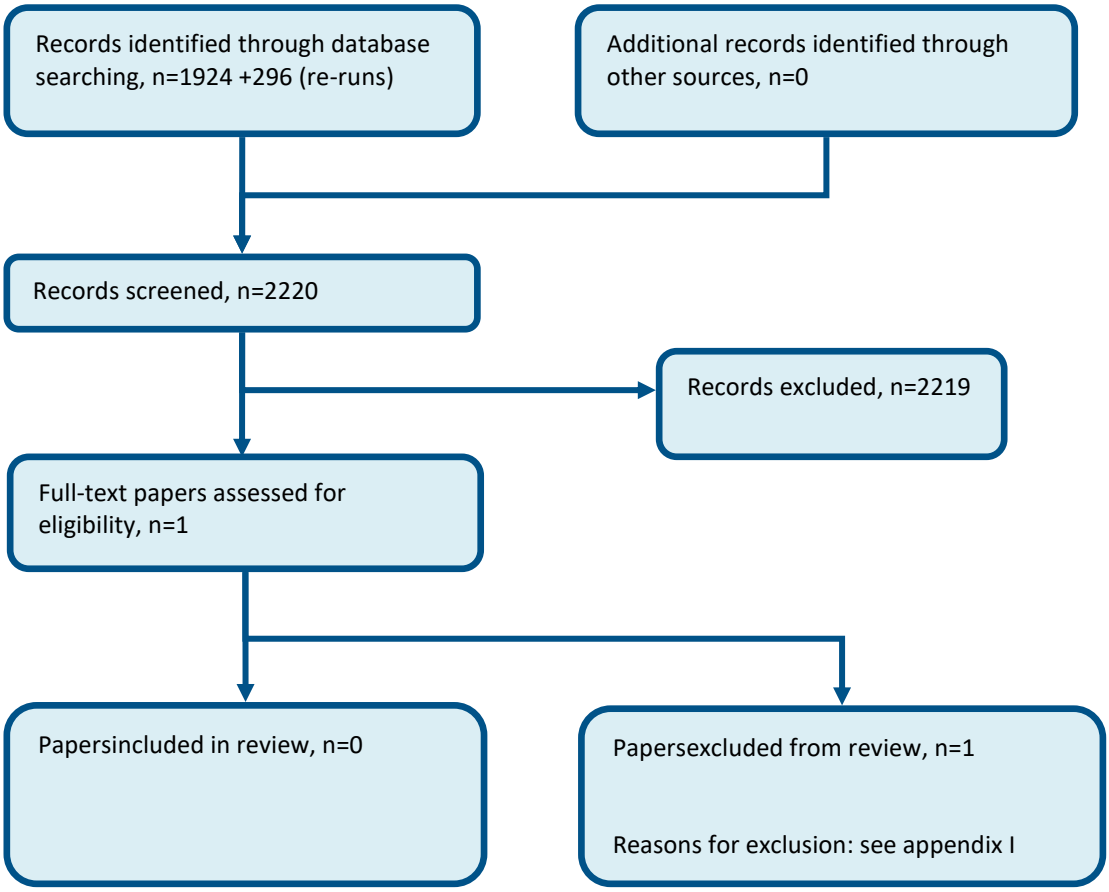
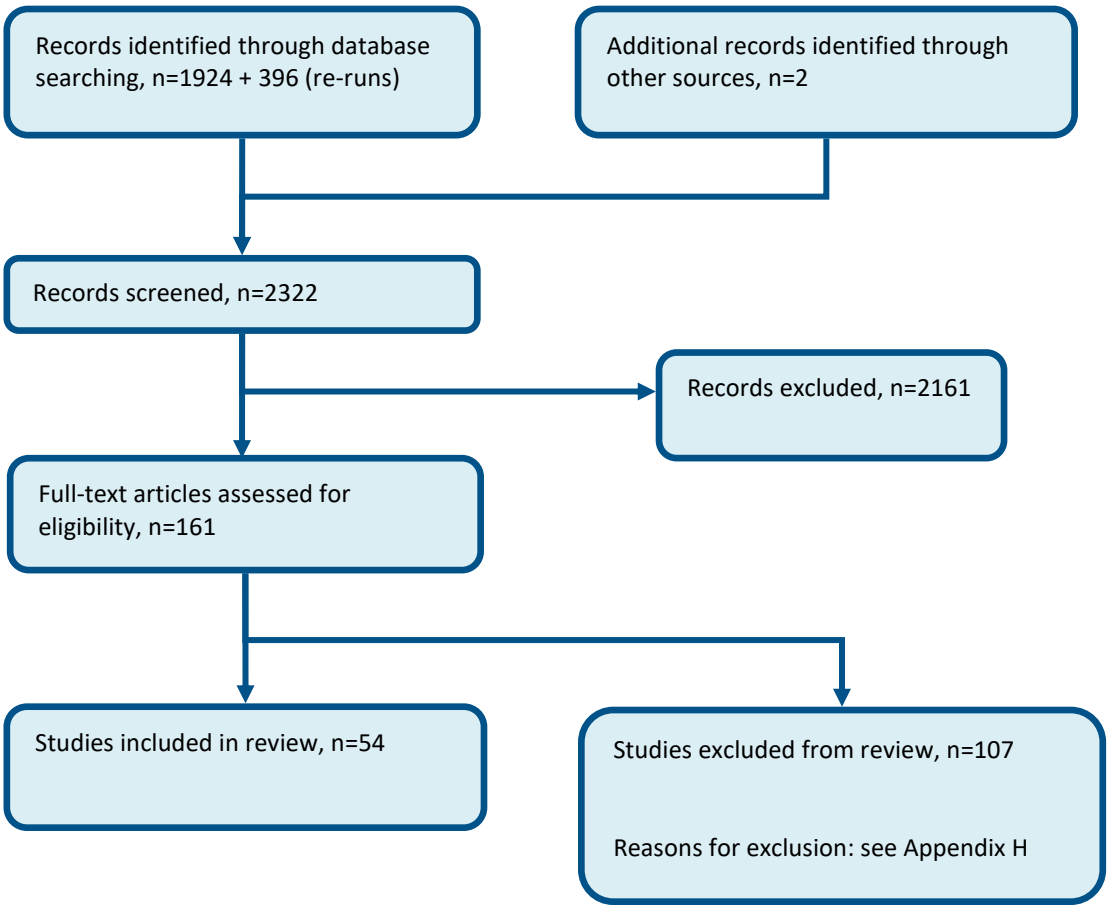
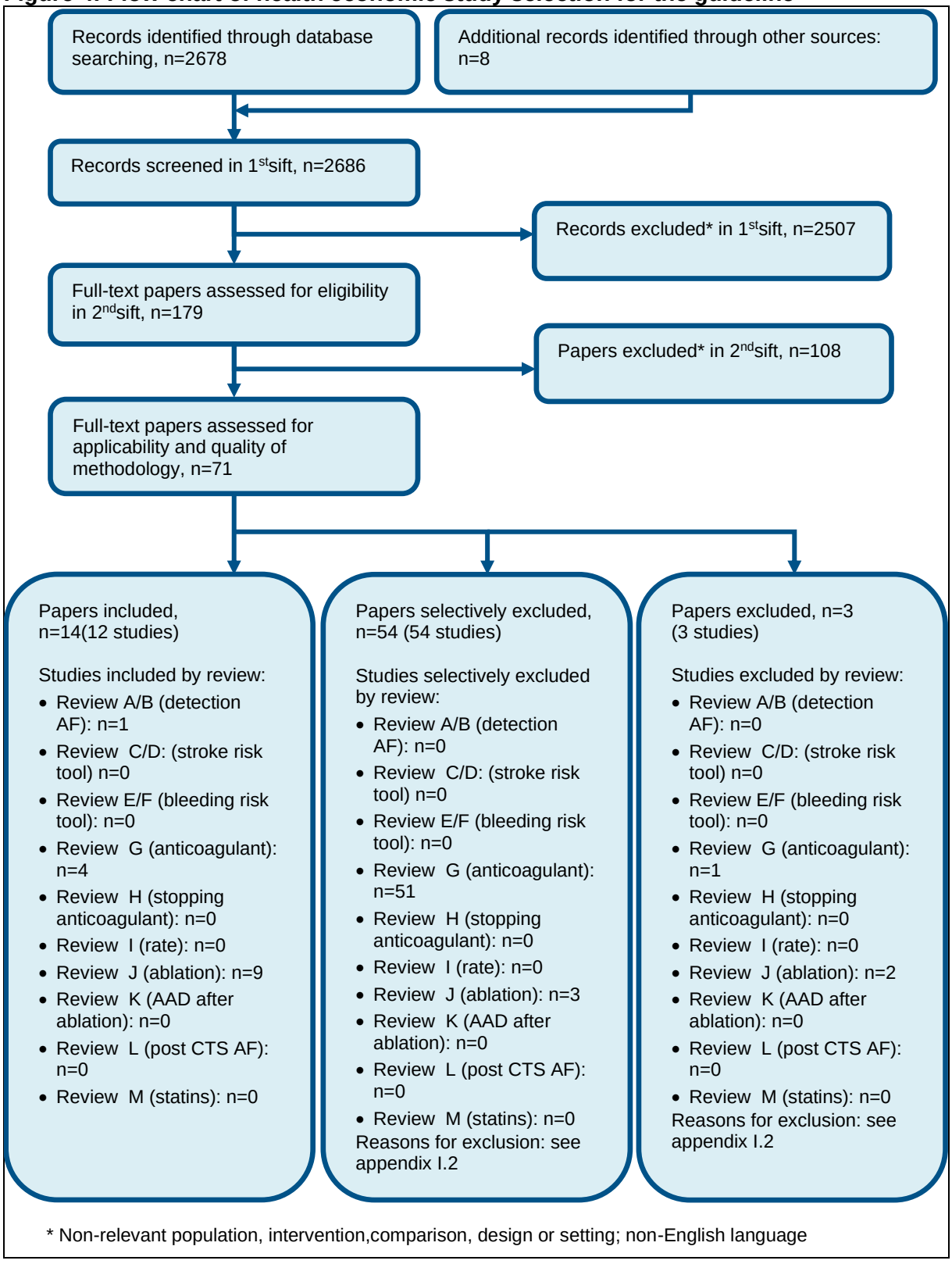


Figure 3: Flow chart of clinical article selection for the review of accuracy of risk tools for prediction of stroke



Appendix D: Economic article selection

Figure 4: Flow chart of health economic study selection for the guideline



Appendix E: Full GRADE tables(Including individual study data)

Table 24: Clinical evidence profile: accuracy of prediction of Major Bleeding in all risk tools featured in the studies (see table 3). Outcomes split across subgroups are only shown if sub-grouping was able to reduce I^2 to <50% in all sub-groups.

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
HAS-BLED	47	532,442	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.65(0.56-0.73)[4] 0.69(0.63-0.75)[8] 0.58(0.46-0.69)[14][Mixed] 0.56(0.55-0.57)[21] 0.54(0.53-0.55)[20] 0.63(0.62-0.65)[23] 0.63(0.56-0.71)[31][Mixed] 0.58(0.55-0.61)[5] 0.61(0.59-0.62)[37] 0.70(0.64-0.76)[39] 0.59(0.56-0.62)[41] 0.60(0.56-0.64)[54] 0.62(0.59-0.65)[54][DOAC] 0.62(0.59-0.64)[52][Mixed] 0.57(0.51-0.64)[58] 0.68(0.63-0.73)[58][DOAC] 0.57(0.50-0.63)[63] 0.66(0.61-0.70)[71] 0.58(0.57-0.59)[77][DOAC] 0.59(0.57-0.61)[91][Mixed] 0.80(0.76-0.83)[95] 0.69(0.59-0.80)[103]	VERY LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
							0.61(0.56-0.67)[110] **[Mixed]** 0.58(0.56-0.60)[116] 0.61(0.58-0.64)[114] **[DOAC]** 0.64(0.62-0.67)[114] **[DOAC]** 0.59(0.57-0.62)[114] 0.58(0.56-0.60)[115] 0.64(0.61-0.66)[117] 0.63(0.60-0.65)[120] 0.71(0.68-0.74)[125] 0.69(0.67-0.72)[126] 0.60(0.56-0.63)[128] 0.59(0.53-0.65)[136] 0.65(0.56-0.73)[137] 0.66(0.62-0.70)[138] 0.61(0.59-0.62)[146] **[Mixed]** 0.64(0.55-0.72)[147] 0.60(0.54-0.67)[154] **[DOAC]** 0.62(0.59-0.66)[154] 0.66(0.64-0.67)[158] **[DOAC]** 0.62 (0.60-0.64)[11] **[Mixed]** 0.60(0.56-0.63)[119] 0.62(0.57-0.68)[88] **[DOAC]** 0.64(0.63-0.65)[25] **[Mixed]** 0.66(0.64-0.68)[30] **[Mixed]** **POOLED RESULT: Random effect: 0.62 (0.61-0.64) [I^2=94%]** **Studies not pooled due to lack of variance measures:** 0.61[56]	

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
HAS-BLED with AS	1	2880	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.68(0.66-0.70)[30] **[Mixed]**	MODERATE
Modified HASBLED[135]	1	9819	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.60(0.55-0.66)[135][Mixed] ('Non-white' participants) 0.57(0.55-0.60)[135][Mixed] ('white' participants)	VERY LOW
HAS-BLED with GDF-15	1	8474	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.69(0.67-0.72)[52] **[Mixed]**	VERY LOW
HAS-BLED with vWF	2	1215	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.61(0.59-0.64)[41] 0.64(0.61-0.67)[119] **POOLED RESULT: Fixed effect: 0.62 (0.60-0.64) [I^2=6%]**	MOD
HAS-BLED + VWF + NT-proBNP	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.64(0.61-0.67)[119]	MOD
HAS-BLED + VWF + NT-proBNP + IL-6	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.64(0.61-0.67)[119]	MOD
HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.64(0.61-0.67)[119]	MOD
HAS-BLED + VWF + NT-	1	940	Serious risk of bias[a]	No serious inconsisten	No serious indirectne	No serious imprecision	0.64(0.60-0.67)[119]	MOD

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
proBNP + IL-6 + Troponin T + BTP				cy	ss			
HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T + BTP + soluble fibrin monomer complex	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.64(0.60-0.67)[119]	MOD
GEN/HAS-BLED	1	652	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.65(0.61-0.68)[138]	MOD
Modified HAS-BLED (multiple additions using biomarkers)	1	1361	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.60(0.56-0.64)[128]	MOD
Modified HAS-BLED (single change of renal dysfunction threshold)	1	231	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.67(0.57-0.75) [147]	VERY LOW
HAS-BED	1	4579	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.58(0.53-0.64)[110] **[Mixed]**	LOW
HAS-BLED with TnI	1	14,821	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.63[56] **[Mixed]**	LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
HEMORRHAGES	19	240,995	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.60(0.51-0.69)[4] 0.66(0.61-0.74)[8] 0.71(0.60-0.82)[14]**[Mixed]** 0.56(0.55-0.57)[21] 0.64(0.63-0.65)[23]**[Mixed]** 0.71(0.69-0.73)[33] 0.63(0.61-0.64)[37] 0.58(0.51-0.65)[58] 0.69(0.64-0.75)[58]**[DOAC]** 0.57(0.50-0.63)[63] 0.61(0.56-0.65)[71] 0.77 (0.73-0.81)[95] 0.64(0.53-0.75)[103]**[Mixed]** 0.61(0.58-0.64)[114]**[DOAC]** 0.66(0.64-0.69)[114]**[DOAC]** 0.59(0.56-0.62)[114] 0.55(0.52-0.57)[120] **POOLED RESULT: Random effect: 0.63 (0.60-0.66) [I^2=97%]** **Studies not pooled due to lack of variance measures:** 0.55[116] 0.67[38]	VERY LOW
HEMORRHAGES with TTR (<65% TTR)	2	4912	Serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.578[116] 0.73(0.70-0.75)[120] **Median: 0.65**	VERY LOW
ATRIA	23	286,664	Very	Very	No	No serious	0.61(0.51-0.70)[4]	VERY LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
			serious risk of bias[a]	serious risk of incon-sistency[b]	serious indirectness	imprecision	0.67(0.61-0.74[8] 0.70(0.58-0.82)[14] **[Mixed]** 0.56(0.55-0.57)[21] 0.65(0.64-0.66)[23] **[Mixed]** 0.74(0.72-0.76)[33] 0.65(0.62-0.67)[36] **[Mixed]** 0.56(0.49-0.63)[58] 0.74(0.68-0.79)[58] **[DOAC]** 0.58(0.51-0.64)[63] 0.59(0.57-0.60)[77] **[DOAC]** 0.60(0.58-0.62)[91] **[Mixed]** 0.59 (0.57-0.61)[116] 0.64(0.61-0.67)[114] **[DOAC]** 0.67(0.65-0.70)[114] **[DOAC]** 0.59(0.57-0.62)[114] 0.74(0.72-0.76)[117] 0.55(0.52-0.57)[120] 0.68(0.65-0.71)[125] 0.61(0.51-0.70)[137] 0.63(0.61-0.65)[146] **[Mixed]** 0.67(0.65-0.69)[158] **[DOAC]** 0.65(0.64-0.67)[30] **[Mixed]** **POOLED RESULT: Random effect: 0.64 (0.61-0.66) [I^2=97%]**	
ATRIA with AS	1	2880	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.67(0.66-0.69)[30] **[Mixed]**	MODERATE
ATRIA with TTR (<65% TTR)	2	4912	Serious risk of bias[a]	Very serious risk of	No serious indirectne	No serious imprecision	0.611[116] 0.75(0.73-0.77)[120] **Median: 0.68**	VERY LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
				incon-sistency[b]	ss			
ORBIT	21	270,606	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.69(0.59-0.80)[14] **[Mixed]** 0.55(0.54-0.56)[21] 0.65(0.64-0.66)[23] **[Mixed]** 0.70(0.62-0.77)[31] **[Mixed]** 0.63(0.58-0.67)[54](Warfarin) 0.70(0.67-0.73)[54] **[DOAC]** 0.68(0.65-0.70)[52] **[Mixed]** 0.56(0.48-0.64)[58] 0.73(0.68-0.78)[58] **[DOAC]** 0.61(0.59-0.62)[77] **[DOAC]** 0.63(0.61-0.65)[91] **[Mixed]** 0.59(0.57-0.61)[116] 0.68(0.65-0.71)[114] **[DOAC]** 0.70(0.68-0.73)[114] **[DOAC]** 0.62(0.59-0.64)[114] 0.57(0.54-0.59)[120] 0.58(0.52-0.64)[136] 0.61(0.51-0.70)[137] 0.66(0.64-0.68)[158] **[DOAC]** 0.64(0.59-0.70)[88] **[DOAC]** 0.67(0.65-0.68)[30] **[Mixed]** **POOLED RESULT: Random effect: 0.64 (0.61-0.67) [I^2=97%]**	VERY LOW
ORBIT with AS	1	2880	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.68(0.67-0.70)[30] **[Mixed]**	MODERATE
ORBIT with TTR (<65%	2	4912	Serious risk of bias[a]	Very serious	No serious	No serious imprecision	0.609[116] 0.73(0.71-0.76)[120]	VERY LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
TTR)				risk of incon-sistency[b]	indirectness		**Median: 0.67**	
ORBIT with GDF-15	1	8474	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.71(0.68-0.73)[52] **[Mixed]**	LOW
CHADS2	5	61,647	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.53(0.47-0.60)[8] 0.58(0.53-0.64)[110] **[Mixed]** 0.65(0.62-0.67)[117] 0.59(0.56-0.62)[126] 0.65(0.63-0.67)[158] **[DOAC]** **POOLED RESULT: Random effect: 0.61 (0.57-0.64) [I^2=85%]**	VERY LOW
CHADSVASC	8	24,402	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.56(0.49-0.62)[8] 0.54(0.48-0.61)[65] 0.56(0.509-0.618)[110] **[Mixed]** 0.65(0.62-0.67)[117] 0.58(0.55-0.60)[126] 0.55(0.51-0.58)[128] 0.68(0.66-0.70)[158] **[DOAC]** **POOLED RESULT: Random effect: 0.59 (0.54-0.64) [I^2=92%]** **Studies not pooled due to lack of variance measures:** 0.591[57] **[Mixed]**	VERY LOW
Modified CHADSVASC	1	1361	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.56(0.53-0.60)[128]	MOD

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
CHADSVASC with TnT	1	14,897	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.63(0.61-0.65) [57]	LOW
GARFIELD	3	62,172	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.61(0.59-0.63)[36] **[Mixed]** 0.56(0.54-0.57)[115] 0.64(0.63-0.65)[25] **[Mixed]** **Pooled effect: Random effects 0.60 (0.56-0.65); I2=96%**	VERY LOW
GARFIELD subgrouped by OAC - VKA	1	3550	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.56(0.54-0.58)[115]	LOW
GARFIELD subgrouped by OAC – Mixed VKA/DOACs	1	7442	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.61(0.59-0.63)[36]	LOW
GARFIELD subgrouped by antiplatelets - <33% with antiplatelets	1	3550	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.56(0.54-0.58)[96]	LOW
GARFIELD subgrouped by antiplatelets – unknown % with antiplatelets	1	7442	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.61(0.59-0.63)[30]	LOW
ABC-bleeding	3	168699	Very serious risk	Very serious risk of	No serious indirectne	Serious imprecision[c]	0.65(0.61-0.70)[54] 0.74(0.71-0.76)[54] **[DOAC]**	VERY LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
			of bias[a]	incon-sistency[b]	ss		0.69(0.66-0.71)[11]**[Mixed]** **POOLED RESULT: Random effect: 0.69(0.65-0.74) [I^2=85%]**	
ABC-bleeding Subgrouped by OAC - VKA	1	2814	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.65(0.61-0.70)[54]	VERY LOW
ABC-bleeding Subgrouped by OAC - Mixed	1	8705	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.69(0.66-0.71)[11]**[Mixed]**	VERY LOW
ABC-bleeding Subgrouped by OAC - NOACs	1	5350	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.74(0.71-0.76)[54]**[DOAC]**	VERY LOW
ABC-bleeding CrC	1	1120	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.52(0.49-0.55)[5]	LOW
ABC-bleeding cTnl-hs	2	8164	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	Serious imprecision	0.65(0.61-0.70)[VKA][54] 0.74(0.71-0.76)[54]**[DOAC]** **POOLED RESULT: Random effect: 0.70 (0.61-0.78) [I2=92%]**	VERY LOW
ABC-bleeding cTnl-hs subgrouped by OAC - VKA	1	2814	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.65(0.61-0.70)[VKA][46]	VERY LOW
ABC-bleeding	1	5350	Very	No serious	No	No serious	0.74(0.71-0.76)[46][DOAC]	LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
cTnl-hs subgrouped by OAC -DOAC			serious risk of bias[a]	inconsistency	serious indirectness	imprecision		
ABC-bleeding cystatin C	2	8164	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	Serious imprecision	0.60(0.54-0.66)[VKA][54] 0.72(0.68-0.75)[54]**[DOAC]** **POOLED RESULT: Random effect: 0.68 (0.65-0.72) [I2=90.6%]**	VERY LOW
ABC-bleeding cystatin C subgrouped by OAC - VKA	1	2814	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.60(0.54-0.66)[VKA][54]	LOW
ABC-bleeding cystatin C subgrouped by OAC - DOAC	1	5350	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.72(0.68-0.75)[54]**[DOAC]**	VERY LOW
ABC-bleeding CKD-EPI	2	8164	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	Serious imprecision	0.65(0.60-0.69)[VKA][54] 0.71(0.69-0.74)[54]**[DOAC]** **POOLED RESULT: Random effect: 0.70 (0.68-0.72) [I2=79%]**	VERY LOW
ABC-bleeding CKD-EPI subgrouped by OAC - VKA	1	2814	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.65(0.60-0.69)[VKA][54]	LOW
ABC-bleeding CKD-EPI subgrouped by OAC - DOAC	1	5350	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.71(0.69-0.74)[54]**[DOAC]**	VERY LOW
vWF	1	1215	Serious risk of bias[a]	No serious inconsistency	No serious indirectne	No serious imprecision	0.61(0.57-0.65)[41]	MOD

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
					ss			
ABS	5	81285	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	0.67(0.65-0.68)[warfarin], 0.72(0.69-0.76)[dabigatran], 0.70(0.68-0.73)[rivaroxaban], 0.72(0.67-0.77) [apixaban] [23]	VERY LOW
OBI	1	3063	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.59(0.58-0.61)[33]	LOW
Kuijer	3	8332	Very serious risk of bias[a]	Serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.56(0.55-0.58)[33] 0.52(0.48-0.56)[71] **POOLED EFFECT: Random effects: 0.54 (0.51-0.58) [I^2=72%]** **Studies not pooled due to lack of variance measures:** 0.58[38]	VERY LOW
Kearon	2	4667	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.69(0.67-0.71)[33] 0.66[38] **Median: 0.675**	LOW
Riete	1	3063	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.68(0.65-0.70)[33]	LOW
Shireman / CBRM	5	12385	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.61(0.51-0.71)[14][Mixed] 0.70(0.68-0.73)[33] 0.57(0.50-0.63)[58] 0.66(0.61-0.71)[58][DOAC] 0.63(0.58-0.67)[71]	VERY LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
							POOLED EFFECT: Random effect: 0.64(0.59-0.69) [I^2=80%]	
mOBRI/Landefield and Goldman and Beyth / Beyth	3	8762	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.56(0.51-0.60)[71] 0.54(0.42-0.66)[14]**[Mixed]** **POOLED EFFECT: Fixed effect: 0.56(0.51-0.60) [I^2=0%].** **Studies not pooled due to lack of variance measures:** 0.65[38]	LOW
TnT	1	14,897	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.62(0.60-0.64)[57]**[Mixed]**	LOW
TnI	1	14,821	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.60[56]**[Mixed]**	LOW
GDF-15	1	8474	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.67(0.65-0.69)[52]**[Mixed]**	LOW
MBR	1	40,450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.53(0.52-0.53)[21]	LOW
HTI	1	208	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.65[19]**[DOAC]**	LOW
Prothrombin	1	208	Very serious risk	No serious inconsisten	No serious	Serious imprecision[c]	0.54(0.47-0.62)[19]**[DOAC]**	VERY LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
time			of bias[a]	cy	indirectness			
Same TTR	1	4637	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.55 (0.54-0.57)[74]	LOW
APTT	1	208	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	0.58(0.50-0.69)[19]**[DOAC]**	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 few 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 around two clinical thresholds: C statisticsof 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 25: Clinical evidence profile: sensitivity and specificityof prediction of Major Bleeding in all risk tools featured in the studies (see table 3). 95% CIs are given for non-pooled results; for meta-analysed results the 95% credible intervals are given for the pooled effect only.

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
HAS-BLED at threshold of ≥1	7	128791	**Threshold at ≥1** 0.948[4] 0.921[20] 0.948[71] 0.992[116] 0.959[136] 0.994[158][DOAC] 0.99[77][DOAC] **Pooled sensitivity: 0.979(0.941-0.993)**	**Threshold at ≥1** 0.0786[4] 0.110[20] 0.209[71] 0.007[116] 0.163[136] 0.060[158][DOAC] 0.050[77][DOAC] **Pooled specificity: 0.070(0.027-0.174)**	**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
HAS-BLED at threshold of ≥2	10	177728	**Threshold at ≥2** 0.968[116] 0.846[4] 0.600[20] 0.847[31][Mixed] 0.625[71] 0.816[95] 0.446[136]	**Threshold at ≥2** 0.068[116] 0.382[4] 0.470[20] 0.320[31][Mixed] 0.560[71] 0.644[95] 0.662[136]	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Very serious imprecision[c]	VERY LOW
					Specificity				

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
			0.915[158] [DOAC] 0.890[77] [DOAC] 0.96[30] [Mixed] **Pooled sensitivity: 0.793(0.570-0.919)**	0.268[158] [DOAC] 0.230[77] [DOAC] 0.17[30] [Mixed] **Pooled specificity: 0.396(0.207-0.624)**	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	VERY LOW
HAS-BLED at threshold of ≥3	13	170197	**Threshold at ≥3** 0.456[31] [Mixed] 0.570[5] 0.338[71] 0.609[110] [Mixed] 0.787[116] 0.652[120] 0.108[136] 0.583[158] [DOAC] 0.465[95] 0.435[4] 0.630[77] [DOAC] 0.330[114] [Mixed] 0.68[30] [Mixed] **Pooled sensitivity: 0.512(0.385-0.637)**	**Threshold at ≥3** 0.706[31] [Mixed] 0.597[5] 0.8186[71] 0.408[110] [Mixed] 0.289[116] 0.598[120] 0.937[136] 0.642[158] [DOAC] 0.688[95] 0.762[4] 0.540[77] [DOAC] 0.820[114] [Mixed] 0.57[30] [Mixed] **Pooled specificity: 0.679(0.554-0.782)**	**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
HAS-BLED at threshold of ≥4	1	3525	**Threshold at ≥4** 0.543(0.453-0.632)[116]	**Threshold at ≥4** 0.591(0.575-0.608)[116]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	Serious imprecision[c]	VERY LOW

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
Modified HASBLED[135] at threshold of ≥1	1	9819	**Threshold at ≥1** 0.925 (0.902-0.945)[135] **[Mixed]**	**Threshold at ≥1** 0.1504(0.143-0.158)[135] **[Mixed]**	**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
Modified HASBLED[135] at threshold of ≥2	1	9819	**Threshold at ≥2** 0.644(0.604-0.682)[135] **[Mixed]**	**Threshold at ≥2** 0.4937(0.483-0.5040[135] **[Mixed]**	**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
Modified HASBLED[135] at threshold of ≥3	1	9819	**Threshold at ≥3** 0.311(0.275-0.349)[135] **[Mixed]**	**Threshold at ≥3** 0.826(0.819-0.834)[135] **[Mixed]**	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
HEMORRHAGES at threshold of ≥1	3	7406	**Threshold at ≥1** 0.794[4] 0.940[38] 0.953[116] **Pooled sensitivity: 0.919(0.658-0.985)**	**Threshold at ≥1** 0.345[4] 0.133[38] 0.091[116] **Pooled specificity: 0.167(0.037-0.5207)**	**Sensitivity**				
					Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecison[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Serious imprecison[c]	VERY LOW
HEMORRHAGES at threshold of ≥2	6	60023	**Threshold at ≥2** 0.358[4] 0.776[38] 0.711[95] 0.480[116] 0.824[120] 0.520[114][Mixed] **Pooled sensitivity: 0.631(0.417-0.798)**	**Threshold at ≥2** 0.768[4] 0.456[38] 0.482[95] 0.582[116] 0.269[120] 0.710[114][Mixed] **Pooled specificity: 0.549(0.349-0.734))**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Serious imprecison[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Serious imprecison[c]	VERY LOW
HEMORRHAGES at	2	5138	**Threshold at ≥3**	**Threshold at ≥3**	**Sensitivity**				

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
threshold of ≥3			0.478(0.354-0.603)[36] 0.171 (0.112-0.250)[108]	0.739(0.716-0.761)[36] 0.886(0.874-0.896)[108]	Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	No serious imprecision	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	No serious imprecision	VERY LOW
ATRIA at threshold of ≥1	4	103289	**Threshold at ≥1** 0.879[4] 0.937[116] 0.983[158]**[DOAC]** 0.930[77]**[DOAC]** **Pooled sensitivity: 0.955(0.864-0.986)**	**Threshold at ≥1** 0.113[4] 0.007[116] 0.100[158]**[DOAC]** 0.210[77]**[DOAC]** **Pooled specificity: 0.132(0.061-0.259)**	**Sensitivity**				
					Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecison[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecison[c]	VERY LOW
ATRIA at threshold of >2	5	103289	**Threshold at > 2** 0.411[4] 0.874[108] 0.776[158]**[DOAC]** 0.750[77]**[DOAC]**	**Threshold at > 2** 0.583[4] 0.615[108] 0.491[158]**[DOAC]** 0.480[77]**[DOAC]**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Serious imprecison[c]	VERY LOW

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
			0.52[114] [Mixed] **Pooled sensitivity: 0.685(0.450-0.848)**	0.71[114] [Mixed] **Pooled specificity: 0.539(0.354-0.716)**	**Specificity**				
					Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	VERY LOW
ATRIA at threshold of ≥3	3	101023	**Threshold at ≥3** 0.385[116] 0.735[158] [DOAC] 0.570[77] [DOAC] **Pooled sensitivity: 0.571(0.212-0.856)**	**Threshold at ≥3** 0.727[116] 0.541[158] [DOAC] 0.640[77] [DOAC] **Pooled specificity: 0.638(0.35446-0.861)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	No serious imprecision	VERY LOW
ATRIA at threshold of ≥4	6	111338	**Threshold at ≥4** 0.346[116] 0.296[120] 0.409[158] [DOAC] 0.300[77] [DOAC] 0.220[114] [Mixed] 0.54[30] [Mixed] **Pooled sensitivity: 0.259(0.096-0.513)**	**Threshold at ≥4** 0.985[116] 0.795[120] 0.772[158] [DOAC] 0.880[77] [DOAC] 0.930[114] [Mixed] 0.70[30] [Mixed] **Pooled specificity: 0.874(0.714-0.941)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	No serious imprecision	VERY LOW
ORBIT at threshold	4	103302	**Threshold at ≥1**	**Threshold at ≥1**	**Sensitivity**				

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
of ≥1			0.700[116] 0.743[136] 0.819[158][DOAC] 0.890[77][DOAC] **Pooled sensitivity: 0.804(0.610-0.916)**	0.432[116] 0.374[136] 0.446[158][DOAC] 0.280[77][DOAC] **Pooled specificity: 0.381(0.217-0.574)**	Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Very serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Serious imprecision[c]	VERY LOW
ORBIT at threshold of ≥2	4	103302	**Threshold at ≥2** 0.417[116] 0.297[136] 0.486[158][DOAC] 0.630[77][DOAC] **Pooled sensitivity: 0.460(0.233-0.692)**	**Threshold at ≥2** 0.722[116] 0.800[136] 0.703[158][DOAC] 0.630[77][DOAC] **Pooled specificity: 0.716(0.528-0.849)**	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	Serious imprecision[c]	VERY LOW
ORBIT at threshold of ≥3	8	114895	**Threshold at ≥3** 0.560[31][Mixed] 0.126[116] 0.34[120] 0.364[158][DOAC] 0.160[137]	**Threshold at ≥3** 0.806[31][Mixed] 0.959[116] 0.789[120] 0.831[158][DOAC] 0.930[137]	**Sensitivity**				
					Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	No serious imprecision	VERY LOW
					Specificity				

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
			0.370[77] **[DOAC]** 0.460[114] **[Mixed]** 0.48[30] **[Mixed]** **Pooled sensitivity: 0.340(0.213-0.493)**	0.840[77] **[DOAC]** 0.800[114] **[Mixed]** 0.75[30] **[Mixed]** **Pooled specificity: 0.845(0.766-0.900)**	Very serious risk of bias[a]	Serious inconsistency[a]	No serious indirectness	No serious imprecision	VERY LOW
CHADS2 at threshold of ≥1	1	39539	**Threshold at ≥1** 0.991(0.981-0.998)[158] **[DOAC]**	**Threshold at ≥1** 0.084(0.081-0.086)[158] **[DOAC]**	**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
CHADS2 at threshold of >2	1	39539	**Threshold at > 2** 0.865(0.836-0.889)[148] **[DOAC]**	**Threshold at > 2** 0.341(0.336-0.346)[148] **[DOAC]**	**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
CHADS2 at threshold of >3	1	39539	**Threshold at ≥3** 0.552(0.513-0.590)[158] **[DOAC]**	**Threshold at ≥3** 0.776(0.775-0.779)[158] **[DOAC]**	**Sensitivity**				
					Very	NA	No serious	No	LOW

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					serious risk of bias[a]		indirectness	serious imprecision	
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
CHADSVASC at threshold of ≥1	1	39539	**Threshold at ≥1** 0.998(0.992-1.00)[158] **[DOAC]**	**Threshold at ≥1** 0.385(0.366-0.404)[158] **[DOAC]**	**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
CHADSVASC at threshold of ≥2	1	39539	**Threshold at ≥2** 0.984(0.970-0.992)[158] **[DOAC]**	**Threshold at ≥2** 0.129(0.125-0.132)[158] **[DOAC]**	**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
CHADSVASC at threshold of ≥3	1	39539	**Threshold at ≥3** 0.929(0.907-0.948)[158] **[DOAC]**	**Threshold at ≥3** 0.271(0.267-0.276)[158] **[DOAC]**	**Sensitivity**				
					Very	NA	No serious	No	LOW

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
					serious risk of bias[a]		indirectness	serious imprecison	
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecison	LOW
ABC-bleedingCrC at threshold ≥2%	1	1120	**Threshold at ≥2** 0.835(0.778-0.884)[5]	**Threshold at ≥2** 0.194(0.169-0.221)[5]	**Sensitivity**				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecison	LOW
					Specificity				
					Serious risk of bias[a]	NA	No serious indirectness	No serious imprecison	LOW
HTIat threshold of ≥117 ng/ml	1	208	**Threshold ≥117 ng/ml** 0.59[19][no raw data or 95% Cis reported in paper]	**Threshold ≥117 ng/ml** 0.71[19][no raw data or 95% Cis reported in paper]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	NA	LOW
					Specificity				
					Very serious risk of bias[a]	NAS	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.Subgrouping to attempt to resolve heterogeneity was not carried out because there would always be <3 studies in any of the constituent sub-group categories, making it not possible to do a further meta-analysis within each sub-group.
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.

Table 26: NRI for major bleeding – HAS-BLED versus other tools.

Prediction tool comparison	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	[NRI(95% CI)](95%CI) [VKA COHORT UNLESS STATED]	Quality
HAS-BLED v HEMORRHAGES	5	50,051	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	+0.068(-0.1 to 0.23)[4] +0.310(0.13 to 0.49)[8] +0.043(0.027 to 0.059)[21] -0.036(-0.189 to 0.117)[63] **Pooled: Random effects NRI: + 0.080(-0.030to +0.190); I^2= 69%** **Studies not pooled due to lack of variance measures:** +0.137[116]	VERY LOW
HAS-BLED v ATRIA	6	50,988	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	+0.090(-0.09 to 0.27)[4] +0.260(0.070to 0.450)[8] +0.049(0.032 to 0.066)[21] -0.063(-0.202 to 0.0759)[63]+0.196 (-0.100to 0.490)[125] **Pooled: Random effects NRI: + 0.070(-0.020to +0.160); I^2= 52%**	VERY LOW

							Studies not pooled due to lack of variance measures: +0.088[116]	
HAS-BLED v MBR	1	40450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.056 (0.043 to 0.068)[21]	LOW
HAS-BLED v CHADS2	3	17529	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.580(0.230to 0.930)[8]+0.3826 (0.150to 0.610)[126] **Pooled fixed effect NRI: +0.440(+0.250to +0.630); I^2=0%** **Studies not pooled due to lack of variance measures:** +0.004[117]	LOW
HAS-BLED v ORBIT	3	46284	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.055 (0.038 to 0.073)[21] -0.037(-0.265 to +0.192)[136] **Pooled fixed effect NRI: +0.050(+0.040to +0.070); I^2=0%** **Studies not pooled due to lack of variance measures:** +0.008[116]	LOW
HAS-BLED v CHADSVASC	3	5518	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.36 (0.15 to 0.57)[8] +0.376 (0.15 to 0.60)[126] **Pooled fixed effect NRI: +0.37 (+0.21 to +0.52); I^2=0%** **Studies notpooled due to lack of variance measures:** +0.020[158]**[DOAC]**	LOW
HAS-BLED v ABC	1	8705	Serious risk of bias[a]	Noserious inconsistency	No serious indirectness	Serious imprecision[c]	-0.138(-0.080to 0.228)[11]	LOW

HAS-BLED v ABCCrC	1	1120	Serious risk of bias[a]	Noserious inconsistency	No serious indirectness	Serious imprecision[c]	+0.137(-0.010to 0.290)[5]	LOW
HAS-BLED v GARFIELD	1	3550	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.042(-0.087 to 0.189)[115]	VERY LOW
HAS-BLED v HAS-BLED with vWF	2	2155	Serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	-0.012(-0.080to 0.060)[41] -0.226(-0.326 to -0.004)[119] **Pooled random effect NRI: -0.12 (-0.33 to +0.09); I^2=92%**	VERY LOW
HAS-BLED v HAS-BLED + VWF + NT-proBNP	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.201(-0.329 to -0.002)[119]	MOD
HAS-BLED v HAS-BLED + VWF + NT-proBNP + IL-6	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.192(-0.325to -0.001)[119]	MOD
HAS-BLED v HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.194(-0.337 to -0.003)[119]	MOD
HAS-BLED v HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T + BTP	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.196(-0.327 to -0.005)[119]	MOD
HAS-BLED v HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T + BTP + soluble fibrin monomer	1	940	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.203(-0.342 to -0.004)[119]	MOD

complex								
HAS-BLED v Recalibrated HAS-BLED	1	Unknown	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.090(-0.123 to -0.0480)[90] **[Mixed]**	LOW
HAS-BLED v modified HAS-BLED (including multiple biomarkers)	1	1361	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.062 (-0.020to 0.140)[128]	LOW
HAS-BLED v modified HAS-BLED (including new renal dysfunction definition)	1	231	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.500(-0.820to -0.180)[147]	LOW
HAS-BLED v GEN/HAS_BLES	1	652	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.044(0.010to 0.080)[138]	MOD
HAS-BLED vs HAS-BLED with AS	1	2880	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.0481(p=0.034)[30] **[Mixed]**	MOD

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) Inconsistency was serious if I2 was 50-74% and very serious if 75% of higher
c) Imprecision serious if the 95% CIs crossed zero.

Table 27: NRI for major bleeding – ATRIA versus other tools

Prediction tool comparison	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	[NRI(95% CI)](95%CI) [VKA COHORT UNLESS STATED]	Quality
ATRIA v CHADS2	2	16159	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.590(0.240to 0.940)[8] +0.280[117] **MEDIAN: +0.43**	LOW
ATRIA v ORBIT	1	3551	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	+0.0355[116]	LOW
ATRIA v CHADSVASC	2	42139	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.590(0.240to 0.940)[8] +0.050[158]**[DOAC]** **MEDIAN:+0.32**	LOW
ATRIA v HEMORRHAGES	5	12664	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	-0.022(-0.080to 0.030)[4] +0.340(0.140to 0.540)[8] +0.027(-0.110to 0.160)[63] **Pooled random effect NRI: +0.090(-0.080to +0.207); I2=83%** **Not pooled due to lack of variance measures:** +0.289[33] +0.3128[116]	VERY LOW
ATRIA v OBI	1	3063	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	+0.505[33]	LOW
ATRIA v Kuijer	1	3063	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	+0.566[33]	LOW
ATRIA v Kearon	1	3063	Very serious	No serious	No serious	NA	+0.277[33]	LOW

			risk of bias[a]	inconsistency	indirectness			
ATRIA v Shireman	1	3063	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	+0.344[33]	LOW
ATRIA v Riete	1	3063	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	+0.448[33]	LOW
ATRIA v ATRIA with TTR<65%	3	4005	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness	No serious imprecision	-0.250[116] -0.1527(-0.240to -0.060)[120] -0.348(-0.560to -0.140)[137] **Pooled random effect NRI: -0.230(-0.410to -0.040); I^2=64%**	VERY LOW
ATRIA v MBR	1	40450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	+0.007 (-0.014 to 0.027)[21]	LOW
ATRIA vs ATRIA with AS	1	2880	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.0645(p=0.025) [30]**[Mixed]**	MOD

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) Inconsistency was serious if I2 was 50-74% and very serious if 75% of higher
c) Imprecision serious if the 95% CIs crossed zero.

Table 28: NRI for major bleeding – HEMORRHAGES versus other tools

Prediction tool comparison	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	[NRI(95% CI)](95%CI) [VKA COHORT UNLESS STATED]	Quality
HEMORRHAGES v CHADS2	1	2600	Very serious	No serious inconsistency	No serious indirectness	No serious imprecision	+0.540(0.220to 0.860)[8]	LOW

			risk of bias[a]					
HEMORRHAGES v CHADSVASC	1	2600	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.590(0.240to 0.940)[8]	LOW
HEMORRHAGES v ORBIT	1	3551	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	-0.216[116]	LOW
HEMORRHAGES v HEMORRHAGES with TTR<65%	2	1712	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.263[116] -0.059(-0.100to -0.020)[120] **MEDIAN: -0.161**	MOD
HEMORRHAGES v MBR	1	40450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.012 (-0.007 to 0.032)[21]	VERY 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) Inconsistency was serious if I2 was 50-74% and very serious if 75% of higher
c) Imprecision serious if the 95% CIs crossed zero.

Table 29: NRI for major bleeding – ORBIT versus other tools

Prediction tool comparison	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	[NRI(95% CI)](95%CI) [VKA COHORT UNLESS STATED]	Quality
ORBIT v ORBIT with TTR<65%	3	4009	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	-0.109 (-0.180to -0.040)[120] -0.348(-0.560to -0.140)[137] **Pooled random effect NRI: -0.21 (-0.44 to 0.02); I^2=77%** **Not pooled due to lack of variance measures:**	VERY LOW

							-0.251[116]	
ORBIT v CHADSVASC	1	39539	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	NA	+0.010[158]**[DOAC]**	LOW
ORBIT v MBR	1	40450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	0.000 (-0.021 to 0.021)[21]	VERY LOW
ORBIT vs ORBIT with AS	1	2880	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision	-0.014(p=0.170) [30]**[Mixed]**	VERY 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) Inconsistency was serious if I2 was 50-74% and very serious if 75% of higher
c) Imprecision serious if the 95% CIs crossed zero.

Table 30: NRI for major bleeding – CHADSVASC versus other tools

Prediction tool comparison	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	[NRI(95% CI)](95%CI) [VKA COHORT UNLESS STATED]	Quality
CHADSVASC v CHADS2	3	55698	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.071 (-0.050to 0.190)[8] -0.129[117] +0.040[158]**[DOAC]** **MEDIAN: +0.040**	VERY LOW
CHADSVASC v modified CHADSVASC (including multiple biomarkers)	1	1361	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.0026 (-0.020to 0.030)[128]	VERY 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) Inconsistency was serious if I2 was 50-74% and very serious if 75% of higher
c) Imprecision serious if the 95% CIs crossed zero.

Table 31: Clinical evidence profile: accuracy of prediction of CRB in all risk tools featured in the studies (see table 3). Outcomes split across subgroups are only shown if sub-grouping was able to reduce I2 to <50% in all sub-groups.

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
HAS-BLED	8	18258	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.60(0.56-0.63)[4] 0.51(0.45-0.58)[14] **[Mixed]** 0.55(0.53-0.56)[32] 0.50(0.47-0.54)[63] 0.58(0.54-0.63)[113] 0.56(0.54-0.58)[115] 0.61(0.58-0.64)[136] 0.59(0.56-0.63)[137] **POOLED RESULT: Random effect: 0.56(0.54-0.59). I^2=83%**	VERY LOW
HEMORRHAGES	3	4467	Very serious risk of bias[a]	Serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.55(0.51-0.59)[4] 0.61(0.55-0.68)[14] **[Mixed]** 0.53(0.50-0.57)[63] **Pooled effect: Random effects 0.56 (0.52-0.60); I2=64%**	VERY LOW
HEMORRHAGES	2	3450	Very serious risk of	No serious risk of	No serious indirectness	No serious imprecision	0.55(0.51-0.59)[4] 0.53(0.50-0.57)[52] **Pooled effect: fixed effect 0.54(0.51-0.56); I2=0%**	LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
subgrouped by OAC - VKA			bias[a]	incon-sistency				
HEMORRHAGES subgrouped by OAC – Mixed VKA/DOAC	1	1157	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.61(0.55-0.68)[14] **[Mixed]**	LOW
HEMORRHAGES subgrouped by antiplatelets - <33%	2	3450	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.55(0.51-0.59)[4] 0.53(0.50-0.57)[52] **Pooled effect: 0.54(0.51-0.56); I2=0%**	LOW
HEMORRHAGES subgrouped by antiplatelets - >33%	1	1157	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.61(0.55-0.68)[14]	LOW
ATRIA	4	6760	Very serious risk of	Serious risk of incon-	No serious indirectness	Serious imprecision	0.50(0.46-0.54)[4] 0.61(0.54-0.67)[14] **[Mixed]**	VERY LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
			bias[a]	sistency[b]			0.52(0.49-0.56)[63] 0.50(0.46-0.53)[137] **Pooled effect: Random Effects 0.52 (0.49-0.56); I^2=63%**	
ATRIA subgrouped by OAC - VKA	3	5743	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	Serious imprecision[c]	0.50(0.46-0.54)[4] 0.52(0.49-0.56)[63] 0.50(0.46-0.53)[137] **Pooled effect: fixed effects 0.51(0.49-0.53); I^2=0%**	VERY LOW
ATRIA subgrouped by OAC – Mixed VKA/DOACs	1	1017	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.61(0.54-0.67)[14]**[Mixed]**	LOW
ATRIA subgrouped by antiplatelets – <33%	4	5743	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	Serious imprecision[c]	0.50(0.46-0.54)[4] 0.52(0.49-0.56)[63] 0.50(0.46-0.53)[137] **Pooled effect: fixed effects 0.51(0.49-0.53); I^2=0%**	VERY LOW
ATRIA subgrouped by antiplatelets – >33%	4	1017	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.61(0.54-0.67)[14]**[Mixed]**	LOW
ORBIT	3	5593	Very serious risk of	Very serious risk of	No serious indirectness	No serious imprecision	0.61(0.54-0.68)[14]**[Mixed]** 0.58(0.55-0.61)[136] 0.52(0.48-0.56)[137]	VERY LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
			bias[a]	incon-sistency[b]			**Pooled effect: Random Effects 0.57(0.52-0.61); I^2=73%**	
ORBIT subgrouped by antiplatelets - <33%	1	2293	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	Serious imprecision[c]	0.52(0.48-0.56)[137]	VERY LOW
ORBIT subgrouped by antiplatelets - >33%	1	1017	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.61(0.54-0.68)[14] **[Mixed]**	LOW
ORBIT subgrouped by antiplatelets – not reported	1	2283	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.58(0.55-0.61)[136]	LOW
CHADS 2	1	2293	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	Serious imprecision[c]	0.51(0.47-0.55)[3]	VERY LOW
CHADS VASC	1	2293	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	Serious imprecision[c]	0.53(0.49-0.57)[3]	VERY LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
GARFIELD	1	3550	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.57(0.55-0.58)[115]	LOW
MBRFS	1	4576	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.53(0.52-0.54)[32]	LOW
mOBRI	1	1017	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.56(0.50-0.62)[14] **[Mixed]**	LOW
CBRM /Shireman	1	1017	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.58(0.54-0.62)[14] **[Mixed]**	LOW
Simplified HAS-BLED	1	1089	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.642(0.60-0.68)[113]	LOW
HAS-BLED with point for sustained AF	1	1089	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.61(0.57-0.65)[113]	LOW

GRADE was conducted with emphasis on C statistics as this was the primary measure discussed in decision making.

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 few 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 around two clinical thresholds: C statisticsof 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 32: Clinical evidence profile: sensitivity and specificityof prediction of clinically relevant bleedingin all risk tools featured in the studies (see table 3). 95% CIs are given for non-pooled results.

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
HAS-BLED at threshold ≥1	2	4566	**Threshold at ≥1** 0.952(0.920-0.980)[4] 0.913(0.880-0.940)[136] **Median[d]: 0.913(0.880-0.940)**	**Threshold at ≥1** 0.081(0.070-0.090)[4] 0.171(0.160-0.190[136] **Median[d]: 0.171(0.160-0.190**	**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[b]	No serious indirectness	No serious imprecision	VERY LOW

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
HAS-BLED at threshold ≥2	2	4566	**Threshold at ≥2** 0.730(0.670-0.790)[4] 0.496(0.440-0.550)[136] **Median[d]: 0.496(0.440-0.550)**	**Threshold at ≥2** 0.390(0.370-0.410)[4] 0.686(0.670-0.710)[136] **Median[d]: 0.686(0.670-0.710)**	**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	No serious imprecision	VERY LOW
HAS-BLED at threshold ≥3	2	4566	**Threshold at ≥3** 0.370(0.310-0.430)[4] 0.110(0.080-0.150)[136] **Median[d]: 0.110(0.080-0.150)**	**Threshold at ≥3** 0.770(0.760-0.790)[4] 0.950(0.940-0.960)[136] **Median[d]: 0.950(0.940-0.960)**	**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	No serious imprecision	VERY LOW
ATRIA at threshold ≥1	1	2268	**Threshold at ≥1** 0.879(0.832-0.917)[4]	**Threshold at ≥1** 0.113(0.099-0.128)[4]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	Serious imprecision[c]	VERY LOW

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
ATRIA at threshold ≥2	1	2268	**Threshold at ≥2** 0.411(0.349-0.475)[4]	**Threshold at ≥2** 0.583(0.561-0.605)[4]	**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
Hemmorhages at threshold ≥1	1	2268	**Threshold at ≥1** 0.742(0.683-0.795)[4]	**Threshold at ≥1** 0.353(0.332-0.374)[4]	**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
Hemmorhages at threshold ≥2	1	2268	**Threshold at ≥2** 0.266(0.212-0.326)[4]	**Threshold at ≥2** 0.779(0.770-0.788)[4]	**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

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
ORBIT at threshold ≥1	1	2283	**Threshold at ≥1** 0.734(0.684-0.779)[136]	**Threshold at ≥1** 0.388(0.367-0.411)[136]	**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
ORBIT at threshold ≥2	1	2283	**Threshold at ≥2** 0.283(0.236-0.334[136]	**Threshold at ≥2** 0.812(0.793-0.829)[136]	**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
CHADS2 at threshold ≥1	1	2293	**Threshold at ≥1** 0.972(0.943-0.988)[3]	**Threshold at ≥1** 0.0230(0.170-0.305)[3]	**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

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
CHADS2 at threshold ≥2	1	2293	**Threshold at ≥2** 0.637(0.575-0.697)[3]	**Threshold at ≥2** 0.385(0.364-0.406)[3]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
CHADSVASC at threshold ≥2	1	2293	**Threshold at ≥2** 0.936(0.899-0.963)[3]	**Threshold at ≥2** 0.079(0.069-0.093)[3]	**Sensitivity**				
					Very serious risk of bias[a]	NA	No serious indirectness	Serious imprecision[c]	VERY LOW
					Specificity				
					Very serious risk of bias[a]	NA	No serious indirectness	No serious imprecision	LOW
CHADSVASC at threshold ≥3	1	2293	**Threshold at ≥3** 0.753(0.695-0.805)[3]	**Threshold at ≥3** 0.292(0.273-0.313)[3]	**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

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 33: NRI for clinically relevant bleeding

Prediction tool comparison	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	[NRI(95% CI)](95%CI) [VKA COHORT UNLESS STATED]	Quality
HAS-BLED v HEMORRHAGES	2	3450	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	+0.103(0.040to 0.160)[4] -0.056(-0.140to 0.028)[63] **Pooled: Random effects NRI: + 0.030(-0.130to +0.180); I^2= 89%**	VERY LOW
HAS-BLED v ATRIA	2	3450	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	+0.130(0.050to 0.210)[4] -0.056(-0.130to 0.014)[63] **Pooled: Random effects NRI: + 0.040(-0.150to +0.220); I^2= 92%**	VERY LOW
ATRIA v HEMORRHAGES	2	3450	Very serious risk of bias[a]	Very serious inconsistency[b]	No serious indirectness	Serious imprecision[c]	+0.130 (0.050to 0.210)[4] +0.0003(-0.076 to 0.076)[63] **Pooled: Random effects NRI: + 0.060(-0.060to +0.190); I2 =**	VERY LOW

							81%	
HAS-BLED v CHADS2	1	2293	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.130(0.050to 0.210)[3]	LOW
HAS-BLED v GARFIELD	1	3550	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	-0.033(-0.129 to 0.094)[115]	VERY LOW
HAS-BLED v CHADSVASC	1	2293	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.130(0.050to 0.210)[3]	LOW
HAS-BLED v ORBIT	1	2283	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.156(0.043 to 0.27)[136]	MOD
ATRIA v ATRIA +TTR	1	2293	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.260 (-0.480to -0.040)[137]	LOW
ORBIT v ORBIT + TTR	1	2293	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.260 (-0.480to -0.040)[137]	MOD

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) Inconsistency was serious if I2 was 50-74% and very serious if 75% of higher
c) Imprecision serious if the 95% CIs crossed zero.

Table 34: Clinical evidence profile: accuracy of prediction of ICH in all risk tools featured in the studies (see table 3). Outcomes split across subgroups are only shown if sub-grouping was able to reduce I2 to <50% in all sub-groups.

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
HAS-BLED	7	110,194	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.53(0.51-0.54)[21] 0.56(0.49-0.63)[5] 0.60(0.58-0.68)[37] 0.52(0.42-0.63)[114]**[DOAC]** 0.56(0.48-0.64)[114]**[DOAC]** 0.57(0.52-0.67)[114] 0.57(0.52-0.63)[142] **Pooled effect: Random effects 0.56(0.53-0.60); I^2=83%**	VERY LOW
HAS-BLED subgrouped by antiplatelets - <33%	1	40,450	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.53(0.51-0.54)[21]	LOW
HAS-BLED subgrouped by antiplatelets - >33%	3	18.113	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.52(0.42-0.63)[114]**[DOAC]** 0.56(0.48-0.64)[114]**[DOAC]** 0.57(0.52-0.62)[114] **Pooled effect: fixed effects 0.56(0.52-0.60); I2=0%**	LOW
HAS-BLED subgrouped by	3	51631	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.56(0.49-0.63)[5] 0.60(0.58-0.68)[37] 0.57(0.52-0.63)[142]	LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
antiplatelets – not reported							**Pooled effect: fixed effects 0.59(0.58-0.61); I2=0%**	
HEMORRHAGES	5	107,162	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.53(0.51-0.54)[21] 0.62(0.60-0.64)[37] 0.54(0.44-0.65)[114]**[DOAC]** 0.61(0.52-0.70)[114]**[DOAC]** 0.60(0.55-0.66)[114] **Pooled effect: Random effects: 0.58(0.52-0.64); I2=93%**	VERY LOW
HEMORRHAGES subgrouped by antiplatelets – <33%	1	40,450	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.53(0.51-0.54)[21]	LOW
HEMORRHAGES subgrouped by antiplatelets – >33%	3	18,113	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.54(0.44-0.65)[114]**[DOAC]** 0.61(0.52-0.70)[114]**[DOAC]** 0.60(0.55-0.66)[114] **Pooled effect: fixed effects 0.59(0.55-0.63); I2=0%**	LOW
HEMORRHAGES	1	48,599	Very serious risk of bias[a]	No serious risk of incon-	No serious indirectness	No serious imprecision	0.62(0.60-0.64)[37]	LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
subgrouped by antiplatelets – not reported				sistency				
ATRIA	4	58,563	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.50(0.49-0.52)[21] 0.59(0.50-0.69)[114]**[DOAC]** 0.59(0.50-0.68)[114]**[DOAC]** 0.58(0.52-0.66)[114] **Pooled effect: Random effects 0.56(0.50-0.61); I2=75%**	VERY LOW
ATRIA subgrouped for antiplatelets - <33%	1	40,450	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	Serious imprecision[c]	0.50(0.49-0.52)[21]	VERY LOW
ATRIA subgrouped for antiplatelets - >33%	3	18.113	Very serious risk of bias[a]	No serious risk of incon-sistency	No serious indirectness	No serious imprecision	0.59(0.50-0.69)[114]**[DOAC]** 0.59(0.50-0.68)[114]**[DOAC]** 0.58(0.52-0.66)[114] **Pooled effect: fixed effects 0.58(0.54-0.63); I2=0%**	LOW
ORBIT	4	58,563	Very serious risk of bias[a]	Very serious risk of incon-sistency[b]	No serious indirectness	No serious imprecision	0.50(0.48-0.51)[21] 0.63(0.55-0.72)[114]**[DOAC]** 0.60(0.50-0.69)[114]**[DOAC]** 0.62(0.57-0.67)[114] **Pooled effect: Random effects 0.58(0.50-0.67); I2=91%**	VERY LOW

Risk tool	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area Under Curve Individual study effects [point estimate (95% Cis)] Pooled effect/range /median [VKA COHORT UNLESS STATED]	Quality
ORBIT subgrouped for antiplatelets - <33%	1	40,450	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	serious imprecision[c]	0.50(0.48-0.51)[21]	VERY LOW
ORBIT subgrouped for antiplatelets - >33%	3	18,113	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	No serious imprecision	0.63(0.55-0.72)[114]**[DOAC]** 0.60(0.50-0.69)[114]**[DOAC]** 0.62(0.57-0.67)[114] **Pooled effect: fixed effects 0.62(0.58-0.66); I2=0%**	LOW
ABCBleeding CrC	1	1120	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	Serious imprecision[c]	0.47(0.40-0.53)[5]	VERY LOW
MBR	1	40450	Very serious risk of bias[a]	No serious risk of inconsistency	No serious indirectness	No serious imprecision	0.52(0.50-0.53)[21]	LOW

GRADE was conducted with emphasis on C statisticsas this was the primary measure discussed in decision making.
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 few 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 around 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 35: Clinical evidence profile: sensitivity and specificityof prediction of intracranial hemmorhagein all risk tools featured in the studies (see table 3). 95% CIs are given for non-pooled results.

Risk tool	No of COHORTS	n	Sensitivity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Specificity (threshold denotes the 'positive' score – i.e. the score indicating a high risk of bleeding) [VKA COHORT UNLESS STATED]	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
HAS-BLEDat threshold ≥3	1		**Threshold ≥3** 0.538(0.410-0.660)[5]	**Threshold ≥3** 0.572(0.540-0.600)[5]	**Sensitivity**				
					Serious risk of bias[a]	NA	No serious indirectnes	Serious imprecision[c]	LOW
					Specificity				
					Serious risk of bias[a]	NA	No serious indirectnes	No serious imprecision	MOD
ABCBleeding CrCat threshold ≥2%	1		**Threshold ≥2** 0.785(0.670-0.880)[5]	**Threshold ≥2** 0.186(0.160-0.210)[5]	**Sensitivity**				
					Serious risk of bias[a]	NA	No serious indirectnes	No serious imprecision	MOD
					Specificity				
					Serious risk of bias[a]	NA	No serious indirectnes	No serious imprecision	MOD

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.

Table 36: NRI for intracranial bleeding

Prediction tool comparison	No of COHORTS	n	Risk of bias	Inconsistency	Indirectness	Imprecision	[NRI(95% CI)](95%CI) [VKA COHORT UNLESS STATED]	Quality
HAS-BLED v HEMORRHAGES	1	40,450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.030(-0.001 to 0.060)[21]	VERY LOW
HAS-BLED v ATRIA	1	40,450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.060(0.026 to 0.093)[21]	LOW
HAS-BLED V ORBIT	1	40,450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	+0.048(0.013 to 0.082)[21]	LOW
HAS-BLED v MBR	1	40,450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	+0.007(-0.018 to 0.033)[21]	VERY LOW
HAS-BLED v	1	1120	Serious	No serious	No serious	Serious	+0.139(-0.010to 0.290)[5]	LOW

ABCbleeding CrC			risk of bias[a]	inconsistency	indirectness	imprecision[c]		
MBR v HEMORRHAGES	1	40,450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	-0.022(-0.062 to 0.017)[21]	VERY LOW
MBR v ATRIA	1	40,450	Very serious risk of bias[a]	No serious inconsistency	No serious indirectness	No serious imprecision	-0.052(-0.094 to -0.011)[21]	LOW
MBR v ORBIT	1	40,450	Serious risk of bias[a]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	-0.040(-0.083 to 0.002)[21]	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) Inconsistency was serious if I2 was 50-74% and very serious if 75% of higher
c) Imprecision serious if the 95% CIs crossed zero.

Appendix F:Forest plots

F.1 C statistics

Note that Forest plots are not shown for tools with only a single study. The sub-grouped analyses are shown regardless of whether the sub-groups succeeded in reducing heterogeneity to I^2<50% in all sub-groups.

C STATISTICS FOR MAJOR BLEEDING

Figure 5: HAS-BLED (sub-grouped for OAC type)

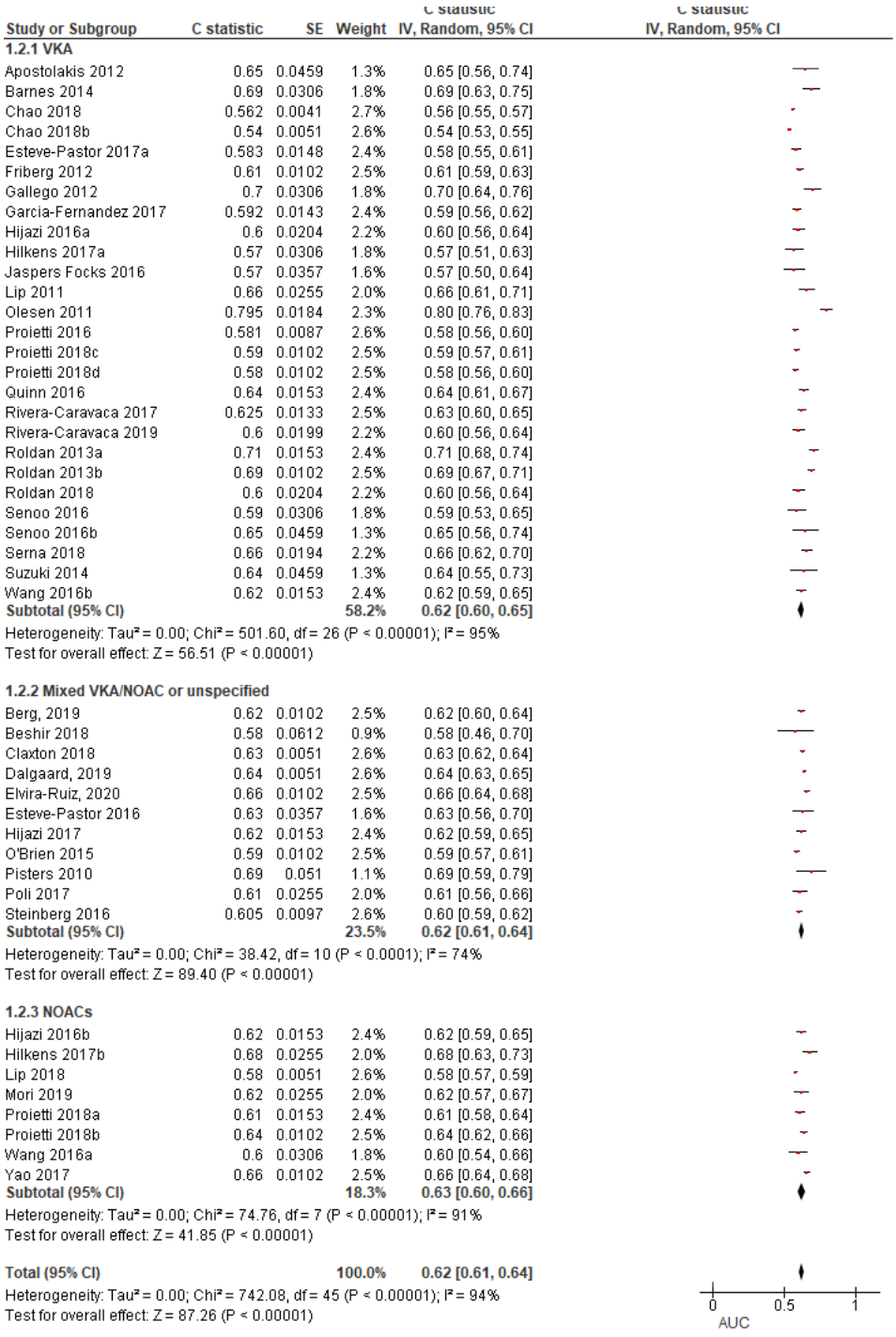


Figure 6: HAS-BLED (sub-grouped for antiplatelets)

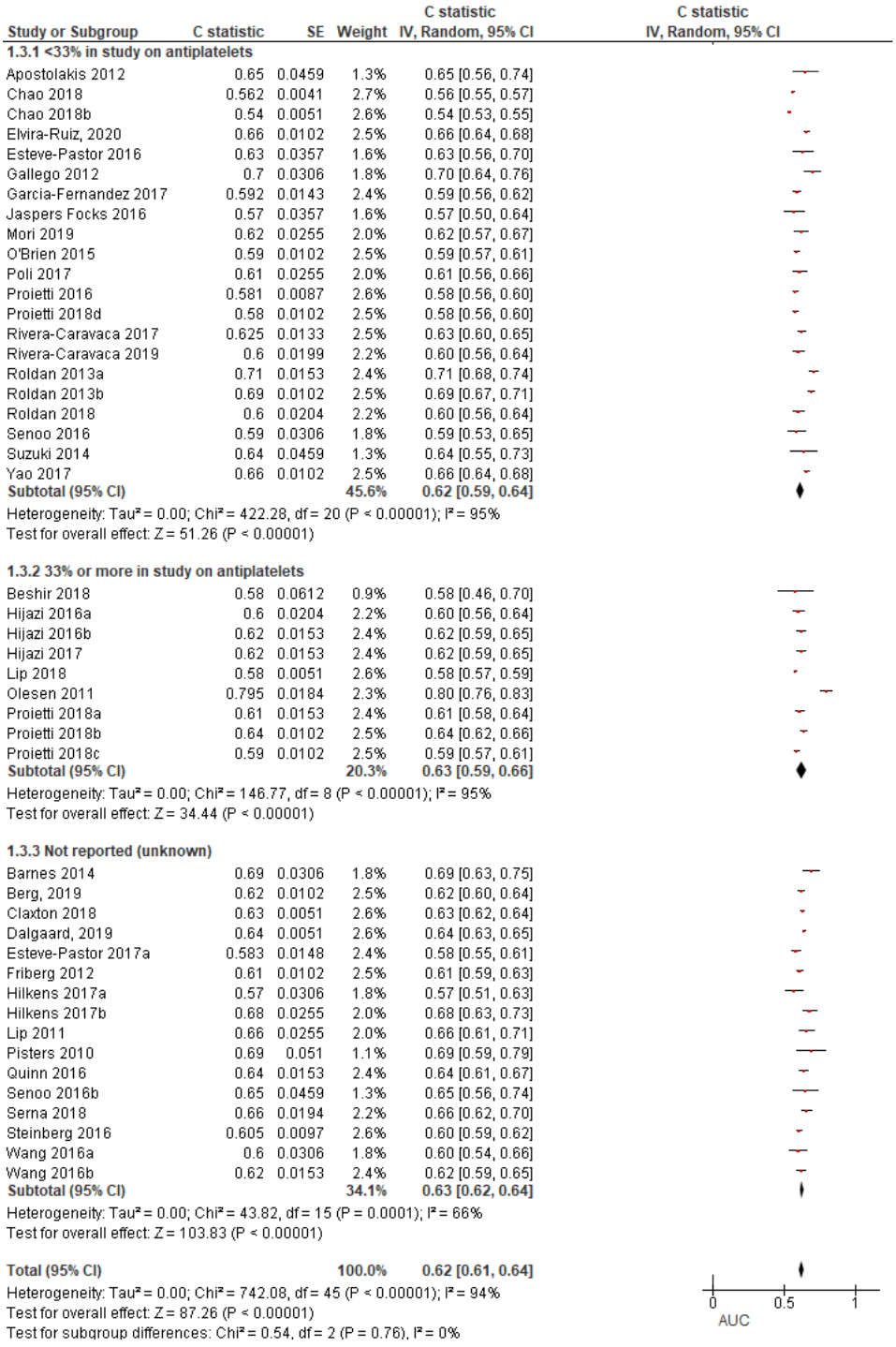


Figure 7: HAS-BLED with vWF (both VKA and <33% antiplatelets)

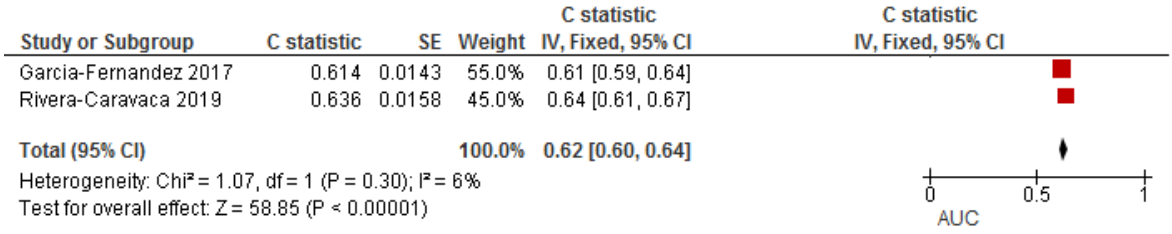


Figure 8: HEMORRHAGES (sub-grouped for OAC type)

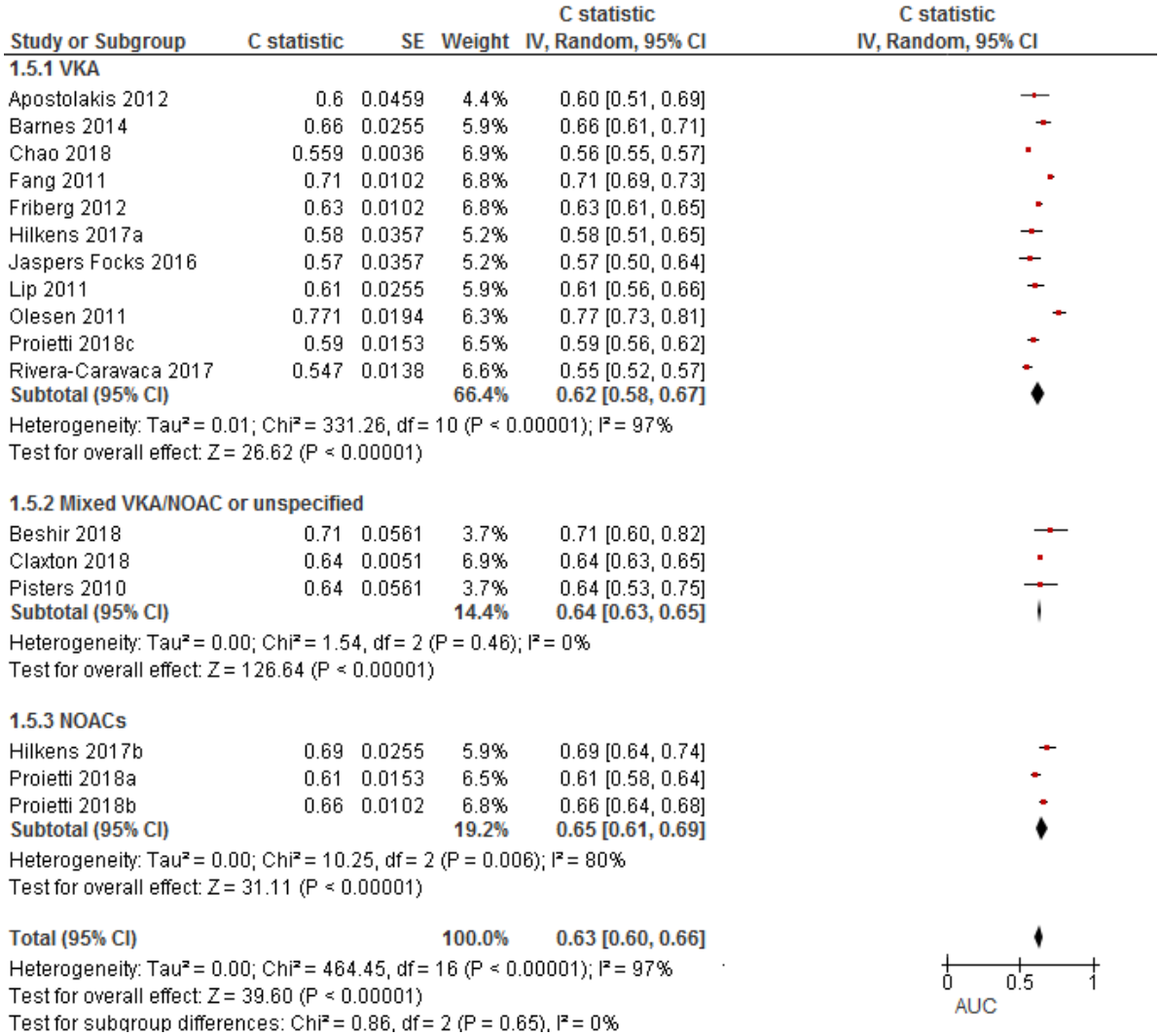


Figure 9: HEMORRHAGES (sub-grouped for antiplatelets)

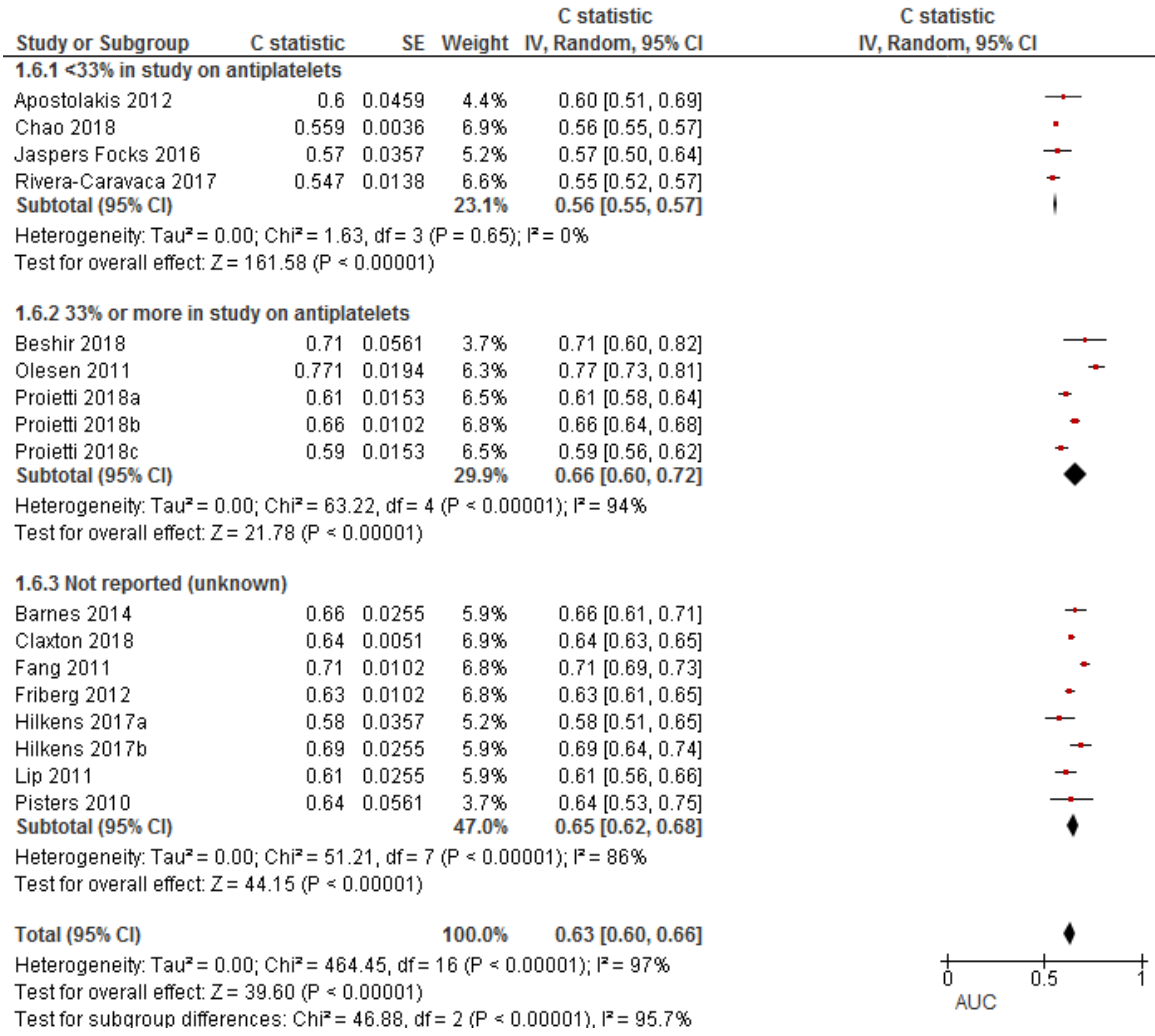

Study or Subgroup
C statistic
SE
Weight
C statistic IV, Random, 95% CI
C statistic IV, Random, 95% CI
1.6.1 <33% in study on antiplatelets
Apostolakis 2012 0.6 0.0459 4.4% 0.60 [0.51, 0.69]
Chao 2018 0.559 0.0036 6.9% 0.56 [0.55, 0.57]
Jaspers Focks 2016 0.57 0.0357 5.2% 0.57 [0.50, 0.64]
Rivera-Caravaca 2017 0.547 0.0138 6.6% 0.55 [0.52, 0.57]
Subtotal (95% CI) 23.1% 0.56 [0.55, 0.57]
Heterogeneity: Tau² = 0.00; Chi² = 1.63, df = 3 (P = 0.65); I² = 0%
Test for overall effect: Z = 161.58 (P < 0.00001)
1.6.2 33% or more in study on antiplatelets
Beshir 2018 0.71 0.0561 3.7% 0.71 [0.60, 0.82]
Olesen 2011 0.771 0.0194 6.3% 0.77 [0.73, 0.81]
Proietti 2018a 0.61 0.0153 6.5% 0.61 [0.58, 0.64]
Proietti 2018b 0.66 0.0102 6.8% 0.66 [0.64, 0.68]
Proietti 2018c 0.59 0.0153 6.5% 0.59 [0.56, 0.62]
Subtotal (95% CI) 29.9% 0.66 [0.60, 0.72]
Heterogeneity: Tau² = 0.00; Chi² = 63.22, df = 4 (P < 0.00001); I² = 94%
Test for overall effect: Z = 21.78 (P < 0.00001)
1.6.3 Not reported (unknown)
Barnes 2014 0.66 0.0255 5.9% 0.66 [0.61, 0.71]
Claxton 2018 0.64 0.0051 6.9% 0.64 [0.63, 0.65]
Fang 2011 0.71 0.0102 6.8% 0.71 [0.69, 0.73]
Friberg 2012 0.63 0.0102 6.8% 0.63 [0.61, 0.65]
Hilkens 2017a 0.58 0.0357 5.2% 0.58 [0.51, 0.65]
Hilkens 2017b 0.69 0.0255 5.9% 0.69 [0.64, 0.74]
Lip 2011 0.61 0.0255 5.9% 0.61 [0.56, 0.66]
Pisters 2010 0.64 0.0561 3.7% 0.64 [0.53, 0.75]
Subtotal (95% CI) 47.0% 0.65 [0.62, 0.68]
Heterogeneity: Tau² = 0.00; Chi² = 51.21, df = 7 (P < 0.00001); I² = 86%
Test for overall effect: Z = 44.15 (P < 0.00001)
Total (95% CI) 100.0% 0.63 [0.60, 0.66]
Heterogeneity: Tau² = 0.00; Chi² = 464.45, df = 16 (P < 0.00001); I² = 97%
Test for overall effect: Z = 39.60 (P < 0.00001)
Test for subgroup differences: Chi² = 46.88, df = 2 (P < 0.00001), I² = 95.7%
0
0.5
1
AUC

Figure 10: ATRIA (sub-grouped for OAC type)

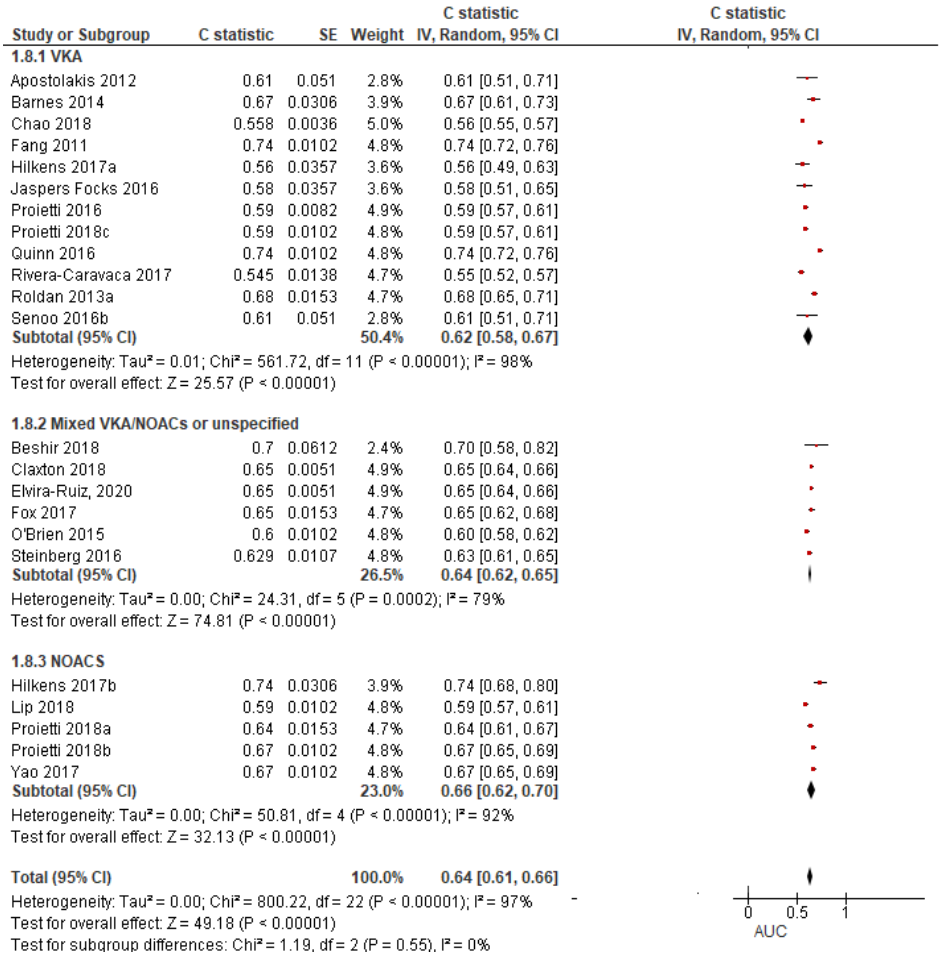


Figure 11: ATRIA (sub-grouped for antiplatelets)

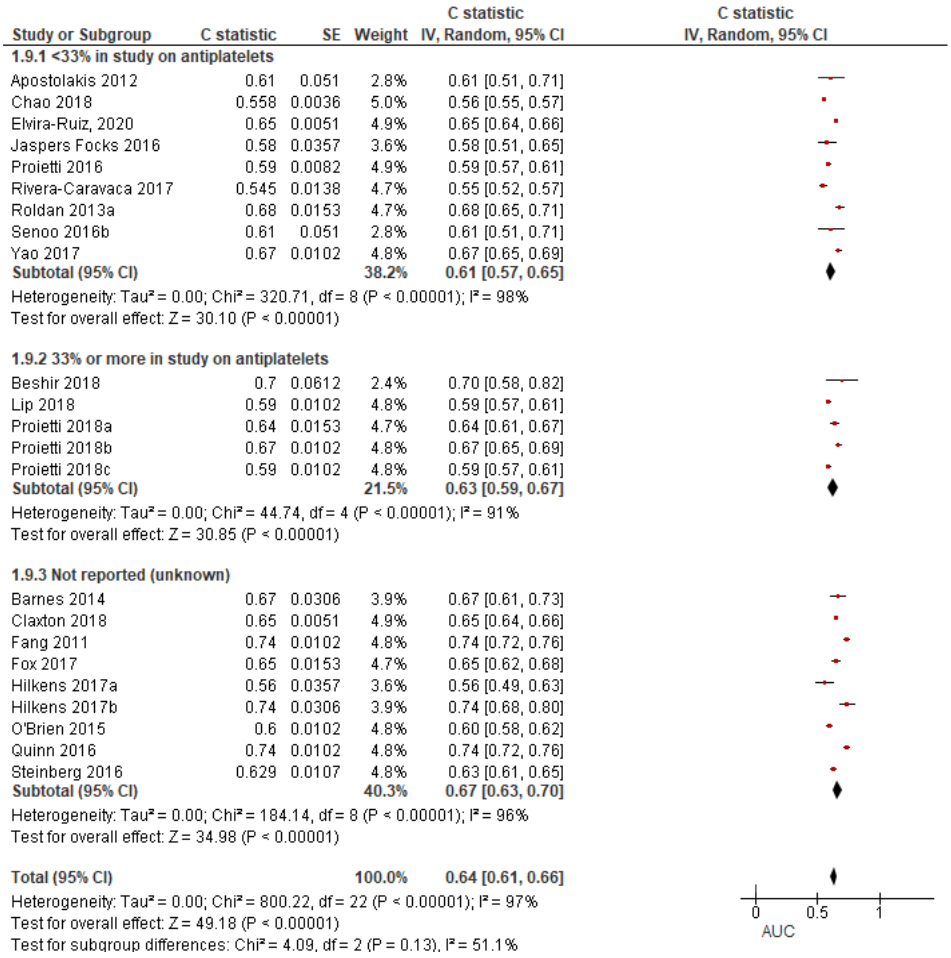


Figure 12: ORBIT (sub-grouped for OAC type)

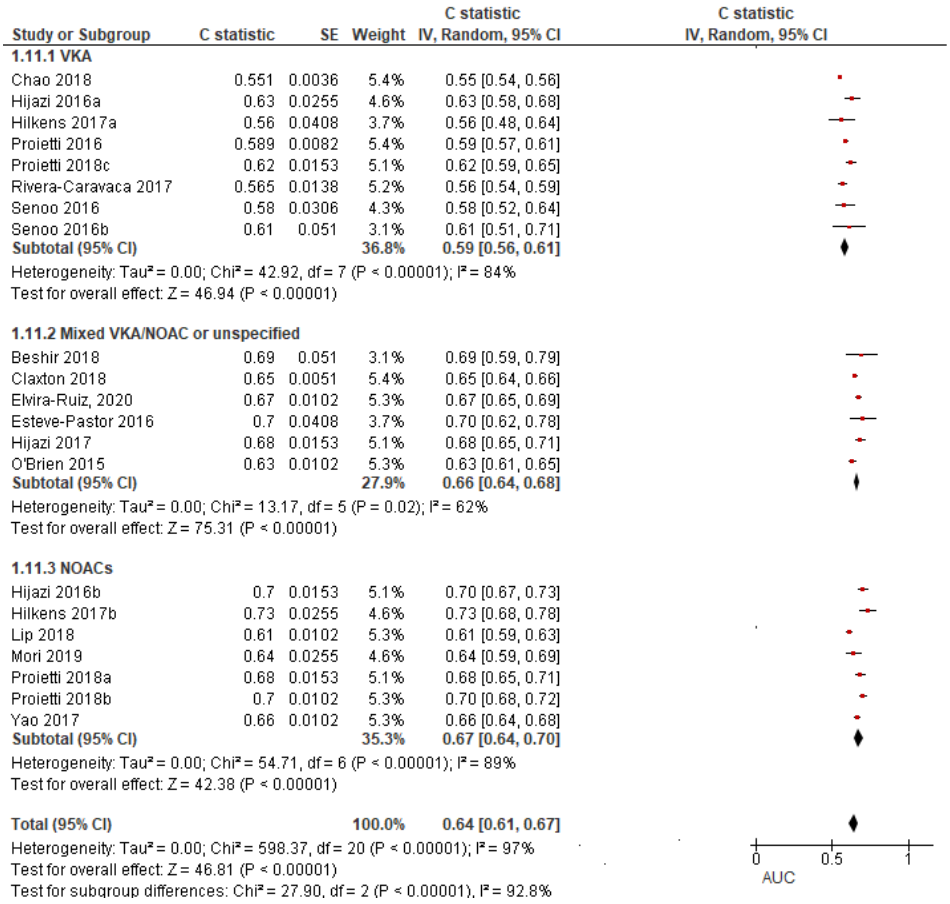


Figure 13: ORBIT (sub-grouped for antiplatelets)

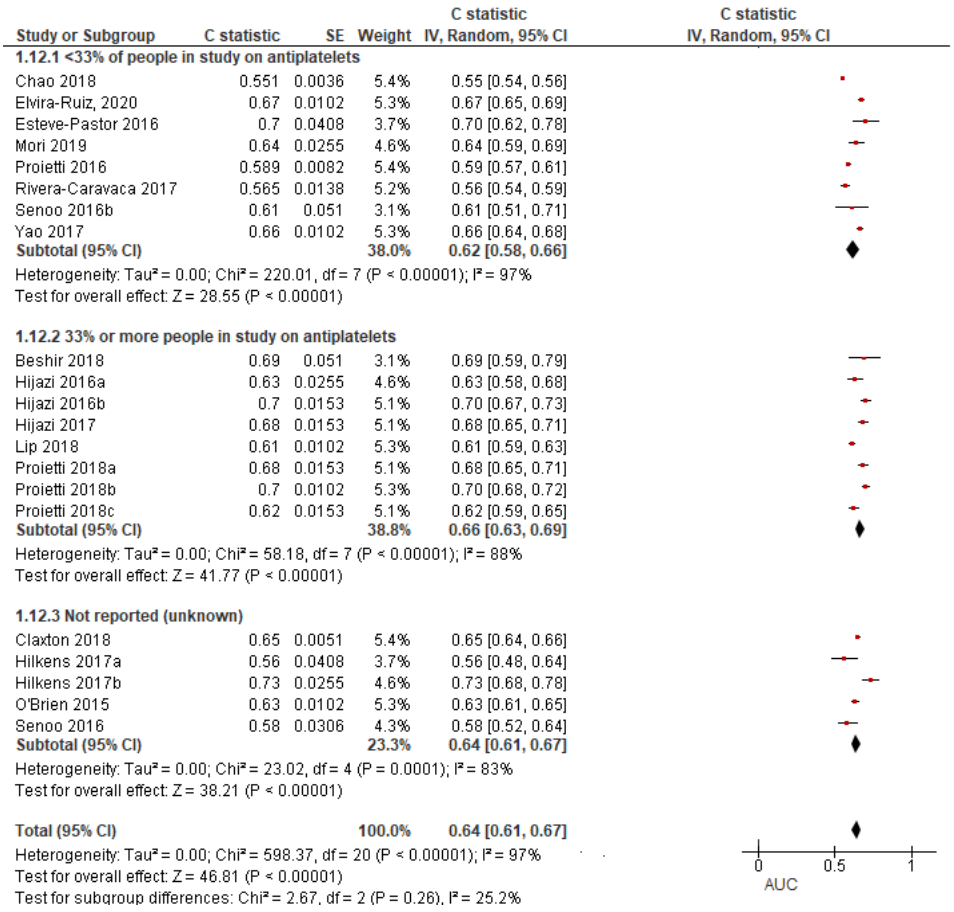


Figure 14: CHADS2(sub-grouped for OAC type)

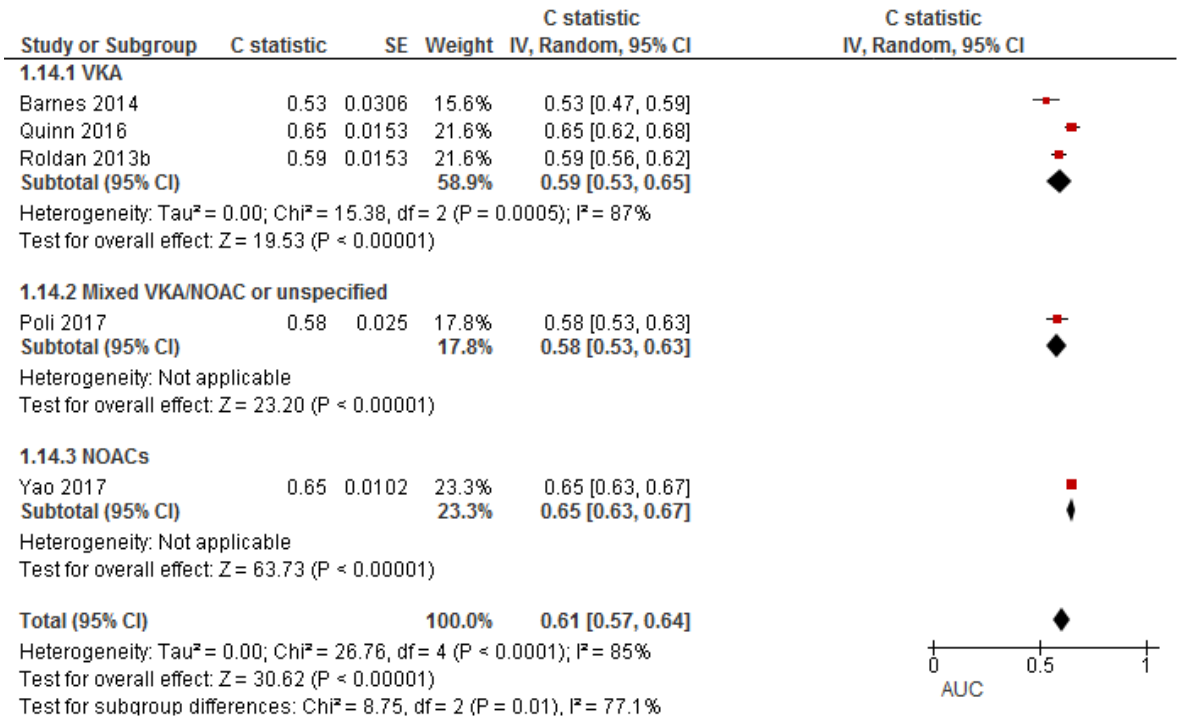


Figure 15: CHADS2(sub-grouped for antiplatelets)

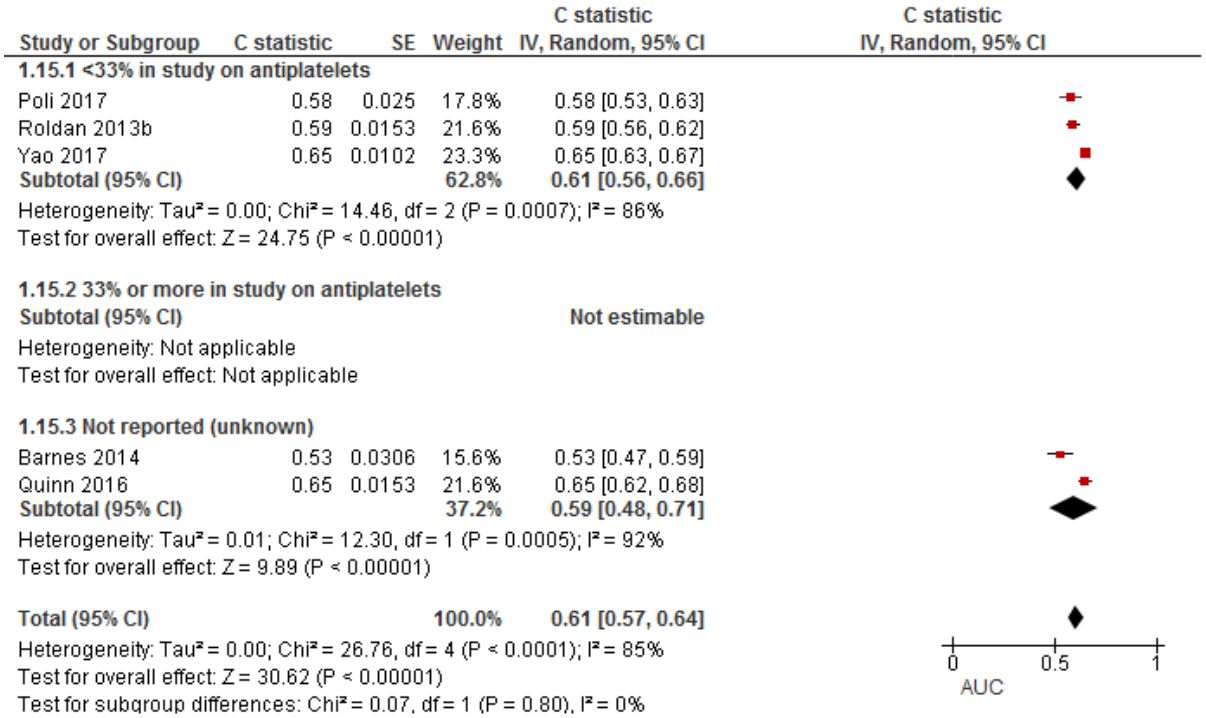


Figure 16: CHADSVASC (sub-grouped for OAC type)

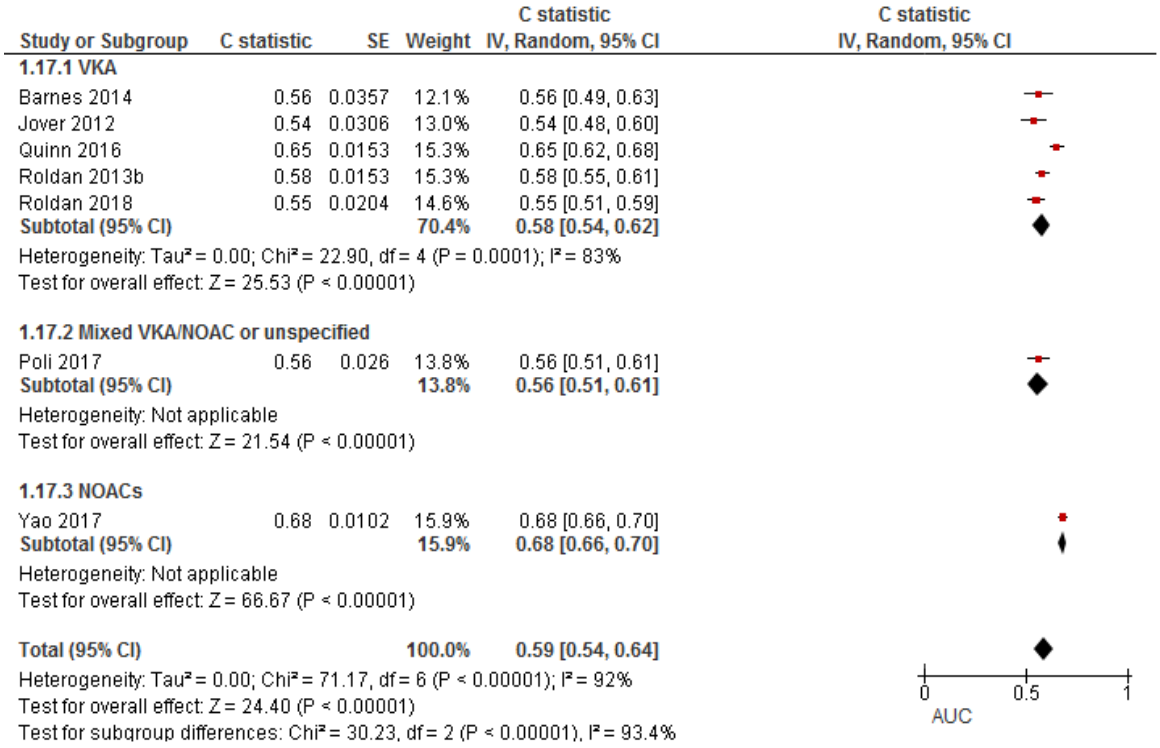


Figure 17: CHADSVASC (sub-grouped for antiplatelets)

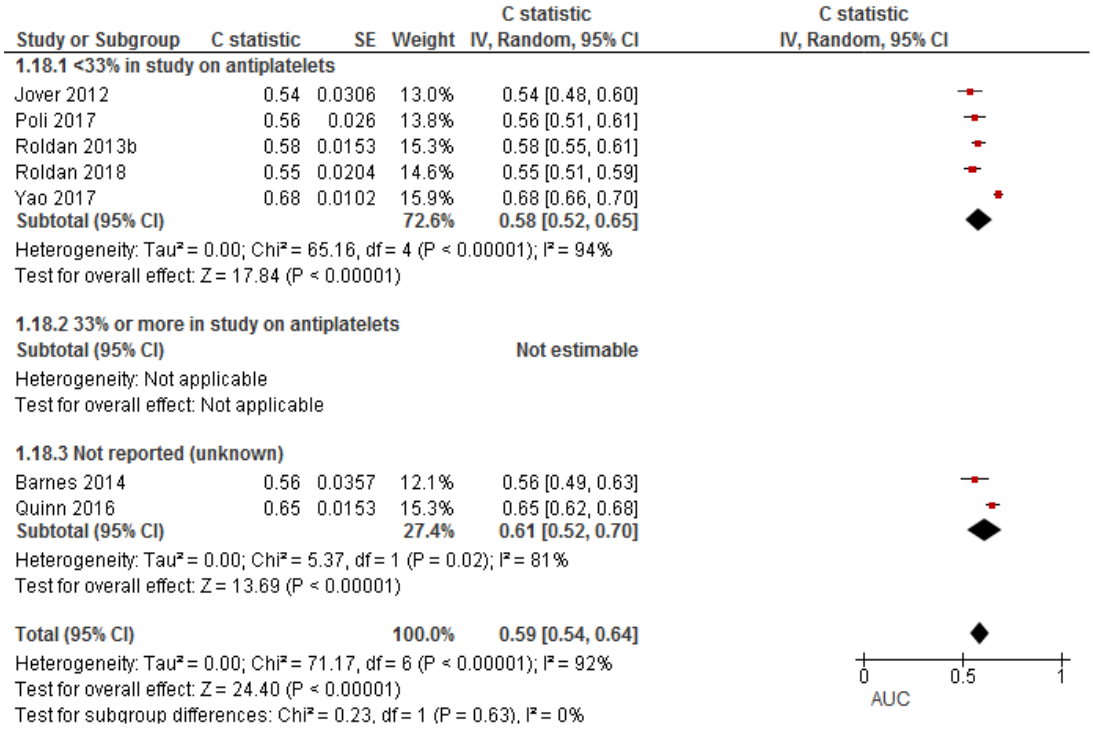


Figure 18: GARFIELD (sub-grouped for OAC type)

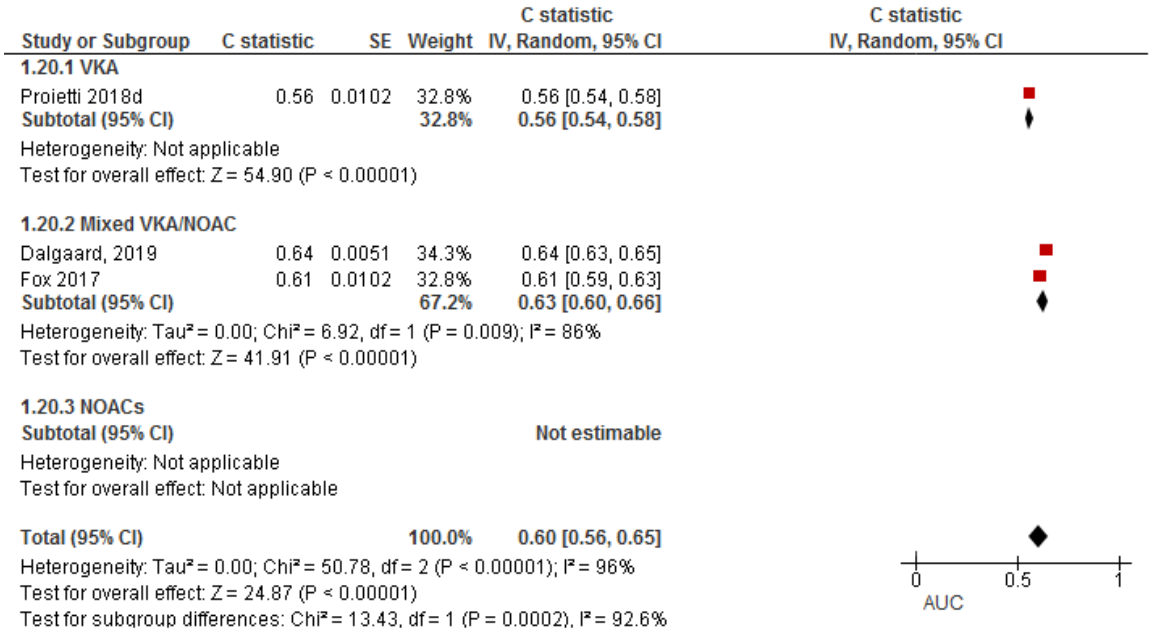


Figure 19: GARFIELD (sub-grouped for antiplatelets)

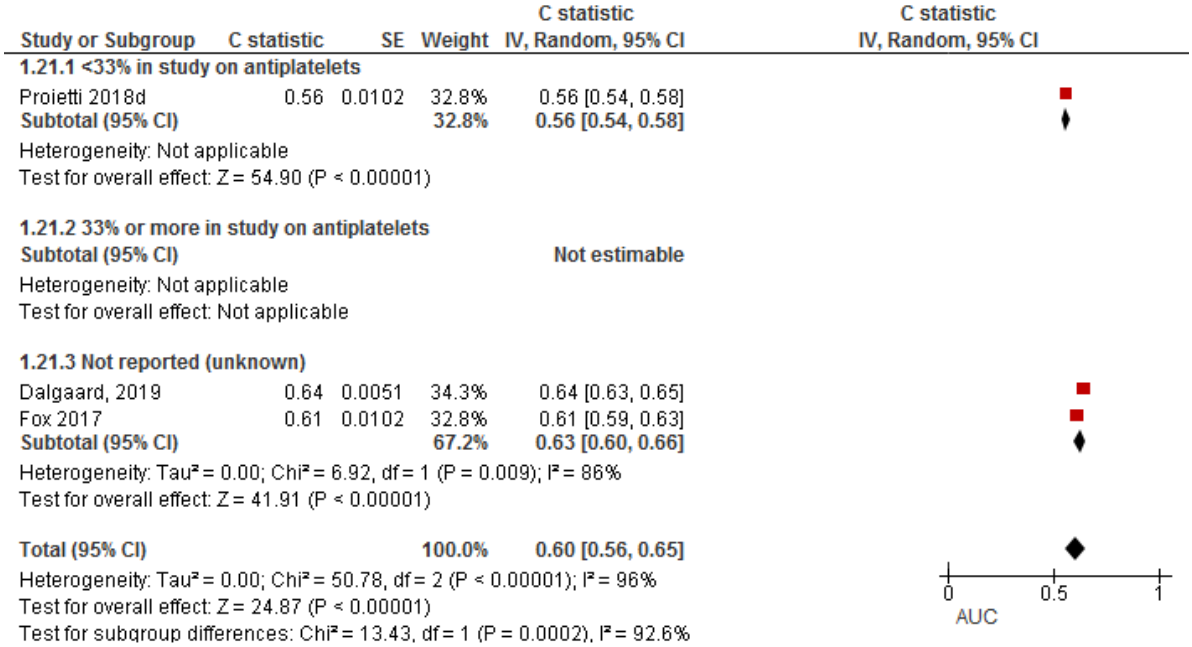


Figure 20: ABC (sub-grouped for OAC type)

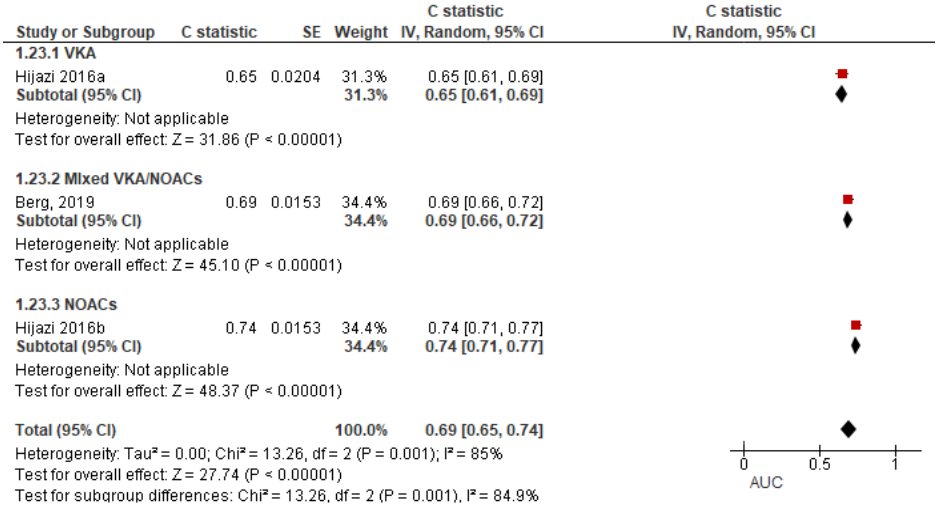


Figure 21: ABC (sub-grouped for antiplatelets)

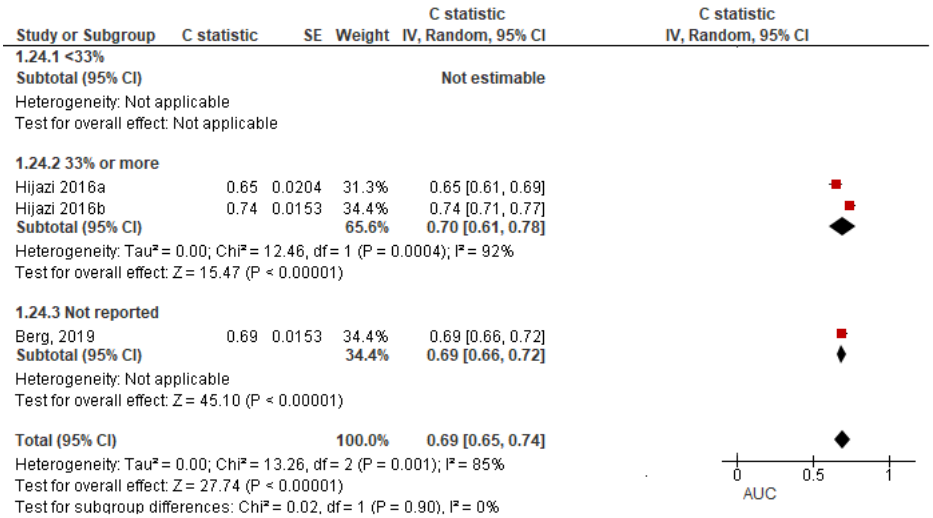


Figure 22: ABC cTnI-hs (sub-grouped for OAC; no sub-grouping for antiplatelets as both studies reporting same (>33%) antiplatelet status)

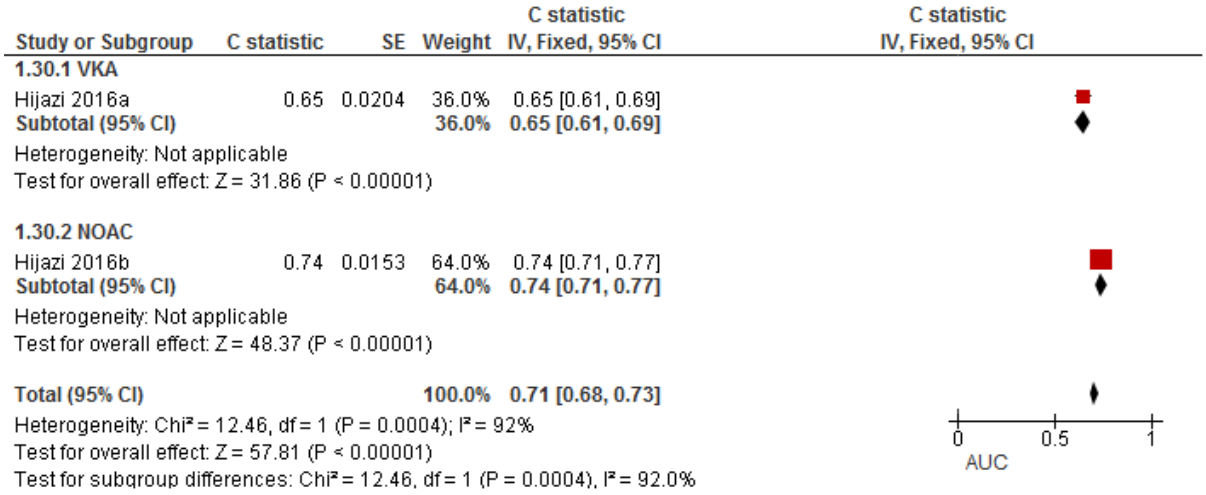


Figure 23: ABC cystatin c (sub-grouped for OAC; no sub-grouping for antiplatelets as both studies reporting same (>33%) antiplatelet status)

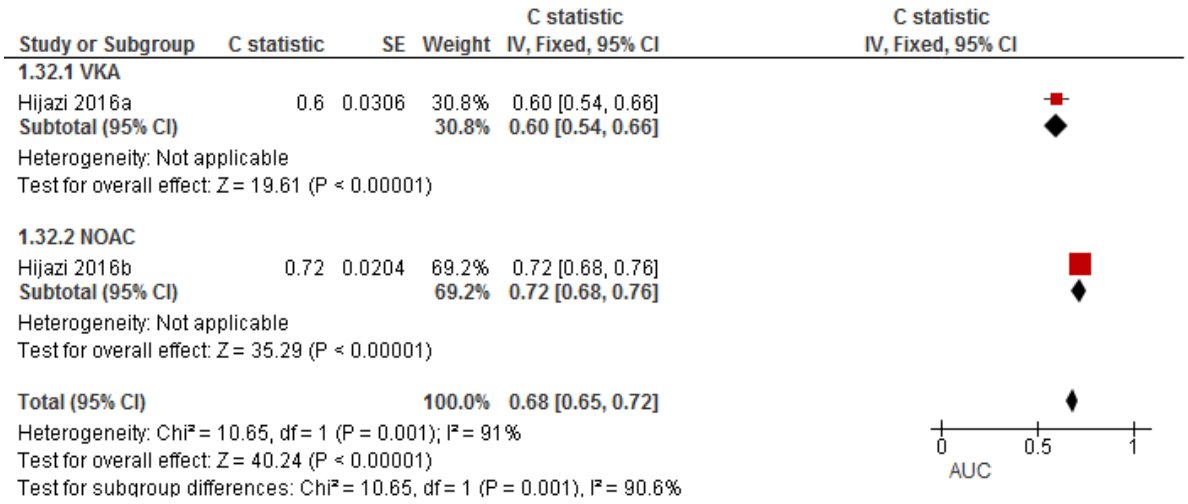


Figure 24: ABC CKD-EPI (sub-grouped for OAC; no sub-grouping for antiplatelets as both studies reporting same (>33%) antiplatelet status)

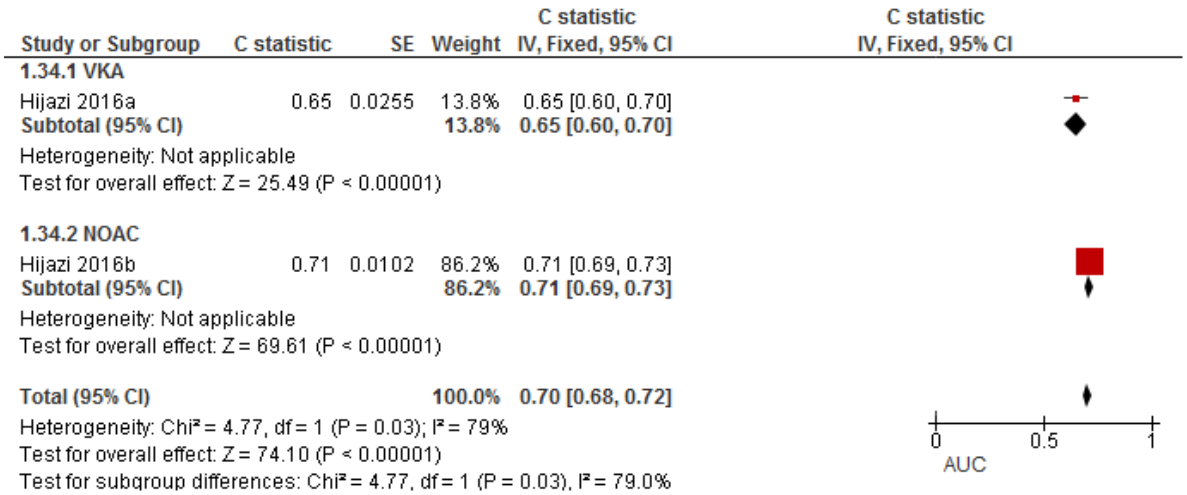


Figure 25: Kuijer (no sub-grouping as both studies involving Warfarin and not reporting antiplatelet status)

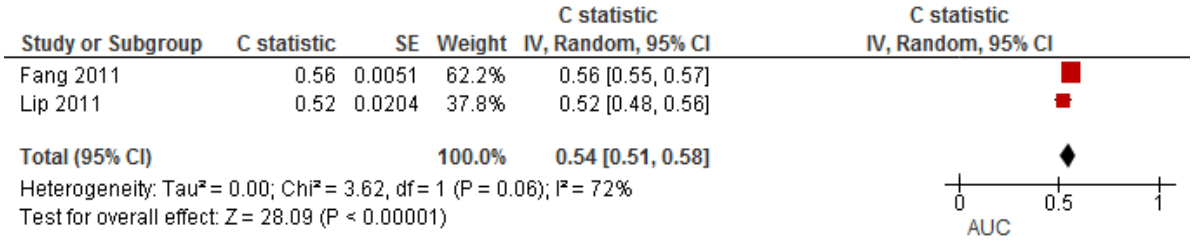


Figure 26: Shireman (sub-grouped for OAC)

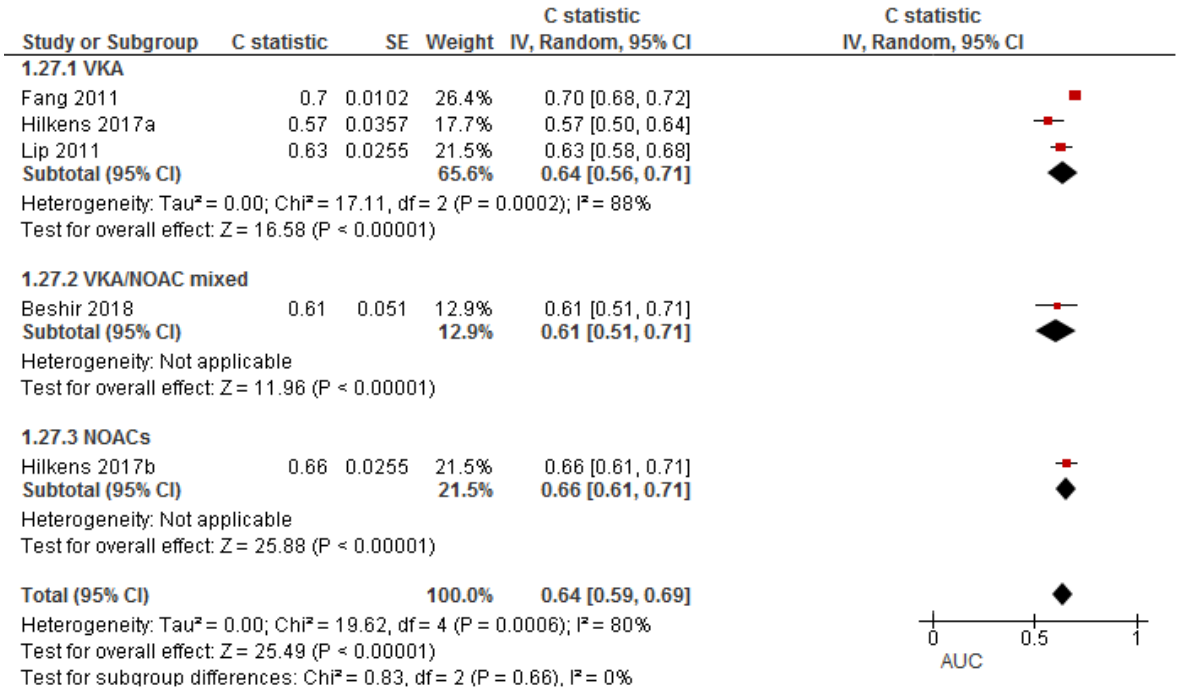


Figure 27: Shireman (sub-grouped for antiplatelets)

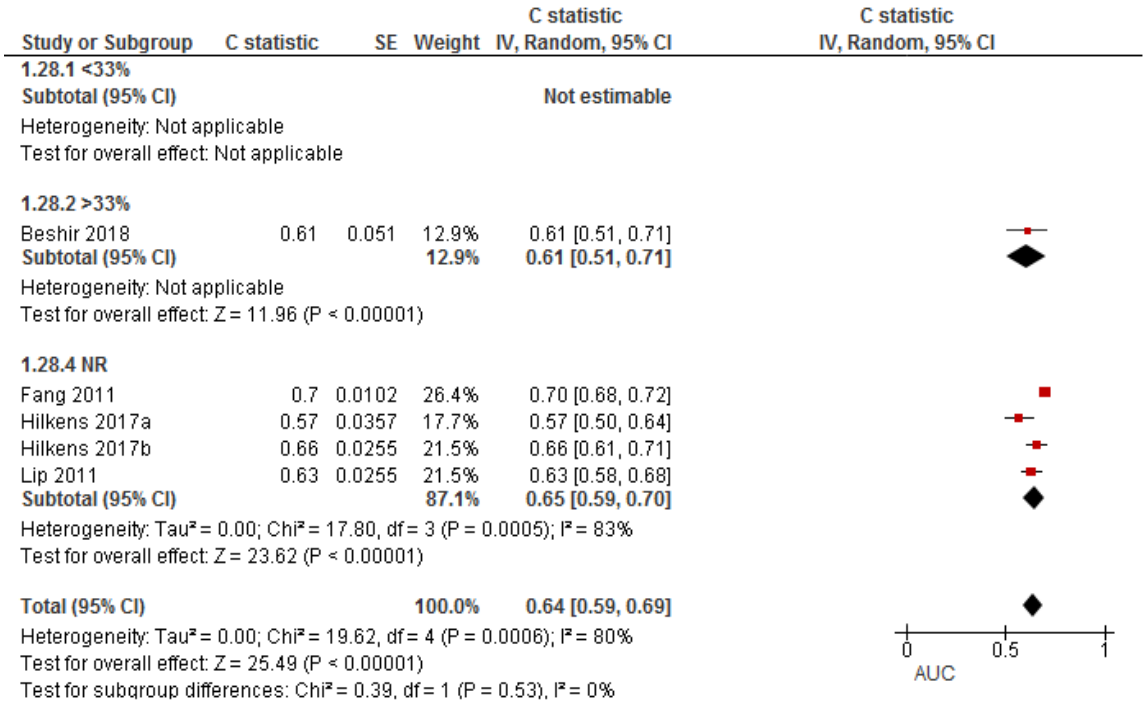
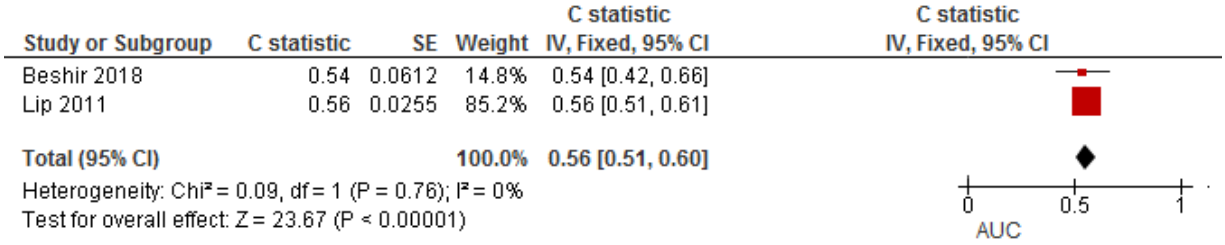


Figure 28: mOBRI (not sub-grouped as no serious heterogeneity)



C statistics for CLINICALLY RELEVANT BLEEDING

Figure 29: HAS-BLED (sub-grouped for OAC type)

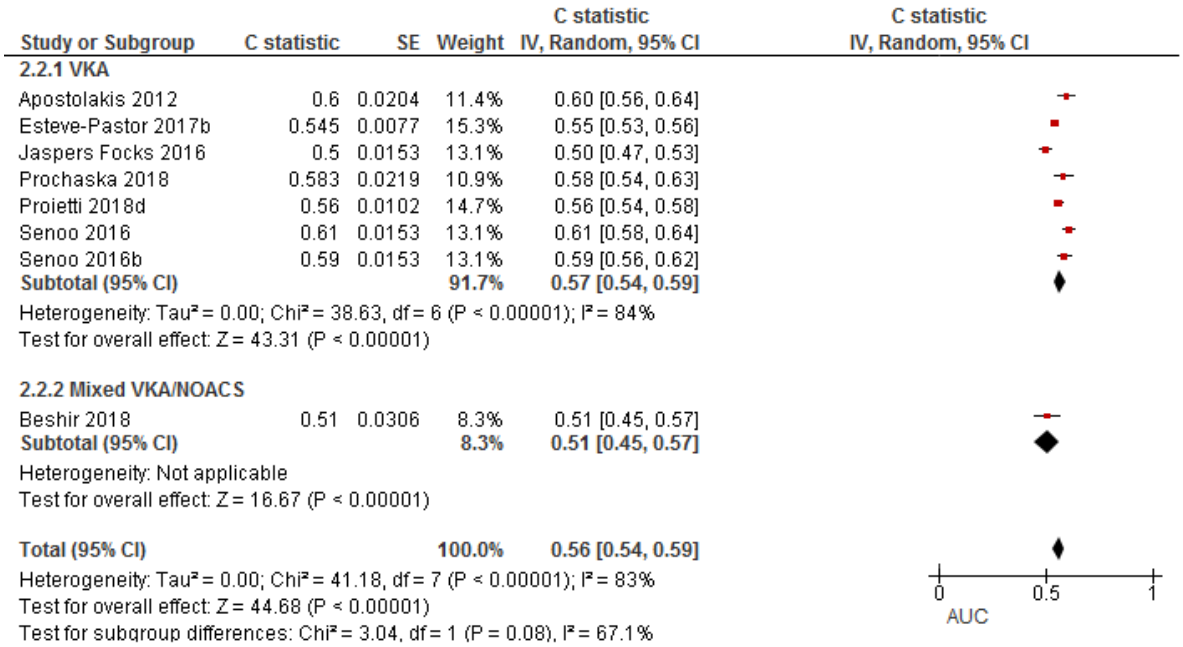


Figure 30: HAS-BLED (sub-grouped for antiplatelets)

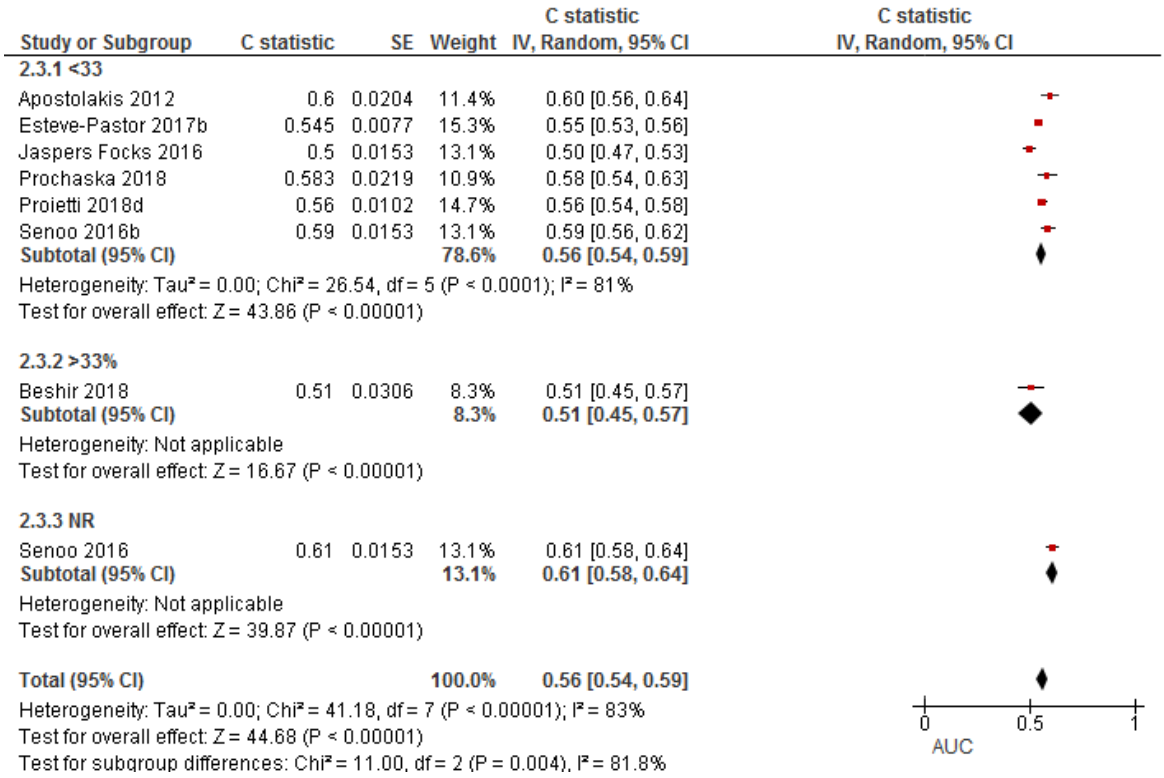


Figure 31: HEMORRHAGES (sub-grouped for OAC type)

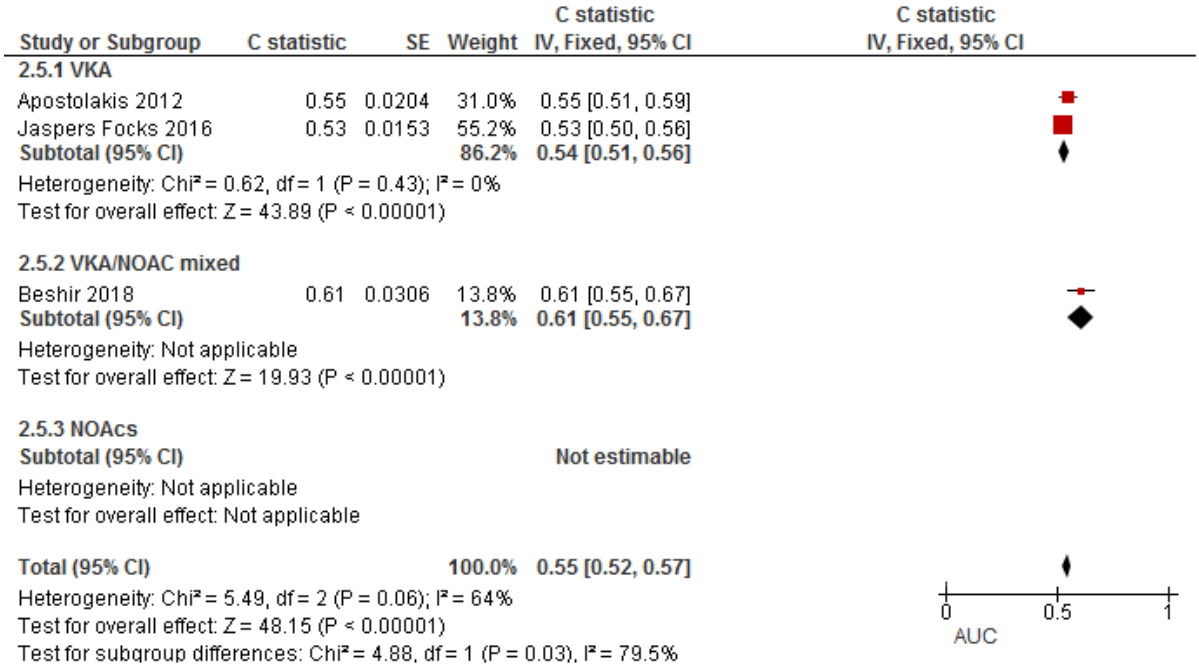


Figure 32: HEMORRHAGES(sub-grouped for antiplatelets)

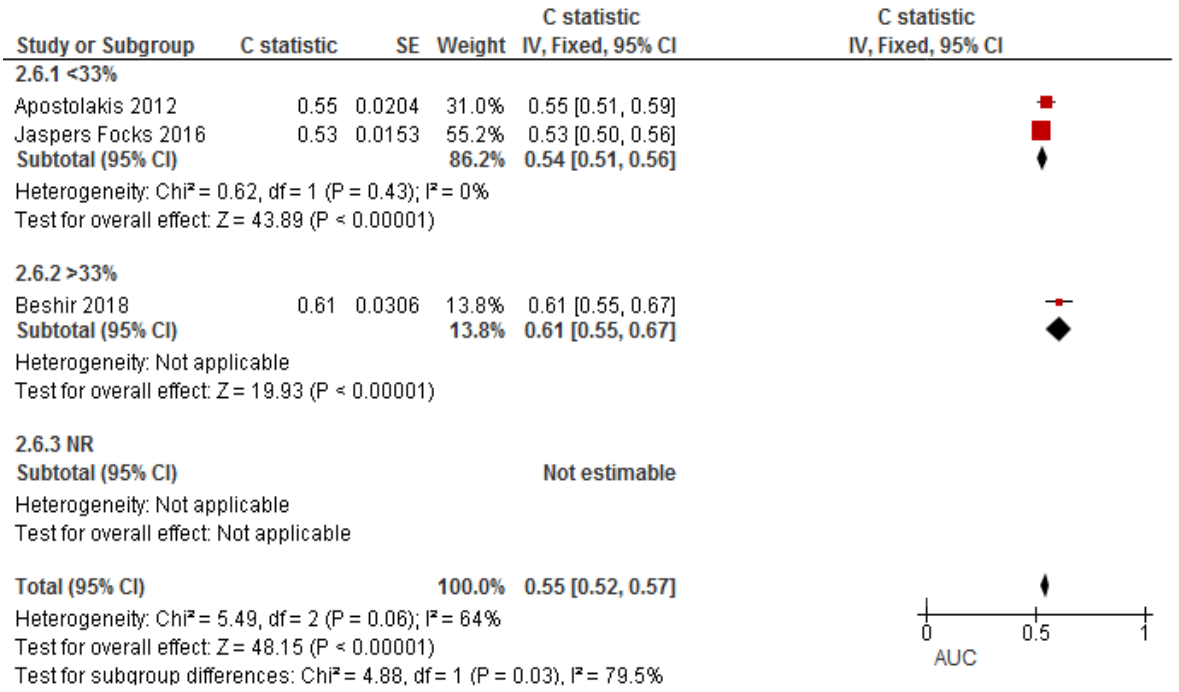


Figure 33: ATRIA(sub-grouped for OAC type)

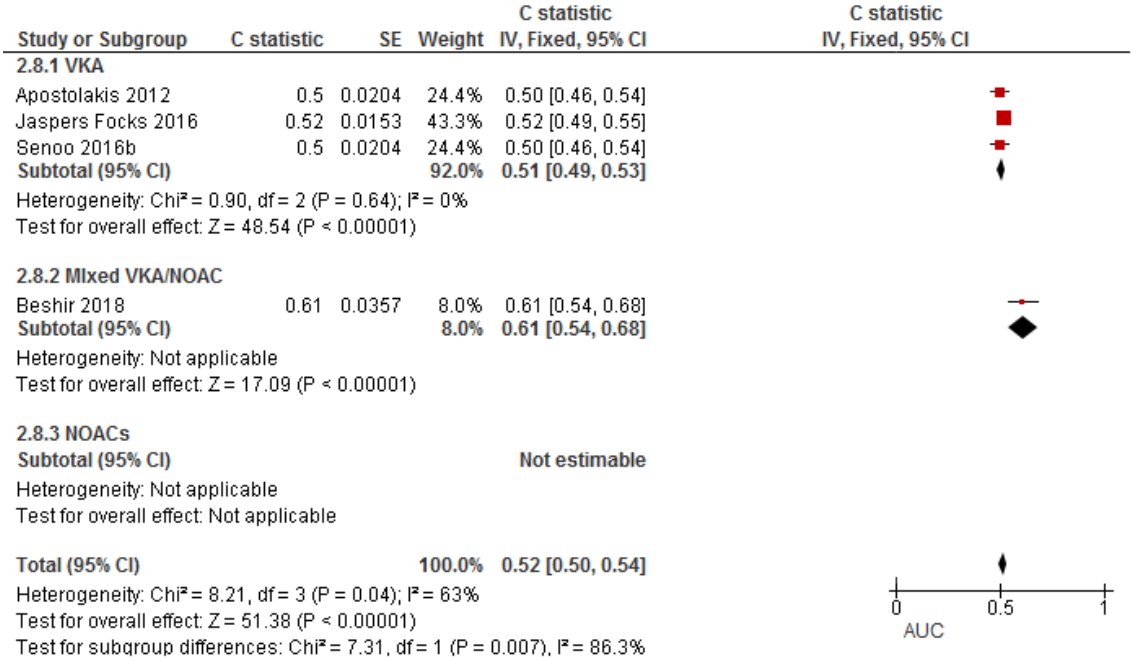


Figure 34: ATRIA(sub-grouped for antiplatelets)

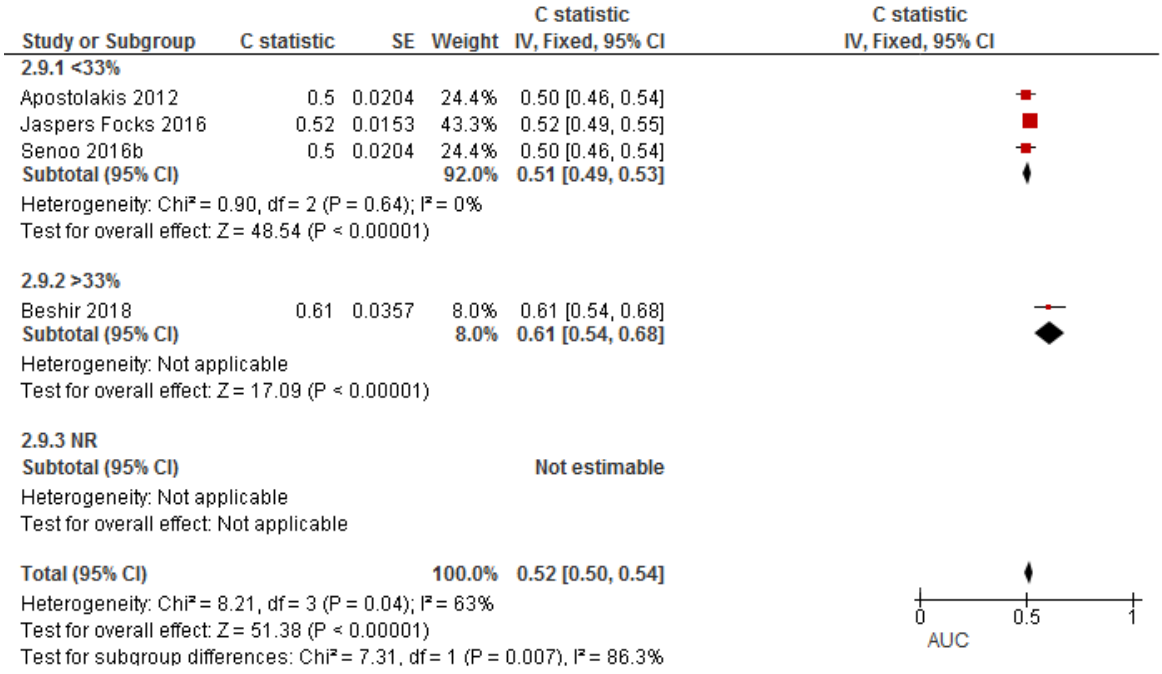


Figure 35: ORBIT (sub-grouped for OAC type)

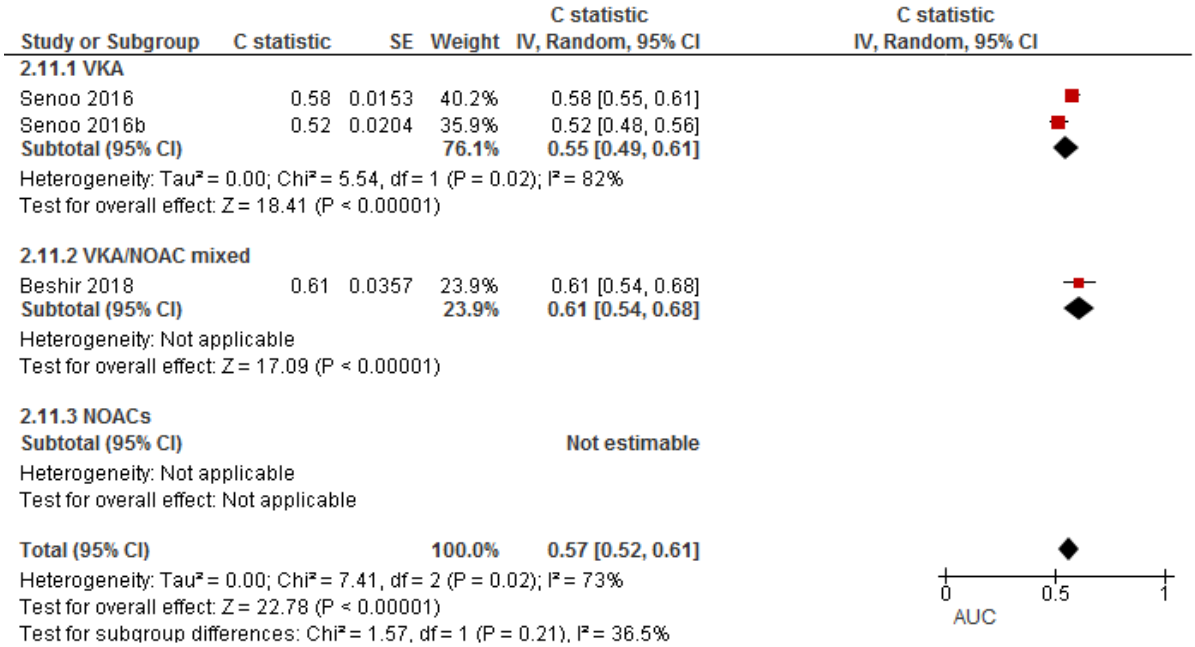
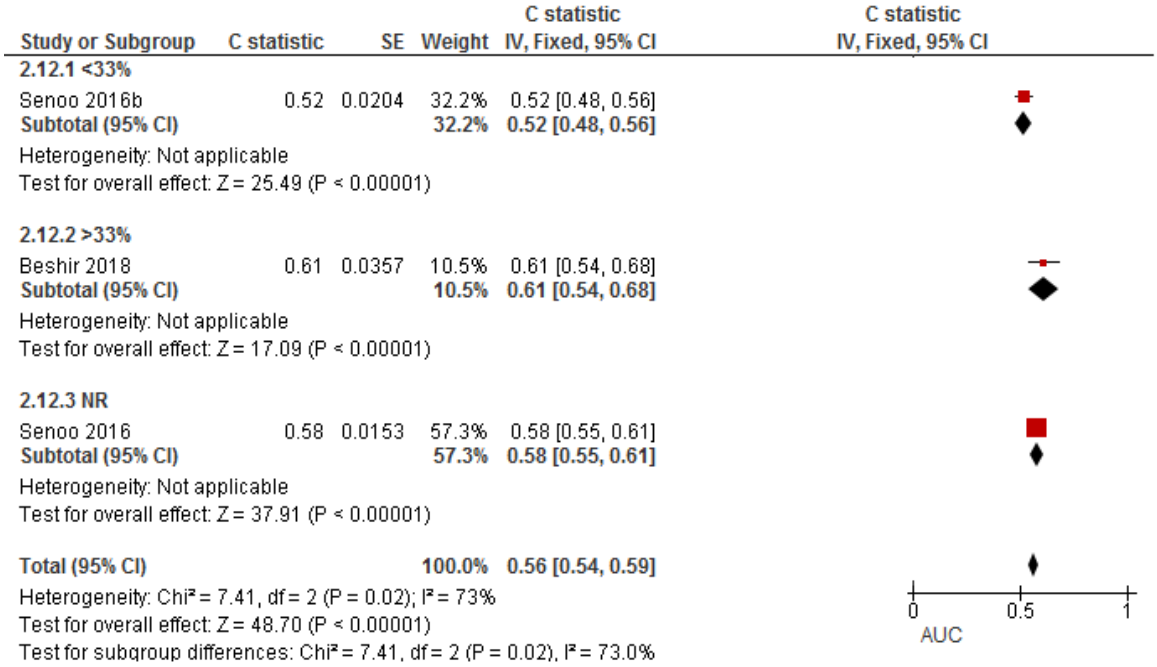


Figure 36: ORBIT (sub-grouped for antiplatelets)



C statistics for INTRACRANIALBLEEDING

Figure 37: HAS-BLED (sub-grouped for OAC type)

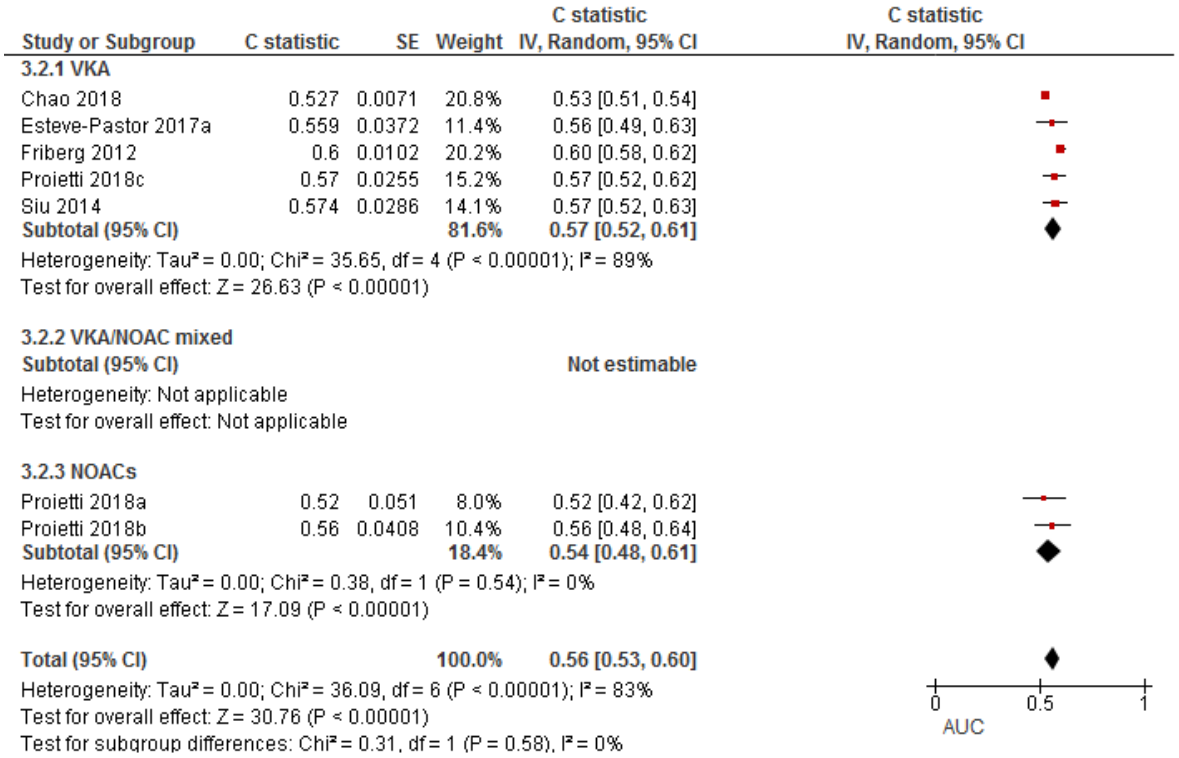


Figure 38: HAS-BLED (sub-grouped for antiplatelets)

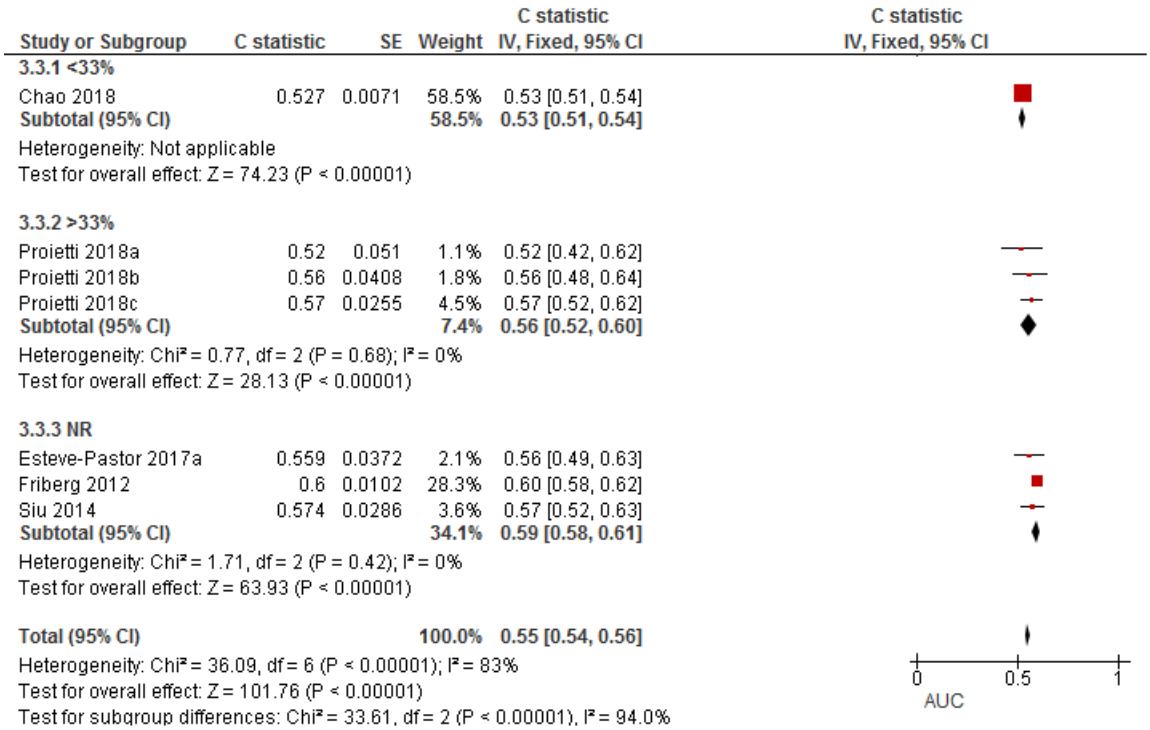


Figure 39: HEMORRHAGES (sub-grouped for OAC type)

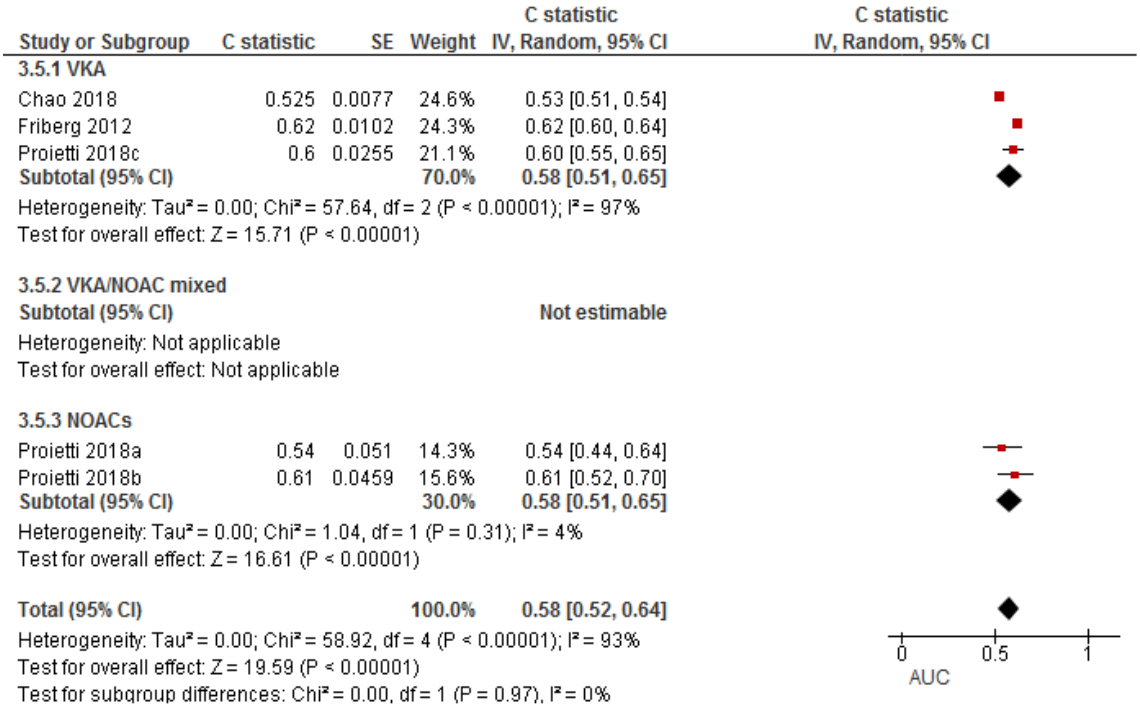


Figure 40: HEMORRHAGES (sub-grouped for antiplatelets)

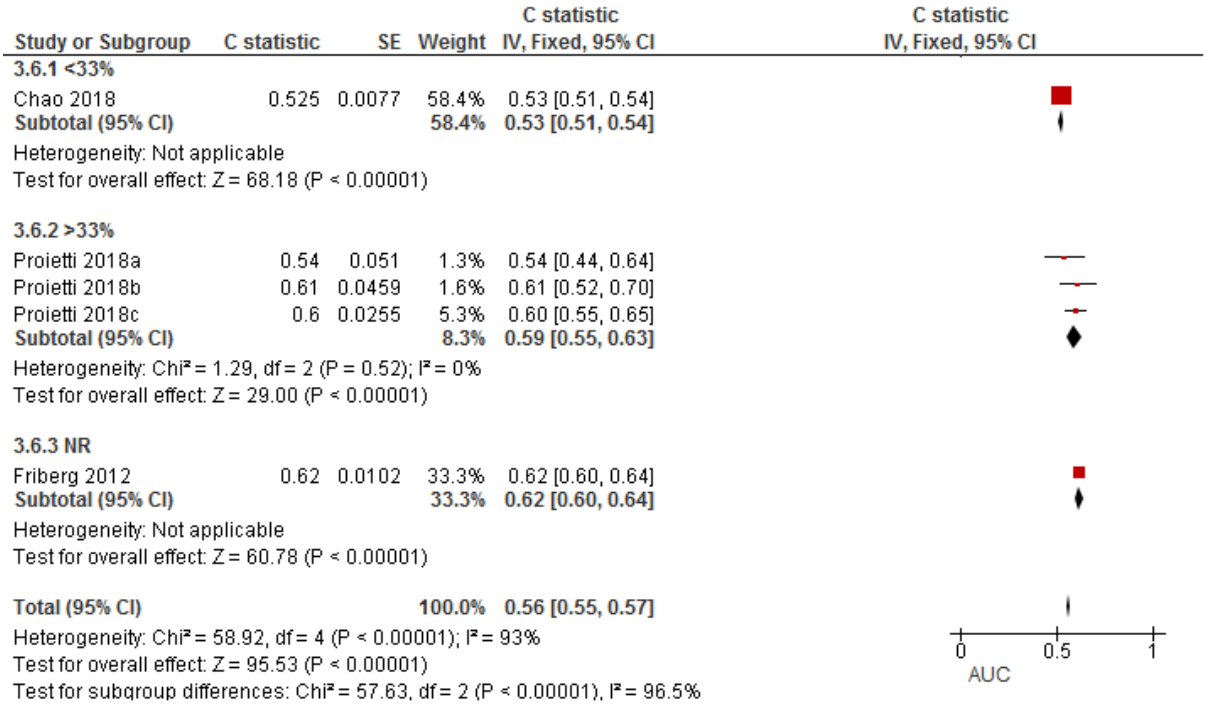


Figure 41: ATRIA (sub-grouped for OAC type)

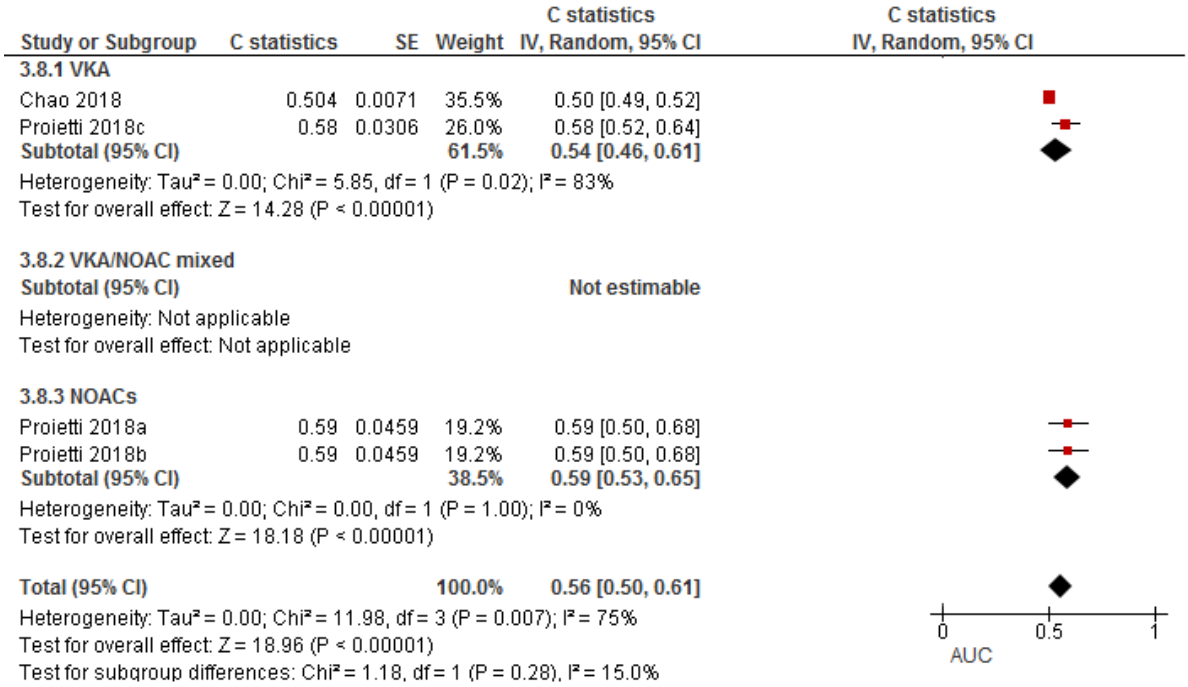


Figure 42: ATRIA (sub-grouped for antiplatelets)

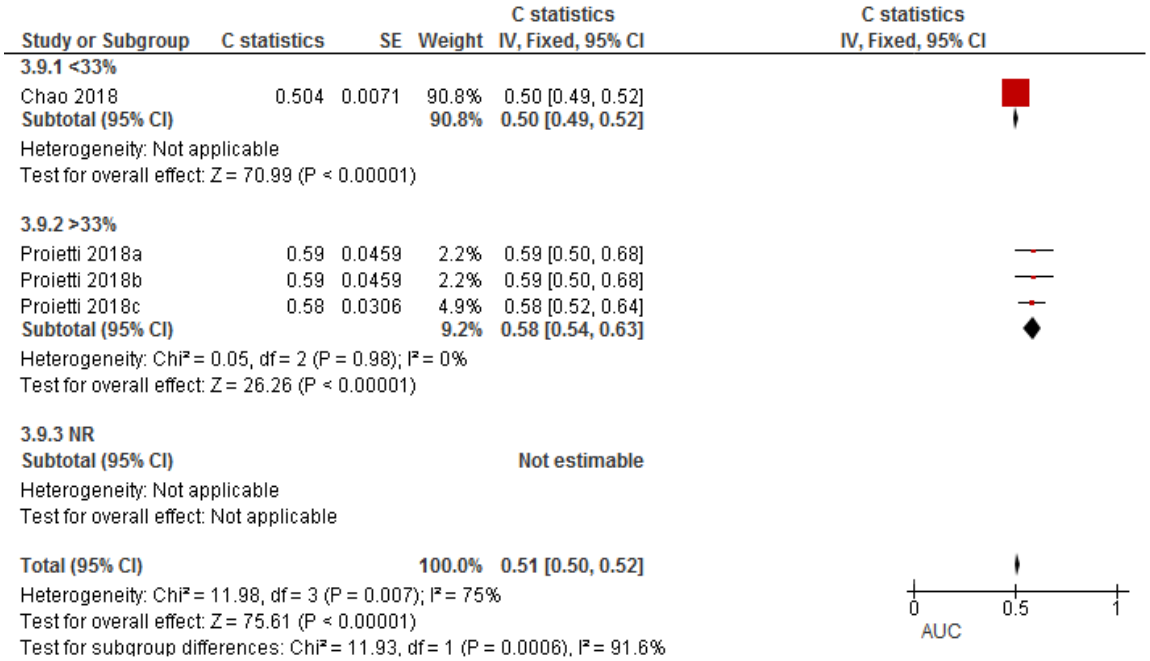


Figure 43: ORBIT (sub-grouped for OAC type)

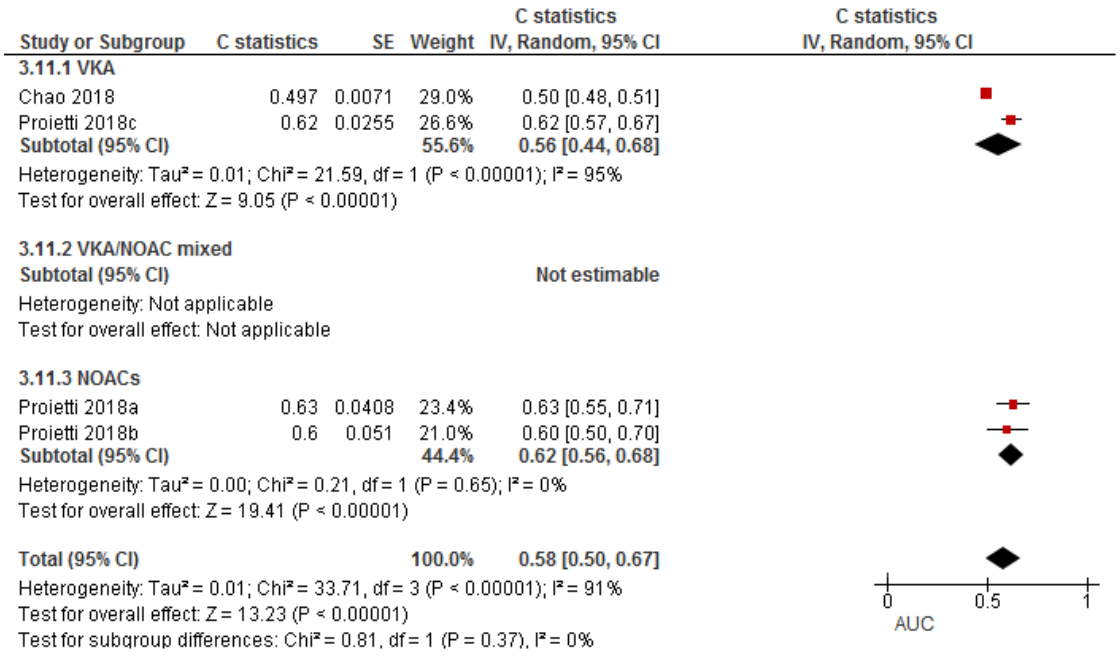
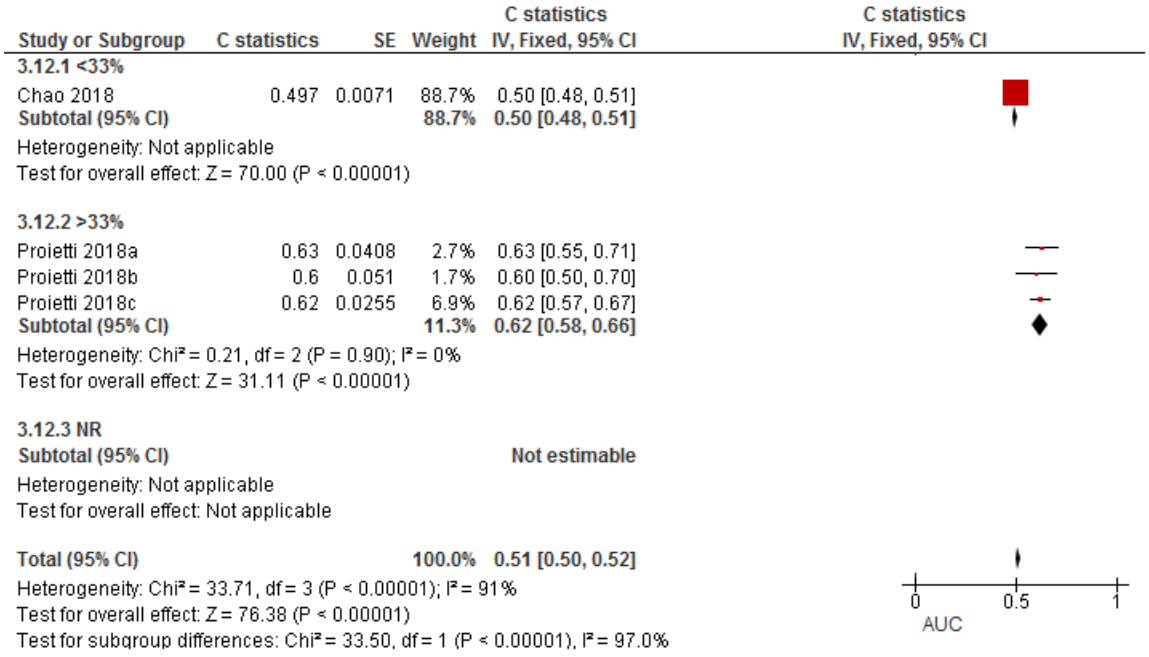


Figure 44: ORBIT (sub-grouped for antiplatelets)



NRI statistics

Note that Forest plots are not shown for comparisons with a single study. Sub-groups are only shown where a sub-group analysis succeeded in reducing heterogeneity to I2<50% in all sub-groups.

Major bleeding

Figure 45: HASBLED v HEMORRHAGE

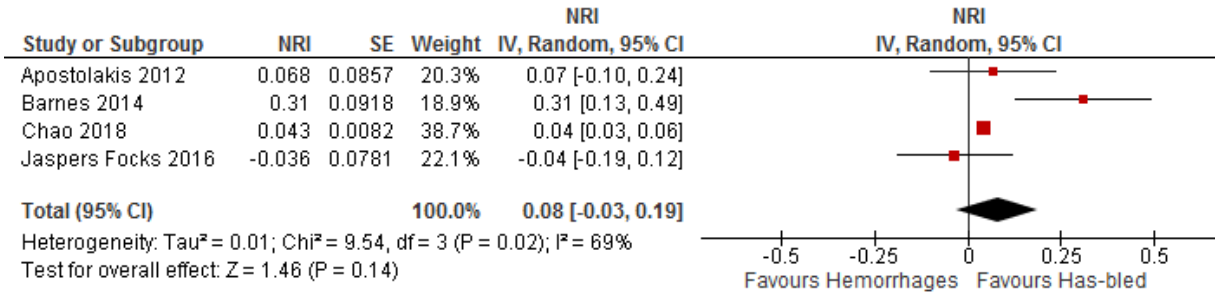


Figure 46: HASBLED v ATRIA

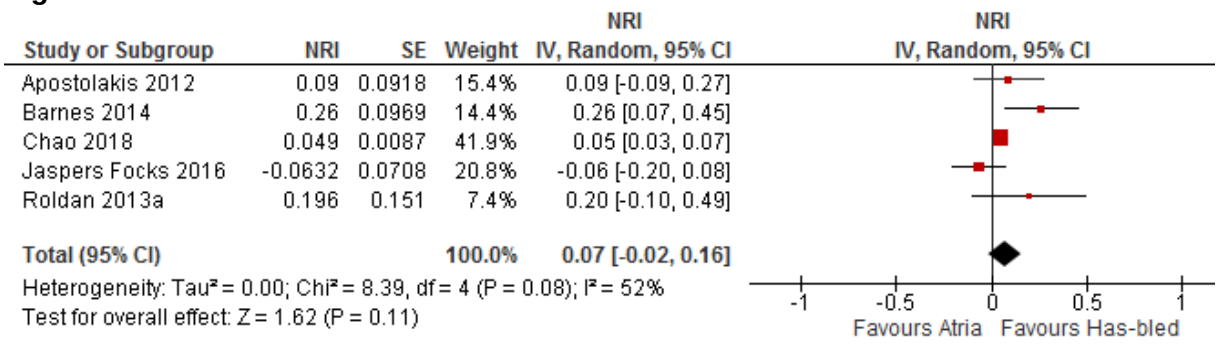


Figure 47: HASBLED v CHADS2

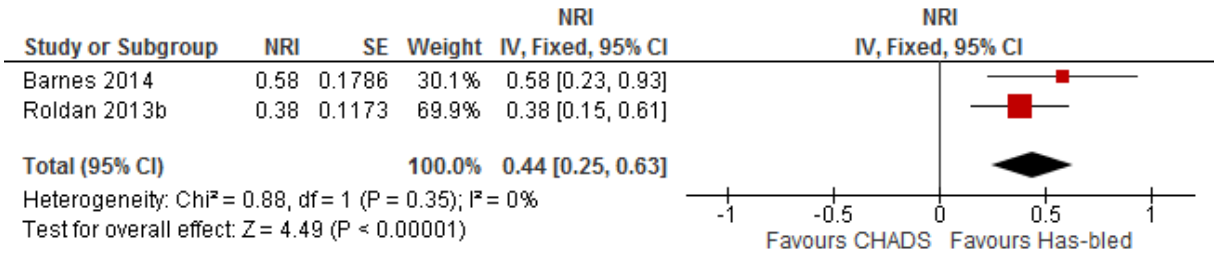


Figure 48: HASBLED v ORBIT

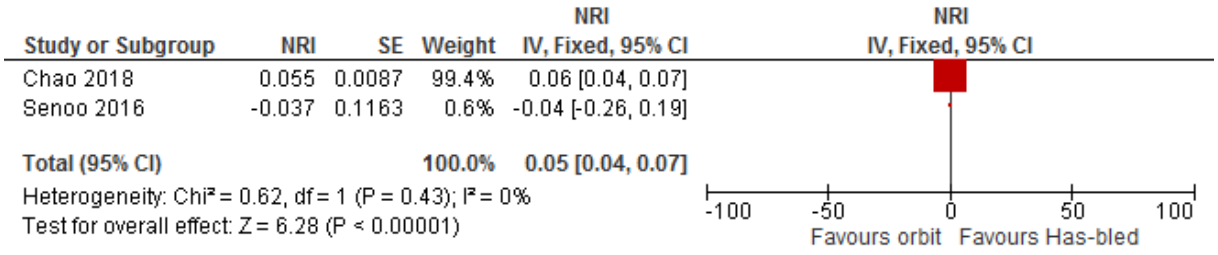


Figure 49: HASBLED v CHADSVASC

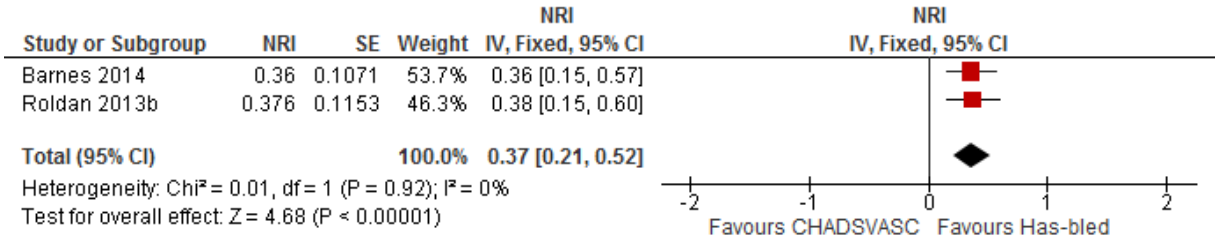


Figure 50: HASBLED v HASBLED with vWF

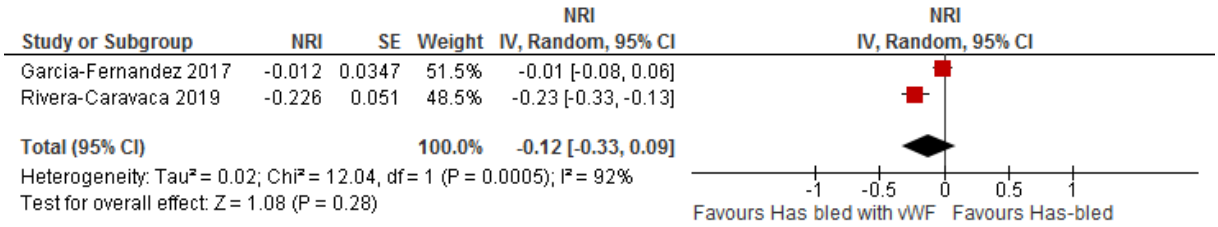


Figure 51: ATRIA v HEMORRHAGES

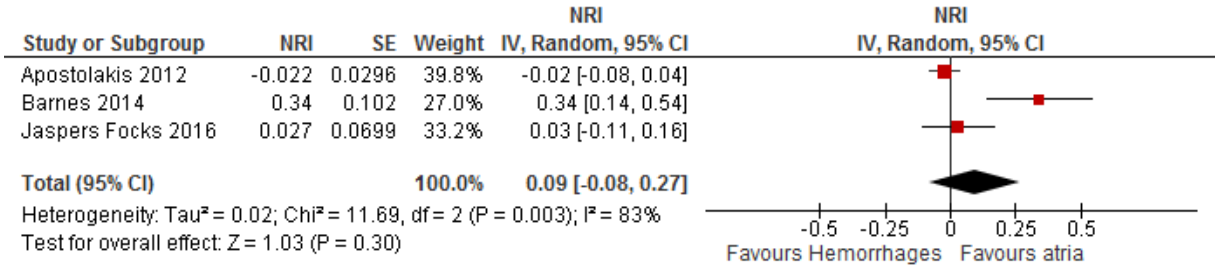


Figure 52: ATRIA v ATRIA with TTR<65%

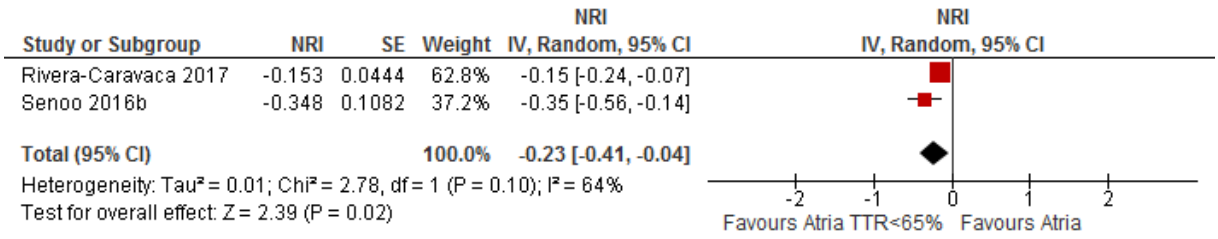
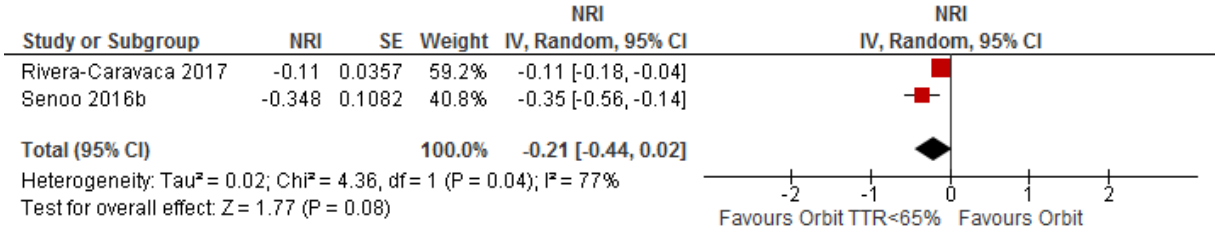


Figure 53: ORBIT v ORBIT with TTR<65%



Clinically relevant bleeding

Figure 54: HASBLED v HEMORRHAGE

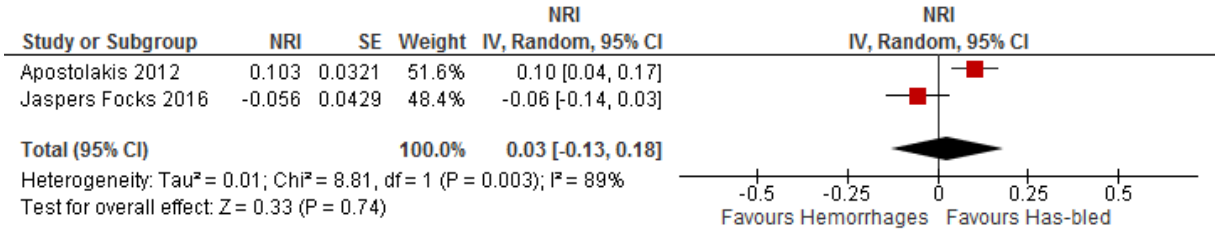


Figure 55: HASBLED v ATRIA

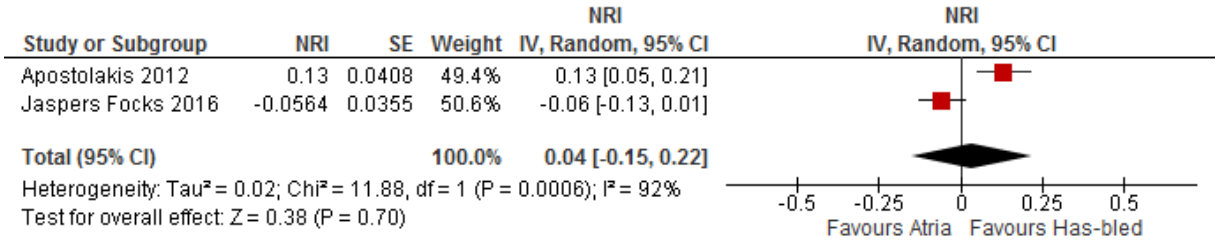
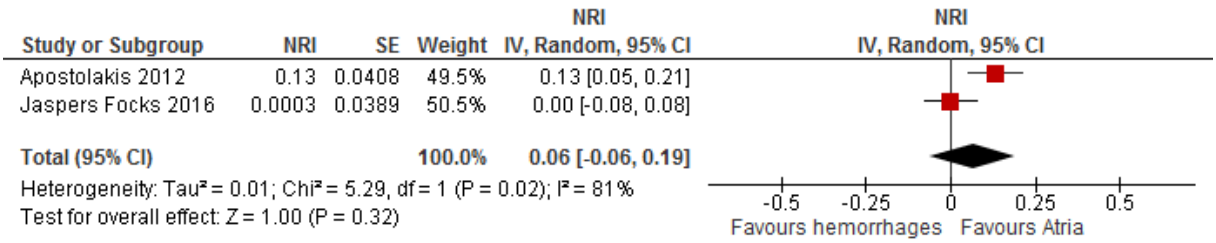


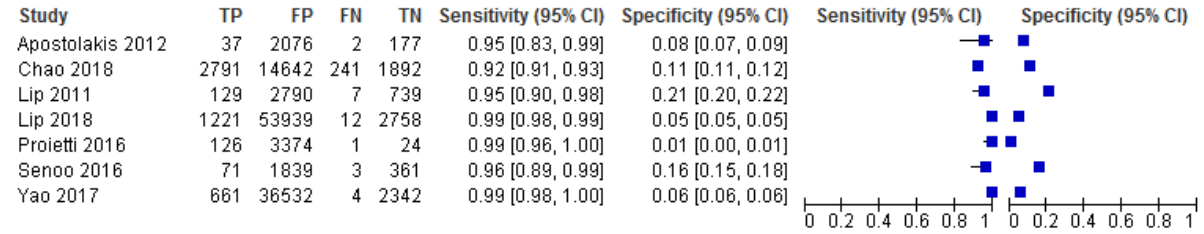
Figure 56: HASBLED v ATRIA

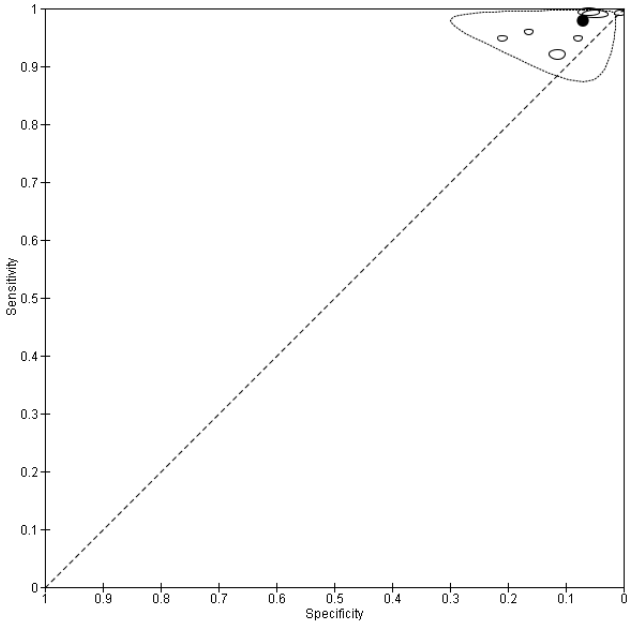


Sensitivity/specificity[only pooled results (n≥3) shown]

Major bleeding

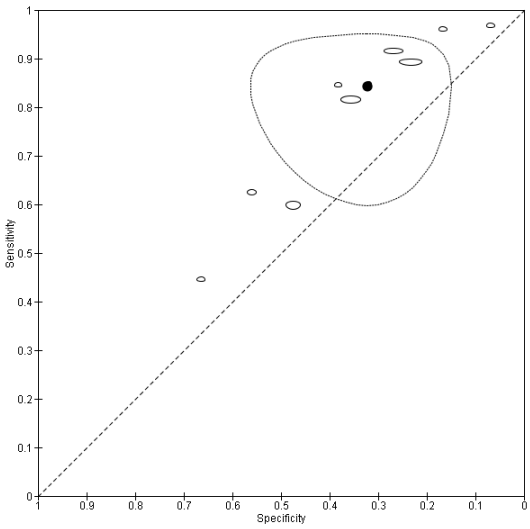
HASBLED at threshold ≥1





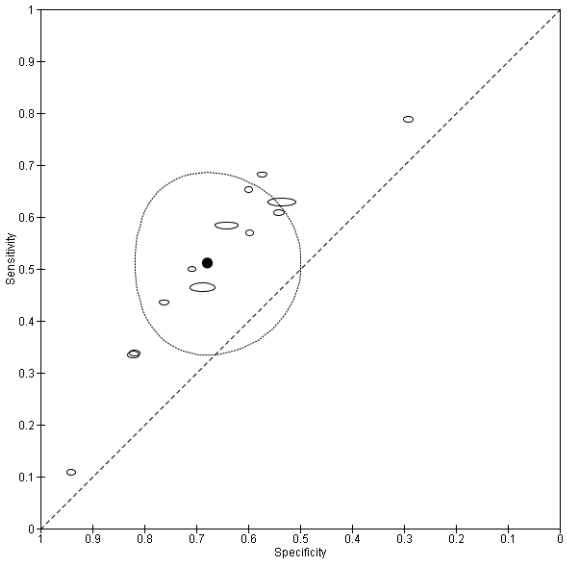
HASBLED at threshold ≥2

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Apostolakis 2012	33	1391	6	862	0.85 [0.69, 0.94]	0.38 [0.36, 0.40]
Chao 2018	1815	8683	1217	7851	0.60 [0.58, 0.62]	0.47 [0.47, 0.48]
Elvira-Ruiz, 2020	172	2180	7	438	0.96 [0.92, 0.98]	0.17 [0.15, 0.18]
Esteve pastor 2016	39	836	7	394	0.85 [0.71, 0.94]	0.32 [0.29, 0.35]
Lip 2011	85	1551	51	1978	0.63 [0.54, 0.71]	0.56 [0.54, 0.58]
Lip, 2018	1110	43493	133	13204	0.89 [0.87, 0.91]	0.23 [0.23, 0.24]
Olesen 2011	1674	27527	377	15193	0.82 [0.80, 0.83]	0.36 [0.35, 0.36]
Proietti 2016	123	3164	4	234	0.97 [0.92, 0.99]	0.07 [0.06, 0.08]
Senoo 2016	33	738	41	1462	0.45 [0.33, 0.57]	0.66 [0.64, 0.68]
Yao 2017	609	28427	56	10447	0.92 [0.89, 0.94]	0.27 [0.26, 0.27]



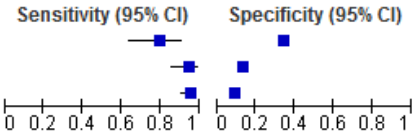
HASBLED at threshold ≥3

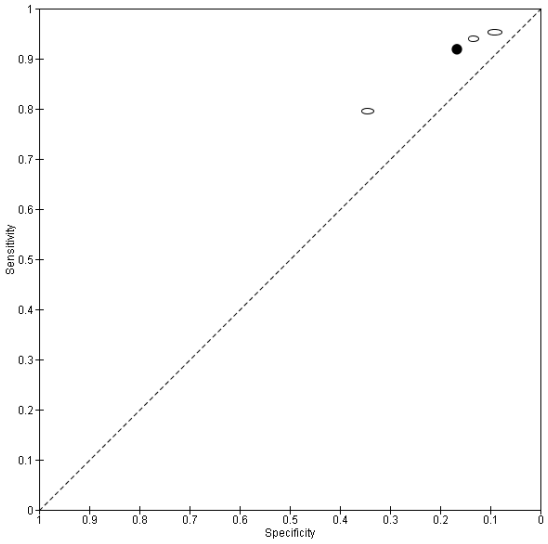
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Apostolakis 2012	17	536	22	1717	0.44 [0.28, 0.60]	0.76 [0.74, 0.78]
Elvira-Ruiz, 2020	122	1119	57	1499	0.68 [0.61, 0.75]	0.57 [0.55, 0.59]
Esteve pastor 2016	23	359	23	871	0.50 [0.35, 0.65]	0.71 [0.68, 0.73]
Esteve Pastor 2017	118	368	89	545	0.57 [0.50, 0.64]	0.60 [0.56, 0.63]
Lip 2011	46	640	90	2889	0.34 [0.26, 0.42]	0.82 [0.81, 0.83]
Lip, 2018	776	26320	457	30377	0.63 [0.60, 0.66]	0.54 [0.53, 0.54]
Olesen 2011	953	13315	1098	29405	0.46 [0.44, 0.49]	0.69 [0.68, 0.69]
Poli 2017	70	1549	45	1824	0.61 [0.51, 0.70]	0.54 [0.52, 0.56]
Proietti 2016	100	2406	27	992	0.79 [0.71, 0.85]	0.29 [0.28, 0.31]
Proietti, 2018	137	1010	272	4657	0.33 [0.29, 0.38]	0.82 [0.81, 0.83]
Rivera Caravaca 2017	163	446	87	665	0.65 [0.59, 0.71]	0.60 [0.57, 0.63]
Senoo 2016	8	130	66	2070	0.11 [0.05, 0.20]	0.94 [0.93, 0.95]
Yao 2017	388	13926	277	24948	0.58 [0.54, 0.62]	0.64 [0.64, 0.65]



Haemmorrhagesat threshold ≥1

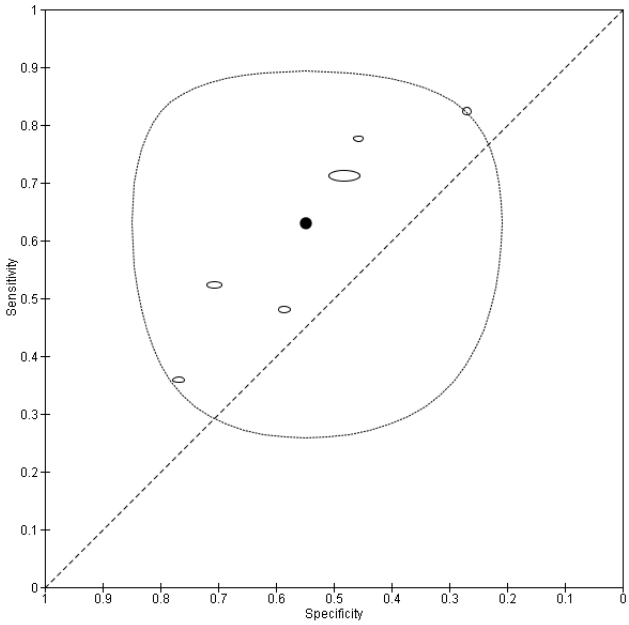
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Apostolakis 2012	31	1460	8	769	0.79 [0.64, 0.91]	0.34 [0.33, 0.37]
Gage 2006	63	1332	4	205	0.94 [0.85, 0.98]	0.13 [0.12, 0.15]
Proietti 2016	121	3097	6	310	0.95 [0.90, 0.98]	0.09 [0.08, 0.10]





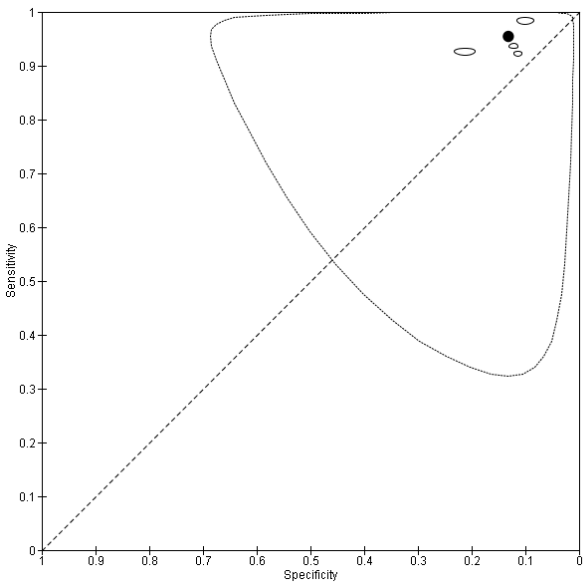
Haemmorrhagesat threshold ≥2

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Apostolakis 2012	14	516	25	1713	0.36 [0.21, 0.53]	0.77 [0.75, 0.79]
Gage 2006	52	835	15	702	0.78 [0.66, 0.87]	0.46 [0.43, 0.48]
Olesen 2011	1459	22127	592	20593	0.71 [0.69, 0.73]	0.48 [0.48, 0.49]
Proietti 2016	61	1413	66	1994	0.48 [0.39, 0.57]	0.59 [0.57, 0.60]
Proietti 2018	214	1785	195	4291	0.52 [0.47, 0.57]	0.71 [0.69, 0.72]
Rivera Caravaca 2017	206	812	44	299	0.82 [0.77, 0.87]	0.27 [0.24, 0.30]



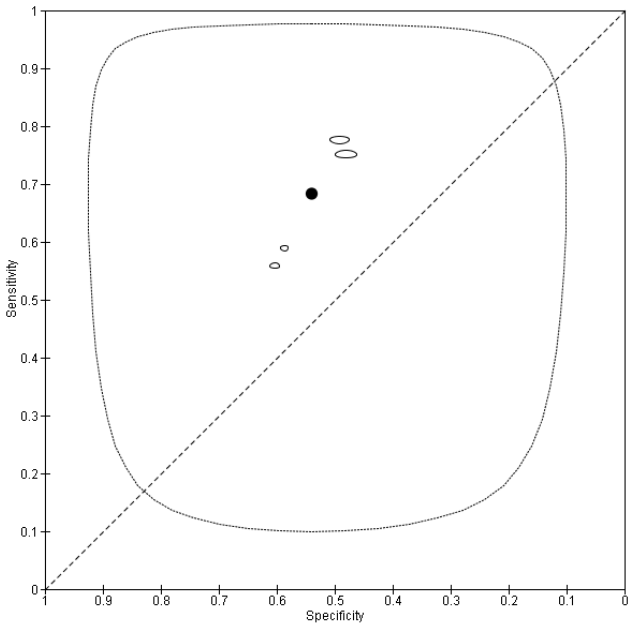
Atria at threshold ≥1

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Apostolakis 2012	36	1974	3	255	0.92 [0.79, 0.98]	0.11 [0.10, 0.13]
Lip 2018	1142	44630	91	12067	0.93 [0.91, 0.94]	0.21 [0.21, 0.22]
Proietti 2016	119	3007	8	418	0.94 [0.88, 0.97]	0.12 [0.11, 0.13]
Yao 2017	654	34986	11	3888	0.98 [0.97, 0.99]	0.10 [0.10, 0.10]



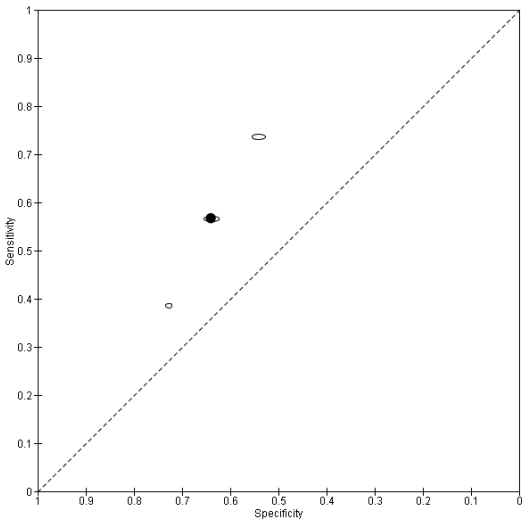
Atria at threshold ≥2

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Apostolakis 2012	23	921	16	1308	0.59 [0.42, 0.74]	0.59 [0.57, 0.61]
Lip 2018	927	29473	306	27224	0.75 [0.73, 0.78]	0.48 [0.48, 0.48]
Proietti 2016	71	1358	56	2067	0.56 [0.47, 0.65]	0.60 [0.59, 0.62]
Yao 2017	516	19774	149	19100	0.78 [0.74, 0.81]	0.49 [0.49, 0.50]



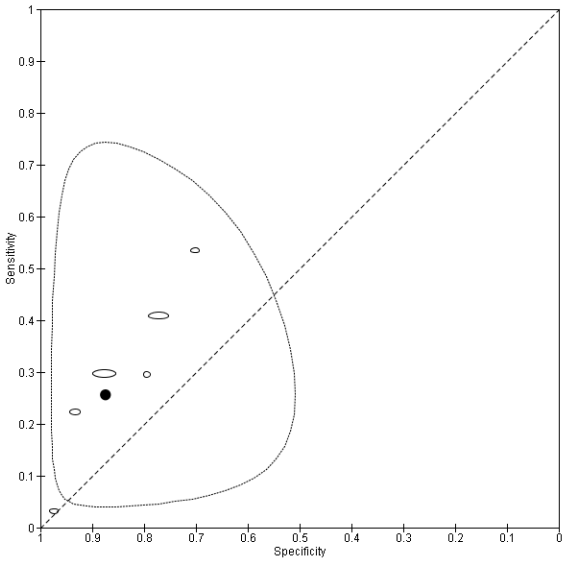
Atria at threshold ≥3

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Lip 2018	698	20494	535	36203	0.57 [0.54, 0.59]	0.64 [0.63, 0.64]
Proietti 2016	49	933	78	2492	0.39 [0.30, 0.48]	0.73 [0.71, 0.74]
Yao 2017	489	17839	176	21037	0.74 [0.70, 0.77]	0.54 [0.54, 0.55]



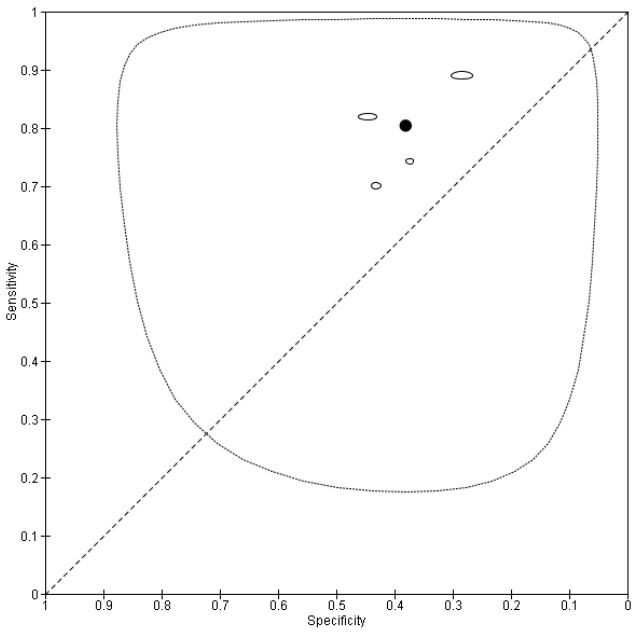
Atria at threshold ≥4

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Elvira-Ruiz, 2020	99	805	86	1889	0.54 [0.46, 0.61]	0.70 [0.68, 0.72]
Lip, 2018	366	7021	867	49676	0.30 [0.27, 0.32]	0.88 [0.87, 0.88]
Proietti 2016	4	90	123	3335	0.03 [0.01, 0.08]	0.97 [0.97, 0.98]
Proietti, 2018	91	380	318	5287	0.22 [0.18, 0.27]	0.93 [0.93, 0.94]
Rivera Caravaca 2017	74	228	176	883	0.30 [0.24, 0.36]	0.79 [0.77, 0.82]
Yao 2017	272	8874	393	30000	0.41 [0.37, 0.45]	0.77 [0.77, 0.78]



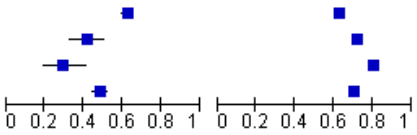
Orbit at threshold ≥1

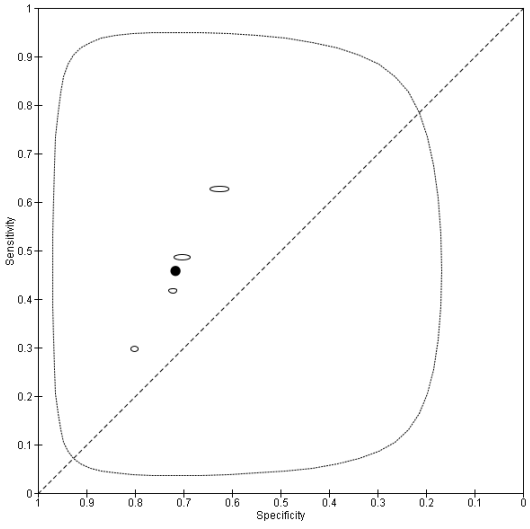
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Lip 2018	1098	40599	135	16098	0.89 [0.87, 0.91]	0.28 [0.28, 0.29]
Proietti 2016	89	1945	38	1478	0.70 [0.61, 0.78]	0.43 [0.42, 0.45]
Senoo 2016	55	1383	19	826	0.74 [0.63, 0.84]	0.37 [0.35, 0.39]
Yao 2017	545	21525	120	17349	0.82 [0.79, 0.85]	0.45 [0.44, 0.45]



Orbit at threshold ≥2

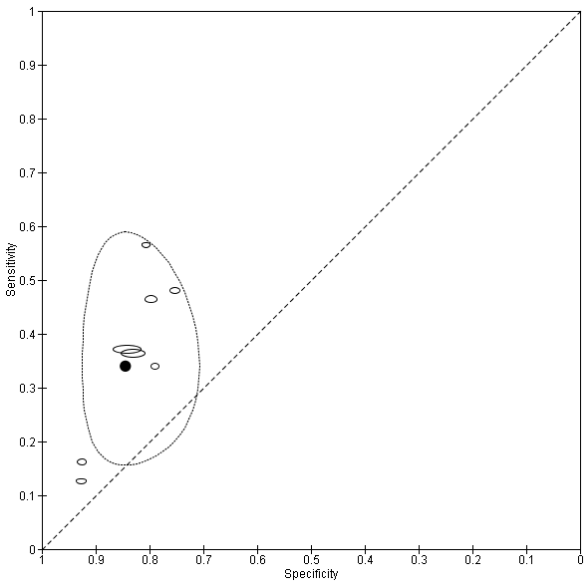
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Lip 2018	773	21223	460	35474	0.63 [0.60, 0.65]	0.63 [0.62, 0.63]
Proietti 2016	53	951	74	2472	0.42 [0.33, 0.51]	0.72 [0.71, 0.74]
Senoo 2016	22	441	52	1768	0.30 [0.20, 0.41]	0.80 [0.78, 0.82]
Yao 2017	323	11565	342	27309	0.49 [0.45, 0.52]	0.70 [0.70, 0.71]





Orbit at threshold ≥3

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Elvira-Ruiz, 2020	89	667	96	2027	0.48 [0.41, 0.56]	0.75 [0.74, 0.77]
Esteve pastor 2016	26	238	20	992	0.57 [0.41, 0.71]	0.81 [0.78, 0.83]
Lip, 2018	457	9022	776	47675	0.37 [0.34, 0.40]	0.84 [0.84, 0.84]
Proietti 2016	16	249	111	3174	0.13 [0.07, 0.20]	0.93 [0.92, 0.94]
Proietti, 2018	190	1148	219	4519	0.46 [0.42, 0.51]	0.80 [0.79, 0.81]
Rivera Caravaca 2017	85	234	165	877	0.34 [0.28, 0.40]	0.79 [0.76, 0.81]
Senoo 2016	12	165	62	2044	0.16 [0.09, 0.27]	0.93 [0.91, 0.94]
Yao 2017	242	6562	423	32312	0.36 [0.33, 0.40]	0.83 [0.83, 0.83]



Appendix G: Clinical evidence tables

Table 37. Apostolakis, 2012[4]

Reference	Apostolakis, 2012[4]
Study type	Retrospective cohort study
Study sample	2,293 patients with AF on VKAs, from AMADEUS RCT trial in UK. Age 70, 65% male, 77% hypertension, 20% DM, 13.5% previous stroke, 31% CAD, **18% antiplatelet treatment**, TTR 0.57. Drops outs NR. No blinding reported.
Inclusion criteria	Not reported
Exclusion criteria	Not reported
Anticoagulants used	Adjustable dose VKA
Risk tools used	HAS-BLED HEMORRHAGES ATRIA
Outcome definition	Serious bleeding – any clinically relevant bleeding (sub-classified as MB and CRNMB)
Mean follow up time	429 days
Number of bleeding events	251 people with 'any clinically relevant bleeding' and 39 with major bleeding
Results	C statistic for any clinically relevant bleeding HEMORRHAGES: 0.55(0.51-0.59) HAS-BLED: 0.60(0.56-0.63) ATRIA: 0.50(0.46-0.54) On head-to head analysis HAS-BLED better than HEMORRHAGES and ATRIA (p<0.002, <0.002) but ATRIA and HEMORRHAGES NS. C statistic for major bleeding HEMORRHAGES: 0.60(0.51-0.69)

Reference	Apostolakis, 2012[4]
	HAS-BLED: 0.65(0.56-0.73) ATRIA: 0.61(0.51-0.70) On head-to head analysis none significantly better than any other Sensitivity/specificity (extracted from tables) for CRB HEMORRHAGES ≥1: 0.742/0.384 ≥2: 0.266/0.77 HASBLED ≥1: 0.952/0.081 ≥2: 0.73/0.39 ATRIA ≥1: 0.879/0.113 ≥2: 0.411/0.583 Sensitivity/specificity (extracted from tables) for MB HEMORRHAGES ≥1: 0.794/0.345 ≥2: 0.358/0.768 HASBLED ≥1: 0.948/0.0786 ≥2: 0.846/0.382 ATRIA ≥1: 0.923/0.010 ≥2: 0.589/0.581 NRI clinically relevant bleeding HAS-BLED v HEMORRHAGES: +0.103 (p<0.001) HAS-BLED v ATRIA: +0.13 (p<0.001)

Reference	Apostolakis, 2012[4]
	ASTRIA v HEMORRHAGES +0.021 (p=0.55) NRI major bleeding HAS-BLED v HEMORRHAGES: +0.068 (p=0.42) HAS-BLED v ATRIA: +0.090 (p=0.33) ATRIA v HEMORRHAGES -0.022 (p=0.82) Calibration Hosmer-Lemeshow goodness of fit statistics showed good calibration for all tools showed by a p value >0.05

Table 38. Apostolakis, 2013[3]

Reference	Apostolakis, 2013[3]
Study type	Retrospective cohort study
Study sample	2,293 patients with AF that had been randomised to VKAs, from AMADEUS RCT trial in UK. Age 70, CHADS2 score 2.1. Age 70, 65% male, 77% hypertension, 20% DM, 13.5% previous stroke, 31% CAD, **18% antiplatelet treatment**, TTR 0.57. Drops outs NR. No blinding reported.
Inclusion criteria	AF on VKAs
Exclusion criteria	Not reported
Anticoagulants used	Warfarin
Risk tools used	HAS-BLED CHADS2 CHADSVASC
Outcome definition	Serious bleeding – any clinically relevant bleeding
Mean follow up time	429 days

Reference	Apostolakis, 2013[3]
Number of bleeding events	251 people with 'any clinically relevant bleeding'. 39 major bleeding
Results	C statistic for clinically relevant bleeding HAS-BLED: 0.60(0.56-0.63) CHADS2: 0.51(0.47-0.55) CHADSVASC: 0.53(0.49-0.57) Head to head: HAS-BLED better than both CHADS2 and CHADSVASC (P<0.001 and 0.001) Sensitivity/specificity (extracted from tables) for CRB HAS-BLED ≥1: 0.952/0.081 ≥2: 0.73/0.39 CHADS ≥1: 0.972/0.0230 ≥2: 0.637/0.385 CHADSVASC ≥2: 0.936/0.079 ≥3: 0.753/0.292 NRI for clinically relevant bleeding (categorical) HAS-BLED v CHADS2: +0.13 (+0.05 to +0.21) HAS_BLED v CHADSVASC: +0.10 (+0.004 to +0.19) NRI for clinically relevant bleeding (continuous) HAS-BLED v CHADS2: +0.16 (+0.03 to +0.29) HAS_BLED v CHADSVASC: +0.29 (+0.16 to +0.42)

Table 39. Barnes, 2014[8]

Reference	Barnes, 2014[8]
Study type	Prospective cohort study
Study sample	2600 patients with NVAF and on warfarin were recruited. USA study. Age 70, 41.7% female, hypertension 75%, DM 25%, CAD 33%, CHF 24.2%, current smoking 6%, renal disease 12%, stroke 11.5%, bleeding diasthesis 31%, HAS-BLED score 2.6, CHADS2 score 3.4. TTR 59.3. **Antiplatelets/NSAIDs not reported**. No blinding. No data loss reported.
Inclusion criteria	Not reported
Exclusion criteria	Not reported
Anticoagulants used	Warfarin
Risk tools used	CHADS2 CHADSVASC HEMORRHAGES HAS-BLED ATRIA
Outcome definition	First major bleeding event, defined according to the International Society of Thrombosis and Haemostasis consensus.
Mean follow up time	Mean of 1 year (2581.6 years of follow up)
Number of bleeding events	110 patients had major bleeding.
Results	C statistics (continuous) for major bleeding at 1 year CHADS2 0.53(0.47-0.60) CHADSVASC 0.56(0.49-0.62) HEMORRHAGES 0.66(0.61-0.74) HAS-BLED 0.69(0.63-0.75) ATRIA 0.67(0.61-0.74) Head to head: sig differences for HAS-BLED v CHADS and CHADSVASC, ATRIA and CHADS and CHADSVASC and HEMORRHASGES v CHADS and CHADSVASC. NRI for major bleeding at one year HAS-BLED v ATRIA: +0.26 (p=0.006) HAS-BLED v HEMMORRHAGES: +0.31 (p=0.001)

Reference	Barnes, 2014[8]
	HAS-BLED v CHADS2: +0.58 (p<0.001) HAS-BLED v CHADSVASC: +0.36 (p<0.001) ATRIA v HEMORRHAGES: +0.34 (p=0.001) ATRIA v CHADS2: +0.59 (p<0.001) ATRIA v CHADSVASC: +0.40 (p<0.001) HEMORRHAGES v CHADS2: +0.54 (p<0.001) HEMORRHAGES v CHADSVASC: +0.54 (p<0.001) CHADS2 v CHADSVASC: -0.071 (p=0.25)

Table 40. Beshir, 2018[14]

Reference	Beshir, 2018[14]
Study type	Retrospective cohort study
Study sample	1017 patients with NVAF and on Warfarin (INR 2-3), dabigatran or rivaroxaban between 2010 and 2015. Malaysia. Age >75: 27%, 52% male, hypertension 82%, IHD 33%, renal impairment 36%, DM 40%, prior stroke/TIA: 22%, CHF: 20%. CHADS2: 2. **35% on antiplatelets**. No blinding. 291 lost to follow up from original sample of 1308 patients.
Inclusion criteria	NVAF, aged >18, using OACS for at least 1 year. If follow up was <1 year but there was an OAC-related bleeding event, then inclusion was also allowed.
Exclusion criteria	<1 year follow up.
Anticoagulants used	Warfarin (n=290), rivaroxaban (n=106), dabigatran (n=621)
Risk tools used	mOBRI CBRM HEMORRHAGES HAS-BLED ATRIA ORBIT
Outcome definition	Major bleeding (ISTH) Clinically relevant non-major bleeding (ISTH)

Reference	Beshir, 2018[14]
	Minor bleeding (ISTH)
Mean follow up time	1 year
Number of bleeding events	Major bleeding: 23 CRNMB: 76
Results	C statistics for major bleeding mOBRI: 0.54(0.42-0.66) CBRM: 0.61(0.51-0.71) HEMORRHAGES: 0.71(0.60-0.82) HAS-BLED: 0.58(0.46-0.69) ATRIA: 0.70(0.58-0.82) ORBIT: 0.69(0.59-0.80) C statistics for CRNMB mOBRI: 0.56(0.50-0.62) CBRM: 0.58(0.54-0.62) HEMORRHAGES: 0.61(0.55-0.68) HAS-BLED: 0.51(0.45-0.58) ATRIA: 0.61(0.54-0.67) ORBIT: 0.61(0.54-0.68) Calibration Hosmer-Lemeshow goodness of fit test: Non significant for all risk tools (no data reported)

Table 41. Berg, 2019[11]

Reference	Berg, 2019[11]
Study type	External validation prospective cohortstudy
Study sample	8705 patients from the ENGAGE trial (sub-study).Details unclear
Inclusion criteria	Patients enrolled on the ENGAGE AF-TIMI 48 trial, who were therefore taking VKAs or edoxaban. Participation in this sub-study was offered to all enrolled patients until recruitment reached 9000 participants
Exclusion criteria	None reported
Anticoagulants used	Warfarin oredoxaban. Numbers unclear
Risk tools used	HAS-BLED ABC-bleeding
Outcome definition	Major bleeding (ISTH definition), adjudicated by an independent clinical events committee.
Mean follow up time	2.8 years
Number of bleeding events	Unclear
Results	**Major bleeding** Harrell's C index HAS-BLED: 0.62(0.60-0.64) ABC-bleeding: 0.69(0.66-0.71) NRI at 3 years for ABC-bleeding vs HAS-BLED + 0.138 (0.080 – 0.228)[predominantly due to correct downclassification] Calibration The Nam-D'Agostino statistics for calibration (nonsignificant P values indicate adequate calibration) for the ABC-bleeding scores at 3 years were 14.6 (p=0.10).

Table 42.Chang, 2016[19]

Reference	Chang, 2016[19]
Study type	Prospective cohort study
Study sample	208 patients (213 enrolled and 5 lost to FU) with NVAF on dabigatran (either 100mg or 150mg/day). Taiwan. Age 74.7, 67.9% male, 36% history of stroke, 24.5% DM, 79.3% hypertension, 18.8% CAD, 16.3% HF, **antiplatelets/NSAIDs 12.5%,**renal disease 0.5%, history of GI bleeding 23.6%, HAS-BLED 1.8. 5 lost to follow up from original cohort of 213. No blinding.
Inclusion criteria	NVAF and on dabigatran
Exclusion criteria	None reported
Anticoagulants used	Dabigatran (110 or 150 mg)
Risk tools used	HTI APTT Prothrombin time
Outcome definition	Major bleeding (2005 ISTH)
Mean follow up time	1 year
Number of bleeding events	17 bleeding events
Results	C statistics Hemoclot thrombin inhibitor levels (HTI): 0.65 (p=0.036) Prothrombin time: 0.54(0.47-0.62) Activated partial thromboplastin time (APTT): 0.58(0.50-0.69) Sensitivity of HTI at cut-off of 117.7 ng/ml: 0.59 Specificity of HTI at cut-off of 117.7 ng/ml: 0.71

Table 43. Chao, 2018[21]

Reference	Chao, 2018[21]
Study type	Retrospective cohort study
Study sample	40,450 AF patients (defined as cases where there had been at least 2 confirmed outpatient diagnoses of AF) receiving warfarin between 1998 and 2011 in Taiwan. Age 67.3, male 55.7%, hypertension 67.4%, abnormal renal function 13.2%, stroke 43%, history of bleeding 18%, use of **antiplatelets 22.7%, NSAIDs 7.2%**, HAS-BLED 2.51. No loss to FU. No blinding reported.

Reference	Chao, 2018[21]
Inclusion criteria	NVAF and on warfarin
Exclusion criteria	Not reported
Anticoagulants used	Warfarin
Risk tools used	Modifiable Bleeding Risk factors score (MBR) HEMORRHAGES HAS-BLED ATRIA ORBIT
Outcome definition	Major bleeding (GI, GU or RT bleeding requiring hospitalisation or transfusion) ICH
Mean follow up time	4.6 years
Number of bleeding events	6889 people with major bleeds including 1581 with ICH.
Results	C statistics major bleeding HEMORRHAGES: 0.559(0.552-0.567) ATRIA: 0.558(0.551-0.565) ORBIT: 0.551(0.544-0.559) MBR: 0.525(0.518-0.533) HAS-BLED: 0.562(0.554-0.569) C statistics ICH HEMORRHAGES: 0.525(0.510-0.539) ATRIA: 0.504(0.490-0.518) ORBIT: 0.497(0.483-0.511) MBR: 0.517(0.502-0.531) HAS-BLED: 0.527(0.513-0.541) NRI for major bleeding HAS-BLED v HEMORRHAGES: +0.043(0.027 to 0.059)

Reference	Chao, 2018[21]
	HAS-BLED v ATRIA: +0.049(0.032 to 0.066) HAS-BLED v ORBIT: +0.055(0.038 to 0.073) HAS-BLED v MBR: +0.056(0.043 to 0.068) MBR v HEMORRHAGES: -0.012(-0.032 to 0.007) MBR v ATRIA: -0.007(-0.027 to 0.014) MBR v ORBIT: +0.000(-0.021 to 0.021) MBR v MBR: -0.056(-0.068 to 0.043) NRI for ICH HAS-BLED v HEMORRHAGES: +0.030(-0.001 to 0.060) HAS-BLED v ATRIA: +0.060(0.026 to 0.093) HAS-BLED v ORBIT: +0.048(0.013 to 0.082) HAS-BLED v MBR: +0.007(-0.018 to 0.033) MBR v HEMORRHAGES: -0.022(-0.062 to 0.017) MBR v ATRIA: -0.052(-0.094 to -0.011) MBR v ORBIT: -0.040(-0.083 to 0.002) MBR v MBR: -0.007(-0.033 to 0.018)

Table 44. Chao, 2018[20]

Reference	Chao, 2018[20]
Study type	Retrospective cohort study
Study sample	19,566 AF patients on Warfarin **and a HAS_BLED score of ≤2**identified from the NHIRD of Taiwan (1998-2011). Age 63.8, male 57.4%, hypertension 52.6%, abnormal renal function 3.4%, stroke 22.6%, bleeding 6.9%, **antiplatelet / NSAID drugs 2.3%.**No loss to FU reported. No blinding reported.
Inclusion criteria	AF, >20 years, CHADSVASC >1 for males and >2 for females, on warfarin, HAS-BLED score ≤2.
Exclusion criteria	None reported
Anticoagulants used	Warfarin
Risk tools used	HAS-BLED baseline HAS-BLED change from baseline (Delta HAS-BLED) HAS-BLED follow up Number of modifiable risk factors
Outcome definition	Major bleeding – bleeding from IC or GI, UG, RT requiring hospitalisation and transfusion.
Mean follow up time	4.8 years
Number of bleeding events	3032 patients with major bleeding events (ICH in 671 of these)
Results	C statistics Baseline HAS-BLED: 0.54(0.53-0.55) Delta HAS-BLED: 0.62(0.61-0.63) HAS-BLED follow up: 0.63(0.62-0.64) Number of modifiable risk factors: 0.49(0.48-0.50) Sensitivity/specificity HAS-BLED ≥1: 0.921/0.175 ≥2: 0.598/0.475 NRI (Follow up HAS-BLED v Delta HAS-BLED): +0.033 (+0.0184 to 0.0476) Note: Although only baseline prediction scores would normally be clinically useful (because it is at baseline where decisions are

Reference	Chao, 2018[20]
	normally made about anticoagulation) this study does show that repeat prediction measures may allow more accurate prediction that can be used to modify management.

Table 45. Claxton, 2018[23]

Reference	Claxton, 2018[23]
Study type	Retrospective cohort study
Study sample	81,285 NVAF patients on Warfarin or DOACs (initiated at baseline). Netherlands. This was an external validation cohort from the Optum Clinformatics database from 2009-2015. For warfarin group (largest) the demographics were: age 73.9, 44% woman, HAS-BLED 2.8, HF 45.5%, CHD: 47.3%, hypertension 89%, DM 39.9%, stroke 33.4%, PAD 25.7%, kidney disease 25.9%, prior GI bleed 16%, prior IC bleed: 2.1%, prior other bleed 16%. No blinding reported. No loss to follow up (as retrospective). **No data on antiplatelets/NSAIDS**
Inclusion criteria	NVAF
Exclusion criteria	None reported
Anticoagulants used	Warfarin (n=49,894), dabigatran (n=9088), rivaroxaban (n=14,043), apixaban (n=8260)
Risk tools used	Anticoagulation-Specific Bleeding Score (ABS) HAS-BLED ATRIA HEMORRHAGES ORBIT
Outcome definition	Major bleeding (with hospitalisation)
Mean follow up time	1 year
Number of bleeding events	3,238 major bleeds (2420 warfarin, 282 dabigatran, 411 rivaroxaban, 125 apixaban)
Results	Model discrimination of ABS in the validation dataset for each anticoagulant (Optum Clinformatics) Warfarin 0.67 (0.65, 0.68)

Reference	Claxton, 2018[23]
	Dabigatran 0.72 (0.69, 0.76) Rivaroxaban 0.70 (0.68, 0.73) Apixaban 0.72 (0.67, 0.77) For the other risk tools, C statistics are only given for all patients (not specified by OAC): Anticoagulation-Specific Bleeding Score (ABS): 0.68(0.67-0.69) HAS-BLED: 0.63(0.62-0.65) ATRIA: 0.65(0.64-0.66) HEMORRHAGES: 0.64(0.63-0.65) ORBIT: 0.65(0.64-0.66) Data for calibration analysis not given, but stated to be adequate for ASBC. Calibration plot given as below:

Reference	Claxton, 2018[23]

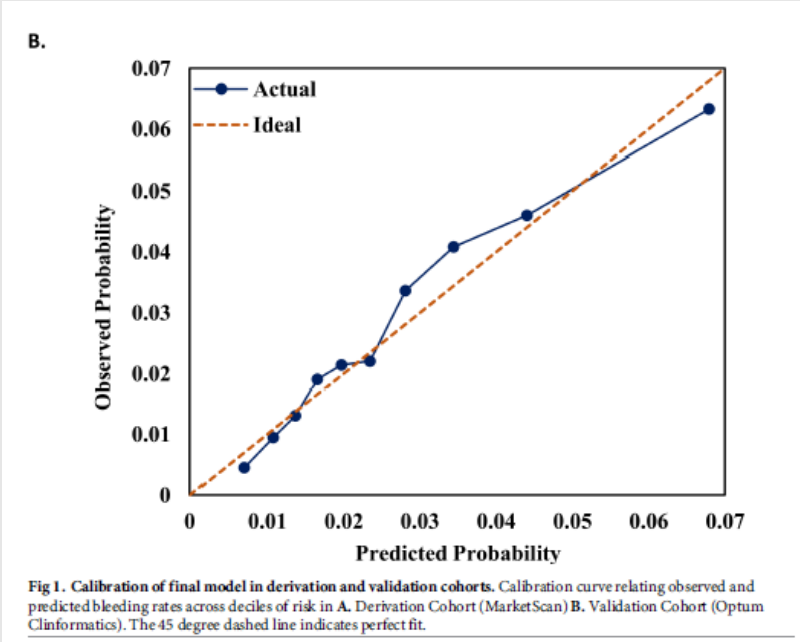


Fig 1. Calibration of final model in derivation and validation cohorts. Calibration curve relating observed and predicted bleeding rates across deciles of risk in **A.** Derivation Cohort (MarketScan) **B.** Validation Cohort (Optum Clinformatics). The 45 degree dashed line indicates perfect fit.

Table 46. Dalgaard, 2019[25]

Reference	Dalgaard, 2019[25]
Study type	Retrospective cohort study
Study sample	51, 180 people with NVAF and on OACs from the Danisjh Nationwide Registries. Taken from a larger cohort of 90,693 which included those not on OACs
Inclusion criteria	Age 18 or over with NVAF
Exclusion criteria	Rheumatic valve disease; valve surgery
Anticoagulants used	Unclear
Risk tools used	GARFIELD-AF

Reference	Dalgaard, 2019[25]
	HAS-BLED
Outcome definition	Major bleeding
Mean follow up time	1 year
Number of bleeding events	1492, but this may include hemorrhagic stroke numbers, so does not necessarily represent major bleeding events
Results	C statistics (major bleeding) GARFIELD 0.64(0.63-0.66) HAS-BLED 0.64(0.63-0.65) No calibration data presented that relates to the relevant group on OACs

Table 47. Esteve-Pastor, 2016[31]

Reference	Esteve-Pastor, 2016[31]
Study type	Prospective cohort study
Study sample	1276 patients with chronic NVAF on VKA or DOAC for at least 6 months before enrolment (FANTASIIA population). SPAIN. There was another cohort of 406 patients in this paper that underwent electrical cardioversion, and they are not included in this extraction. Age 74, 44% male, 80.6% hypertensive, 30% HF, 29.3% DM, 6.6% VD, 12.9% previous embolism, 3.8% previous bleeding, 10% renal impairment, 1.3% liver impairment, 77.4% VKA, 22.6% DOACs, **10.9% on NSAIDS / antiplatelets**. HAS-BLED score: 2. TTR 60.9. No blinding. No loss to FU reported.
Inclusion criteria	On VKA or DOAC for at least 6 months before enrolment
Exclusion criteria	None reported
Anticoagulants used	VKA and DOACS
Risk tools used	HAS-BLED ORBIT
Outcome definition	Major bleeding (2005 ICTH)
Mean follow up	1 year

Reference	Esteve-Pastor, 2016[31]
time	
Number of bleeding events	46 patients with major bleeding events
Results	C statistics major bleeding HAS-BLED: 0.63(0.56-0.71) ORBIT 0.70(0.62-0.77) Sensitivity/specificity HASBLED ≥2: 0.847/0.320 ≥3: 0.456/0.706 ORBIT ≥3: 0.560/0.806 ≥4: 0.413/0.904

Table 48. Esteve-Pastor, 2017a[5]

Reference	Esteve-Pastor, 2017a[5]
Study type	Prospective cohort study
Study sample	1,120 patients with paroxysmal, persistent or permanent AF, stable on VKAs (INR 2-3). Spain. Age 76, 49.5% male, 82% hypertension, 27%DM, 33% dyslipidaemia, 15.5% current smoker, 31.2% HF, 19.6% CAD, 19% previous stroke, 8.4% previous bleeding. TTR at 6 months 80, CHADSVASC 4, HAS-BLED 2, ABC 16.5. Number on antiplatelets – not reported. No loss to FU reported. No blinding.
Inclusion criteria	TTR 100%
Exclusion criteria	Rheumatic valve disease, prosthetic heart valves, haemodynamic instability, ACS, or hospital admission/surgery in past 6 months
Anticoagulants used	VKAs

Reference	Esteve-Pastor, 2017a[5]
Risk tools used	ABC-bleedingCrC HAS-BLED
Outcome definition	Major bleeding (2005 ICTH)
Mean follow up time	6.5 years
Number of bleeding events	207 patients with MB events. Of these, there were 65 ICH, 85 GI bleeding.
Results	C index major bleeding ABC-bleedingCrC: 0.518(0.488-0.548) HAS-BLED: 0.583(0.554-0.612) C index ICH ABC-bleedingCrC: 0.465(0.399-0.530) HAS-BLED: 0.559(0.486-0.632) C index GI bleeding ABC-bleedingCrc: 0.569(0.504-0.635) HAS-BLED: 0.606(0.539-0.673) Sensitivity/specificity HAS-BLED Major bleeding ≥3: 0.570/0.597 ABCCrCMajor bleeding >2%: 0.835/0.194 HAS-BLED ICH ≥3: 0.538/0.572 ABCCrC ICH >2%: 0.785/0.186 NRI major bleeding ABCCrCvs HAS-BLED: -0.1374(p=0.005)

Reference	Esteve-Pastor, 2017a[5]
	NRI ICH ABCCrCvs HAS-BLED: -0.1396(p=0.075) NRI GI bleeding ABCCrCvs HAS-BLED: -0.08174(p=0.362)

Table 49. Esteve-Pastor, 2017b[32]

Reference	Esteve-Pastor, 2017b[32]
Study type	Retrospective cohort study
Study sample	4576 patients with paroxysmal, persistent or permanent AF. 2283 on warfarin and 2293 on Idraparinux. Taken from the multinational AMADEUS database. Spain. Age 71, 66.5% male, **21.4% on anti-platelets or NSAID**, 77% hypertensive, 20%DM, 23% HF, 31% CAD, 13% previous stroke, TTR 58, CHADSVASC 3, HAS-BLED 2, Modifiable bleeding risks score 1. No loss to FU reported. **Assessors BLINDED.**
Inclusion criteria	In AMADEUS trial
Exclusion criteria	Contraindications to OACs, alcohol abuse, terminal renal dysfunction, breastfeeding, pregnancy and recent or anticipated hospital admission/surgery with potential for uncontrolled bleeding.
Anticoagulants used	VKAs
Risk tools used	HAS-BLED Modifiable bleeding risk factors score
Outcome definition	Major bleeding (2005 ICTH) Clinically relevant non-major bleeding event (repetitive epistaxis for >5mins in 24 hours, or haematuria, haemetmesis and subcutaneous haematomas of >25cm2 (spontaneous) or >100cm2 if after trauma.
Mean follow up time	347 days

Reference	Esteve-Pastor, 2017b[32]
Number of bleeding events	113 patients with MB events and 597 with any clinically relevant bleeding event.
Results	C index any clinically relevant bleeding HAS-BLED: 0.545(0.530-0.559) Modifiable bleeding risk factors score: 0.530(0.515-0.544) Head-to-head: HAS-BLED significantly better than MBRF score (p=0.04)

Table 50. Fang, 2011[33]

Reference	Fang, 2011[33]
Study type	Retrospective cohort study
Study sample	3063 patients in the validation cohort, taken from 9,186 patients with NVAF on warfarin (median exposure 3.5 years), taken from the ATRIA study (USA). AF defined as any ICD-9 codes. Demographic data not given for validation cohort. No blinding or loss to FU reported.
Inclusion criteria	Not reported
Exclusion criteria	Not reported
Anticoagulants used	Warfarin
Risk tools used	ATRIA Outpatient Bleeding Index Kuijer et al. Kearon et al. HEMORRHAGES Shireman Riete risk scheme
Outcome definition	Major bleeding, defined as fatal, requiring transfusion of >2 U packed cells, or haemorrhage into a critical anatomical site (ie intracranial or retroperitoneal). Only bleeding events occurring within 5 days of preceding Warfarin exposure were included.
Mean follow up time	Approximately 3 years

Reference	Fang, 2011[33]
Number of bleeding events	154 first major bleed
Results	C statistics on validation dataset (continuous scores) ATRIA: 0.74(0.72-0.76) Outpatient Bleeding Index: 0.68(0.65-0.70) Kuijer et al.: 0.57(0.54-0.59) Kearon et al.: 0.69(0.67-0.71) HEMORRHAGES: 0.71(0.69-0.73) Shireman: 0.70(0.68-0.73) Riete risk scheme: 0.68(0.65-0.70) C statistics on validation dataset (categorical scores) ATRIA: 0.69(0.66-0.71) Outpatient Bleeding Index: 0.59(0.58-0.61) Kuijer et al.: 0.56(0.55-0.58) Kearon et al.: 0.67(0.65-0.69) HEMORRHAGES: 0.67(0.65-0.70) Shireman: 0.64(0.61-0.66) Riete risk scheme: 0.63(0.61-0.66) NRI on validation dataset (versus ATRIA). NB: In paper signs given as positive but clear from text that they should be negative. Outpatient Bleeding Index: -0.505 Kuijer et al.: -0.566 Kearon et al.: -0.277 HEMORRHAGES: -0.289 Shireman: -0.344 Riete risk scheme:-0.448

Table 51. Fox, 2017[36]

Reference	Fox, 2017[36]
Study type	Retrospective Cohort study
Study sample	25,285 patients with AF that were on OACs. 8804 on DOACs and 16,491 on VKAs. Details of the characteristics of these patients are not reported. No blinding reported.
Inclusion criteria	People with incident or prevalent AF
Exclusion criteria	Not reported
Anticoagulants used	DOAC(undefined) and VKA
Risk tools used	GARFIELD AF Risk HAS-BLED
Outcome	Major bleeding (undefined, but includes haemorrhagic stroke)
Mean follow up time	Up to 3 years
Number of bleeding events	305 at 1 year and 625 at 3 years (based on N of 7442 – unclear why this is not 25,285 referred to above, but may relate to these being the number with a 3 year follow up)
Results	C statistics GARFIELD-AF risk model ATRIA score 1-yr Major bleed (treated patients) 0.61 (0.58-0.64) 0.65 (0.62-0.68) 3-yr Major bleed (treated patients) 0.61 (0.59-0.63) 0.65 (0.62-0.67)

Table 52. Friberg, 2012[37]

Reference	Friberg et al. 2012[37]
Study type	Retrospective cohort study.
Study sample	48, 599 patients with AF (defined by ICD-10 code 1489 with or without subscales A-F) using Warfarin at baseline identified from the Swedish National Discharge Registry. Demographic data stated to be in supplementary file but not available in that file who were on warfarin. This subset was taken from an overall cohort of 170 291 which included those not on anticoagulants. No blinding reported.
Inclusion criteria	All individuals with a diagnosis of AF, between July 2005 and December 2008 who were known to have used Warfarin or other OACs at baseline. A further subset of people using OACS and aspirin were analysed separately and these are not included.

Reference	Friberg et al. 2012[37]
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.
Anticoagulants used	Warfarin
Risk tools used	HAS-BLED and HEMORRHAGES
Outcome definition	Primary: Intracranial haemorrhage (defined by ICD-10 code I60-62). Secondary: major bleeding (including all IC bleeds, all GI bleeds and diagnosis of anaemia secondary to bleeding). A blanking period of 14 days was also used, that excluded events occurring in first 14 days.
Mean follow up time	1.5 years
Number of bleeding events	0.6 IC bleeds per year and 1.9 major bleeds per year in those taking OACs.
Results	C statistics for IC and major bleeding IC bleeding HAS-BLED: 0.60 (0.58-0.68) HEMORRHAGES: 0.62 (0.60-0.64) Major bleeding HAS-BLED: 0.61 (0.59-0.62) HEMORRHAGES: 0.63 (0.61-0.64)

Table 53. Gage, 2006[38]

Reference	Gage, 2006[38]
Study type	Retrospective cohort study
Study sample	1604 medicare beneficiaries on NRAF (USA) with chart-confirmed AF on warfarin. 69.2% aged > 75 years, 7.9% hepatic or renal

Reference	Gage, 2006[38]
	disease, 4.8% malignancy, 37.2% previous stroke, 0.4% uncontrolled hypertension. Also on Aspirin: 7.04%. No blinding or loss to FU reported.
Inclusion criteria	Not reported
Exclusion criteria	Not reported
Anticoagulants used	Warfarin
Risk tools used	Landefeld and Goldman and Beyth et al: 0.65 Kuijer et al: 0.58 Kearon et al: 0.66 HEMORRHAGES: 0.67
Outcome definition	Major bleeding
Mean follow up time	Unclear, but appears to be around 1 year
Number of bleeding events	4.9 bleeds per 100 patient-years
Results	C statistics Landefeld and Goldman and Beyth et al: 0.65 Kuijer et al: 0.58 Kearon et al: 0.66 HEMORRHAGES: 0.67 Sensitivity/specificity HEMORRHAGES ≥1:0.94/0.133 ≥2:0.776/0.456 ≥3:0.478/0.739

Table 54. Gallego, 2012[39]

Reference	Gallego, 2012[39]
Study type	Retrospective cohort study
Study sample	965 consecutive anticoagulated people with permanent or paroxysmal AF, with at least 6 months of anticoagulation with acenocoumarol (INR 2-3). 50% male, mean age 76, hypertension 57%, DM 25.5%, HF 36.5%, prev. stroke/TIA 19%, renal impairment 10%, CAD 4%, hypercholesterolemia 31%, current smoking 14%, previous bleeding 8.5%, median HAS-BLED 2, CHADS2 score 2. Antiplatelet therapy 16.6%. 95 died during FU. No blinding reported.
Inclusion criteria	INR 2-3
Exclusion criteria	Prosthetic heart valves, ACS, stroke, valvular AF, haemodynamic instability, any surgical treatment of hospital admission in past 6 months.
Anticoagulants used	VKA (acenocoumarol)
Risk tools used	HAS-BLED
Outcome definition	Major bleeding – 2005 International Society on Thrombosis and Haemostasis criteria.
Mean follow up time	861 days
Number of bleeding events	75 people had major bleeding (15 ICH)
Results	C statistic major bleeding HAS-BLED: 0.70 (0.64-0.76)

Table 55. Garcia-Fernandez, 2017[41]

Reference	Garcia-Fernandez, 2017[41]
Study type	Prospective cohort study
Study sample	1215 patients with NVAF on VKA at INR 2-3. Age 76, male 49.3%, hypertension 82.5%, DM 26.4%, HF 31.1%, IHD 19%, previous stroke 18.4%, previous bleeding 8.4%, renal disease 10.3%, **antiplatelet drugs 17.8%,**HAS-BLED score 2. No loss to FU or blinding reported.

Reference	Garcia-Fernandez, 2017[41]
Inclusion criteria	NVAF, INR 2-3
Exclusion criteria	Valvular AF; prosthetic valve replacements; or acute coronary syndrome, stroke, hemodynamic instability, hospital admissions or surgical interventions in previous 6 months
Anticoagulants used	VKA
Risk tools used	vWF HAS-BLED HAS-BLED + vWF
Outcome definition	Major bleeding
Mean follow up time	2373 days
Number of bleeding events	222 people with major bleeding
Results	C statistics vWF: 0.61(0.57-0.65) [ROC curve indicated optimum cut off at 197 UI/dL] HAS-BLED: 0.592(0.564-0.620) HAS-BLED + vWF: 0.614(0.586-0.641) IDI HAS-BLED v HAS-BLED +vWF = 0.0105 (p=0.056) NRI HAS-BLED with vWF v HAS-BLED +0.012 (p=0.735)

Table 56. Hijazi, 2014a[57]

Reference	Hijazi, 2014a[57]
Study type	Retrospective cohort study
Study sample	14,897 patients with AF on apixaban or warfarin, from the ARISTOTLE trial. Likely to be a multinational multi-centre trialbut not

Reference	Hijazi, 2014a[57]
	reported. Ranges of baseline data given as data given for different categories of TnT. Age 64-74, male 53.8-74.6%, CHF 28-47%, hypertension 87%, DM 18-32%, Prior stroke/TIA 16-21%, MI 6-19%. Aspirin 28-34%. Warfarin 53.2-55.7%. BLINDED ASSESORS of BLEEDING. No loss to FU reported.
Inclusion criteria	Not reported
Exclusion criteria	Not reported
Anticoagulants used	Apixaban and warfarin
Risk tools used	CHADSVASC CHADSVASC with TnT
Outcome definition	Major bleeding (2005 ISTH)
Mean follow up time	Median 1.9 years
Number of bleeding events	674
Results	C statistic for major bleeding (not differentiated according to OAC) CHADSVASC: 0.591 CHADSVASC with TnT 0.629(0.609-0.650) TnT alone:0.617(0.596-0.637)

Table 57. Hijazi, 2014[56]

Reference	Hijazi, 2014[56]
Study type	Retrospective cohort study
Study sample	14,821 patients with AF on apixaban or warfarin, from the ARISTOTLE trial. Overlap with Hijazi, 2014[57]in terms of sample, but this study used a different risk tool. Likely to be a multinational multi-centre trial but not reported. Ranges of baseline data given as data given for different categories of TnI. Age 66-72, male 6--70%, CHF 24-51%, hypertension 87%, DM 21-28%, Prior stroke/TIA 16-21%, MI 6-19%. **Aspirin 29-34%**. Warfarin 49.9-56.5%. BLINDED assessors. No loss to FU reported.
Inclusion criteria	Not reported

Reference	Hijazi, 2014[56]
Exclusion criteria	Not reported
Anticoagulants used	Apixaban and warfarin
Risk tools used	HAS-BLED HAS-BLED with TnI
Outcome definition	Major bleeding (2005 ISTH)
Mean follow up time	Median 1.9 years
Number of bleeding events	674
Results	C statistic for major bleeding (not differentiated according to OAC) HAS-BLED: 0.606 HAS-BLED with TnI 0.630 TnI alone: 0.598

Table 58. Hijazi, 2016[54]

Reference	Hijazi, 2016[54]
Study type	Retrospective cohort study
Study sample	External validation in 8468 patients with AF (67% permanent or persistent) randomised to dabigatran and warfarin in the multinational RE-LY trial. Age 72, 26% women, 44% on antiplatelets or NSAISs, 8% current smokers, 22% DM, 79% hypertension, 29% CHF, 13% previous clinically relevant bleeding, 19% previous stroke/TIA, 17% previous MI, 4% previous PAD, 19% vascular disease, Renal function CKD-EPI 68.2. ASSESSOR BLINDING. No loss to FU reported.
Inclusion criteria	Not reported
Exclusion criteria	Not reported
Anticoagulants used	Dabigatran and Warfarin

Reference	Hijazi, 2016[54]
Risk tools used	HAS-BLED ORBIT ABC-bleeding ABC-bleeding (cTnI-hs) ABC-bleeding (cystatin C) ABC-bleeding (CKD-EPI)
Outcome definition	Major bleeding: 2005 ISTH, adjudicated by a **blinded**clinical events committee.
Mean follow up time	1.9 years
Number of bleeding events	463 (all) 159 (warfarin) 304 (DOAC: dabigatran)
Results	C statistics ALL patients n=8468 ABC-bleeding: 0.71(0.68-0.73) ABC-bleeding: (cTnI-hs) 0.71(0.68-0.73) ABC-bleeding (cystatin C): 0.68(0.64-0.71) ABC-bleeding (CKD-EPI): 0.69(0.66-0.71) ORBIT: 0.68(0.65-0.70) HAS-BLED: 0.62(0.59-0.64) Warfarin patients n=2814 ABC-bleeding: 0.65(0.61-0.70) ABC-bleeding: (cTnI-hs) 0.65(0.61-0.70) ABC-bleeding (cystatin C): 0.60(0.54-0.66) ABC-bleeding (CKD-EPI): 0.65(0.60-0.69) ORBIT: 0.63(0.58-0.67) HAS-BLED: 0.60(0.56-0.64) DOAC(dabigatran) patients n=5350

Reference	Hijazi, 2016[54]
	ABC-bleeding: 0.74(0.71-0.76) ABC-bleeding: (cTnI-hs) 0.74(0.71-0.76) ABC-bleeding (cystatin C): 0.72(0.68-0.75) ABC-bleeding (CKD-EPI): 0.71(0.69-0.74) ORBIT: 0.70(0.67-0.73) HAS-BLED: 0.62(0.59-0.65) Calibration ABC showed good discriminative ability in the different sub-groups of patients with AF. Calibration plot in Appendix but cannot access.

Table 59. Hijazi, 2017[52]

Reference	Hijazi, 2017[52]
Study type	Retrospective cohort study
Study sample	8,474 AF patients (with at least 1 additional risk factor for stroke) taken from the RE-LY study, on dabigatran or warfarin. Baseline characteristics given as ranges as sub-grouped by GDF-15. Age 69-75, male 61-67%, sbp 130, DM 11-35%, HF 25-34%, hypertension 78-80%, previous stroke/TIA 20-22%, prior MI 12-21%, prev PAD/MI/CAD 23-38%, aspirin 36-41%. CHADS2 $\geq$3 22-43%. No blinding/loss to FU reported.
Inclusion criteria	Not reported
Exclusion criteria	Not reported
Anticoagulants used	Dabigatran (110 or 150mg twice daily) or adjusted dose warfarin (INR 2-3)
Risk tools used	HAS-BLED ORBIT (with or without GDF-15)
Outcome definition	Major bleeding (2005 ISTH)
Mean follow up	Median 1.9 years

Reference	Hijazi, 2017[52]
time	
Number of bleeding events	458
Results	C statistic major bleeding not differentiated by OAC HAS-BLED: 0.62(0.59-0.64) HAS-BLED with GDF-15: 0.69(0.67-0.72) ORBIT:0.68(0.65-0.70) ORBIT with GDF-15:0.71(0.68-0.73) GDF15 alone: 0.67(0.65-0.69)

Table 60. Hilkens, 2017[58]

Reference	Hilkens, 2017[58]
Study type	Retrospective cohort study
Study sample	3623 patients with AF on warfarin or dabigatran, from the RE-LY trial in Holland. No baseline data available. No report of blinding/loss to FU.
Inclusion criteria	Documented AF in preceding 6 months; history of stroke or TIA
Exclusion criteria	
Anticoagulants used	Warfarin and dabigatran
Risk tools used	HEMORRHAGERS Shireman HAS_BLED ATRIA ORBIT (score) ORBIT (equation)
Outcome definition	Major bleeding, defined as reduction in Hb level of >20 g/L, transfusion of >2 U of blood or symptomatic bleeding in a critical area/organ.
Mean follow up time	2 years

Reference	Hilkens, 2017[58]
Number of bleeding events	266
Results	C statistic for major bleeding on warfarin (n=1195) HEMORRHAGES: 0.58(0.51-0.65) Shireman: 0.57(0.50-0.63) HAS-BLED: 0.57(0.51-0.64) ATRIA: 0.56(0.49-0.63) ORBIT: 0.56(0.48-0.64) C statistic for major bleeding on dabigatran (n=2428) HEMORRHAGES: 0.69(0.64-0.75) Shireman: 0.66(0.61-0.71) HAS-BLED: 0.68(0.63-0.73) ATRIA: 0.74(0.68-0.79) ORBIT: 0.73(0.68-0.78) C statistic for major bleeding on dabigatran or warfarin at 1 year (n=3623) HEMORRHAGES: 0.65(0.61-0.69) Shireman: 0.62(0.58-0.66) HAS-BLED: 0.64(0.60-0.68) ATRIA: 0.67(0.62-0.71) ORBIT: 0.66(0.62-0.71) C statistic for major bleeding on dabigatran or warfarin at 2 years (n=3623) HEMORR2HAGES: 0.63 (0.59-0.66) Shireman: 0.61 (0.57-0.64) HAS-BLED: 0.62 (0.58-0.65) ATRIA: 0.66 (0.62-0.69) ORBIT (score): 0.66 (0.62-0.69) ORBIT (equation): 0.66 (0.62-0.69)

Reference	Hilkens, 2017[58]

Calibration
ORBIT had best calibration at 2 years.

Supplemental Figure II. Calibration of risk scores for major bleeding in patients with a TIA or stroke on oral anticoagulants at two years

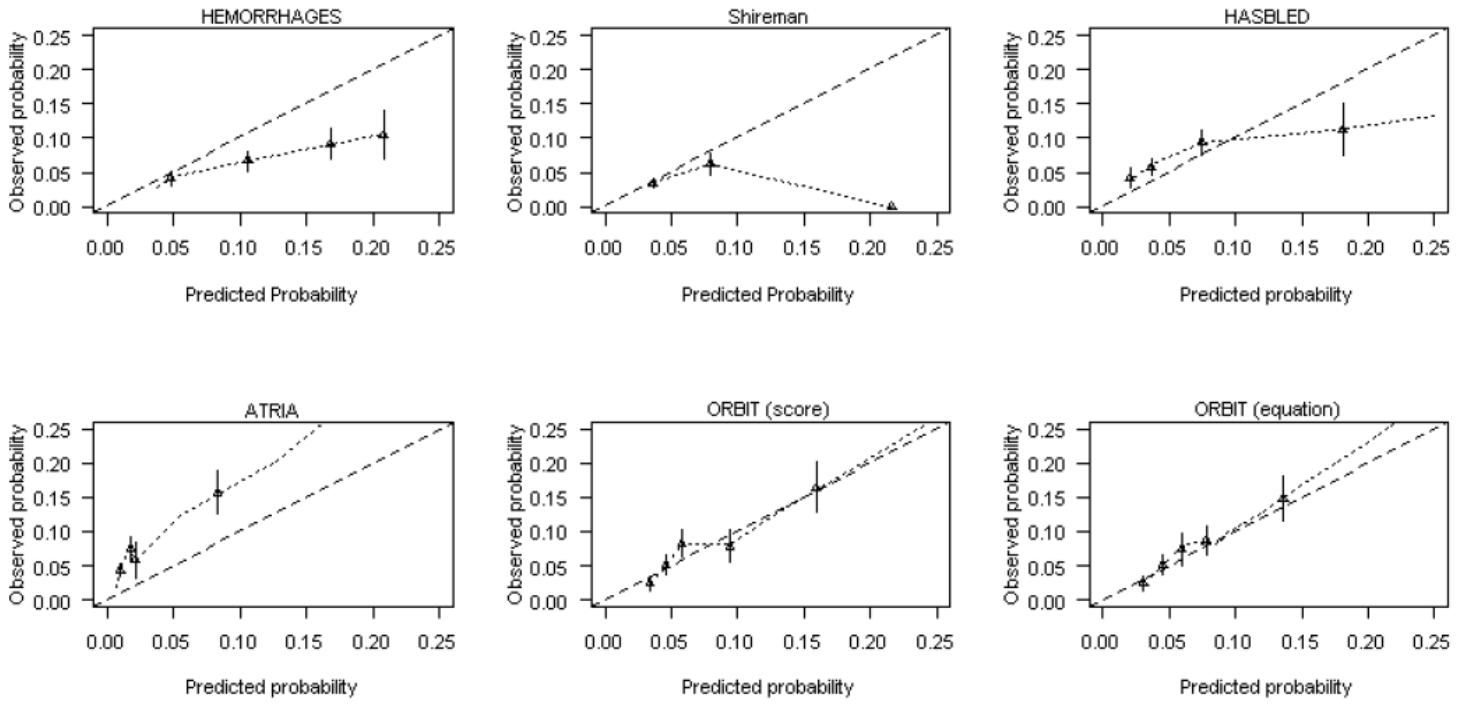


Table 61. Jaspers Focks, 2016[63]

Reference	Jaspers Focks, 2016[63]
Study type	Prospective cohort study
Study sample	1157 AF patients aged >80 years, using a VKA from 2011-2014 in the Netherlands. Median age 84, 42.6% male, 37 months on VKA, 65.8% hypertension, 22% previous stroke/TIA, 9.8% LVEF<40%, 26.6% CAD, 25.7% DM, 21.8% previous bleeding, 5.3% recent or active malignancy, **4.1% on antiplatelets**and 2.1% on NSAIDS. HAS-BLED score 2.23. No blinding reported. 735 completed 3 year follow up (367 patients died and 55 patients moved out of the area or discontinued VKA treatment
Inclusion criteria	NVAF, ≥80 years
Exclusion criteria	Mechanical heart valve problems and/or clinically significant mitral valve stenosis.
Anticoagulants used	VKA
Risk tools used	HAS-BLED ATRIA HEMORRHAGES
Outcome definition	Major bleeding (2005 ICTH) and Clinically relevant bleeding
Mean follow up time	30 months
Number of bleeding events	80 major bleeds in 77 patients
Results	Major bleeding C statistics HAS-BLED: 0.57(0.50-0.63) ATRIA: 0.58(0.51-0.64) HEMORRHAGES: 0.57(0.50-0.63) NRI HAS-BLED v ATRIA: -0.0632 (SE: 0.071) HAS-BLED v HEMORRHAGES: -0.0360 (0.078) HEMORRHAGES v ATRIA: -0.0272 (0.069) Clinically relevant bleeding C statistics HAS-BLED: 0.50(0.47-0.54) ATRIA: 0.52(0.49-0.56)

Reference	Jaspers Focks, 2016[63]
	HEMORRHAGES: 0.53(0.50-0.57) NRI HAS-BLED v ATRIA: -0.0564 (SE: 0.036) HAS-BLED v HEMORRHAGES: -0.0561 (0.043) HEMORRHAGES v ATRIA: -0.0003 (0.039) Any bleeding C statistics HAS-BLED: 0.51(0.47-0.54) ATRIA: 0.53(0.50-0.57) HEMORRHAGES: 0.53(0.50-0.57) NRI HAS-BLED v ATRIA: -0.0851 (SE: 0.033) HAS-BLED v HEMORRHAGES: -0.0372 (0.038) HEMORRHAGES v ATRIA: -0.0479 (0.035) Calibration The calibration of all models was reported as 'adequate' (Hosmer-Lemeshow goodness of fit significance level >0.05

Table 62. Jover, 2012[65]

Reference	Jover, 2012[65]
Study type	Prospective cohort study
Study sample	933 patients with permanent or paroxysmal NVAF on acenocoumarol OAC (INR 2-3) for at least 6 months. Age 76, 46% male, 85% hypertension, 27% DM, 32% hypercholesterolemia, 14% current smokers, 39% CHF, 20% prior stroke/TIA, 20% CAD, 9% PAD, 17% on antiplatelets. CHADS2 score 2, CHADSVASC score 4. No blinding reported. No loss to FU reported.
Inclusion criteria	CHADSVASC ≥2; age >18
Exclusion criteria	Haematologic disorder or contraindications to OACs in past 6 months, ischaemic events requiring hospitalisation in previous 6 months, rheumatic AF, prosthetic heart valves.

Reference	Jover, 2012[65]
Anticoagulants used	Acenocoumarol
Risk tools used	CHADSVASC
Outcome definition	Major bleeding (2005 ISTH)
Mean follow up time	Median 2.5 years
Number of bleeding events	80 patients with major bleeding
Results	C statistic major bleeding CHADSVASC: 0.54(0.48-0.61)

Table 63. Lip, 2011[71]

Reference	Lip, 2011[71]
Study type	Retrospective cohort study
Study sample	7,329 people with NVAF on warfarin or ximelagatran. Taken from the SPORTIF III and V cohorts (Multinational cohort). Following data are for those who developed a major bleed/no major bleed: age 73.9/70.9, female 31/31%, paroxysmal AF 11/12%, hypertension 77/77%, DM 29/23%, CAD 50/45%, LV dysfunction 44/36%, stroke/TIA 26/21%, CHADS 2.6/2.2.**Blinded assessors**.
Inclusion criteria	>18 years, persistent or paroxysmal AF, NVAF, on warfarin or ximelagatran; at least one of the following stroke risk factors: hypertension, age 75 or older, previous stroke/TE, LV dysfunction, age >65 with CAD, age >65 with DM
Exclusion criteria	Not reported
Anticoagulants used	Warfarin or ximelagatran
Risk tools used	HAS-BLED Shireman HEMORRHAGE Beyth et al. Kuijer et al.

Reference	Lip, 2011[71]
Outcome definition	Major bleeding (2005 ICTH) [BLINDED by central adjudication committee].
Mean follow up time	499 days
Number of bleeding events	136 people had major bleeding
Results	C statistics for major bleeding in warfarin patients (n=3665) HAS-BLED: 0.66(0.61-0.70) Shireman: 0.63(0.58-0.67) HEMORRHAGE: 0.61(0.56-0.65) Beyth et al. : 0.56(0.51-0.60) Kuijer et al.: 0.52(0.48-0.56) C statistics for major bleeding in warfarin AND ximelagatran patients (n=7329) HAS-BLED: 0.65(0.61-0.68) Shireman: 0.64(0.61-0.68) HEMORRHAGE: 0.62(0.58-0.65) Beyth et al. : 0.57(0.53-0.60) Kuijer et al.: 0.49(0.46-0.52) Sensitivity/specificity HAS-BLED (n=3665) ≥1: 0.948/0.209 ≥2: 0.625/0.560 ≥3: 0.338/0.8186 Calibration Hosmer-Lemeshow showed all tools had adequate calibration (all p>0.05).

Table 64. Lip, 2014[74]

Reference	Lip, 2014[74]
Study type	Retrospective cohort study
Study sample	4,637 patients with AF (n=572 had valvular AF) who were receiving OACs. FRANCE. Mean age 71, 35% female, 60% HF, 28% CAD, 12% previous MI, 6% previous CABG, 44% hypertensive, 9% previous stroke, 9% renal insufficiency. **17% on antiplatelets, 15% on Aspirin, 6% clopidogrel, 4% DAT.**Mean CHADSVASC score 3.2, Mean HAS-BLED score 1.6.. Not blinded.
Inclusion criteria	Patients given a diagnosis of NVAF or atrial flutter between 2000 and 2010 at Cardiology department in France.
Exclusion criteria	For this analysis, those not on OACs
Anticoagulants used	VKAs
Risk tools used	SAMe-TT2R2 score
Outcome definition	Severe bleeding – defined as decrease in blood Hb level of >5 g/dL, or the need for transfusion of 2 or more units of blood, or the need for corrective surgery, or the occurrence of an IC or retroperitoneal haemorrhage. Major bleeding – defined using BARC definition: IC haemorrhage, intraocular bleeding compromising vision, overt bleeding plus Hb drop of >5 g/dL, tamponade, bleeding requiring surgical or percutaneous control or inotropes, or any transfusion with overt bleeding, fatal bleeding. Both identified by hospital ICD coding.
Mean follow up time	1016 days (2.78 years).
Number of bleeding events	480 developed severe bleeding, of whom 144 had major (BARC) bleeding.
Results	Harrel C statistic for severe bleeding SAMe-TT2R2 score (cont): 0.552 (0.537 to 0.566) SAMe-TT2R2 score (3 cats – low 0-1, mod 2, high >2): 0.552 (0.538 to 0.566) SAMe-TT2R2 score (2 cats – low 0-2, high >2): 0.552 (0.538 to 0.567) Harrel C statistic for major bleeding SAMe-TT2R2 score (cont): 0.574 (0.560 to 0.589) SAMe-TT2R2 score (3 cats – low 0-1, mod 2, high >2): 0.576 (0.561 to 0.590) SAMe-TT2R2 score (2 cats – low 0-2, high >2): 0.571 (0.557 to 0.586)

Table 65. Lip, 2018[77]

Reference	Lip, 2018[77]
Study type	Retrospective cohort study
Study sample	57,930 patients with NVAF on DOACs. Taken from 3 Danish nationwide databases. Age 73.5, female 44.6%, HF 22.5%, DM 15.2%, Vascular diseases 16.2%, hypertension 59%, CPD 13.3%, prior bleeding 14.2%, kidney diseases 3.4%, **Aspirin use 39.1%, NSAIDs 22.4%.** Not blinded. Loss to FU not reported.
Inclusion criteria	OAC naïve at baseline; NVAF.
Exclusion criteria	Prior exposure to any OAC inclusive doses within 1 year; valvular AF; venous thromboembolism.
Anticoagulants used	DOACs
Risk tools used	HAS-BLED ATRIA ORBIT
Outcome definition	Combined bleeding endpoint: IC, GI, traumatic IC, and clinically relevant non-major bleeding.
Mean follow up time	1 year **(2.5 year data available in online supplement but no access possible**).
Number of bleeding events	2.41 / 100 person-years
Results	C statistics HAS-BLED: 0.58(0.57-0.59) ATRIA: 0.59(0.57-0.60) ORBIT: 0.61(0.59-0.62) Sensitivity and specificity [%] HAS-BLED: ≥3: 62.8 and 53.5 ATRIA: ≥4: 29.7 and 87.6 ORBIT: ≥3: 31.1 and 84.0 Sensitivity and specificity [%] (at intermediate/high threshold – actual thresholds not described) HAS-BLED: - ATRIA: 17.9 and 93.1

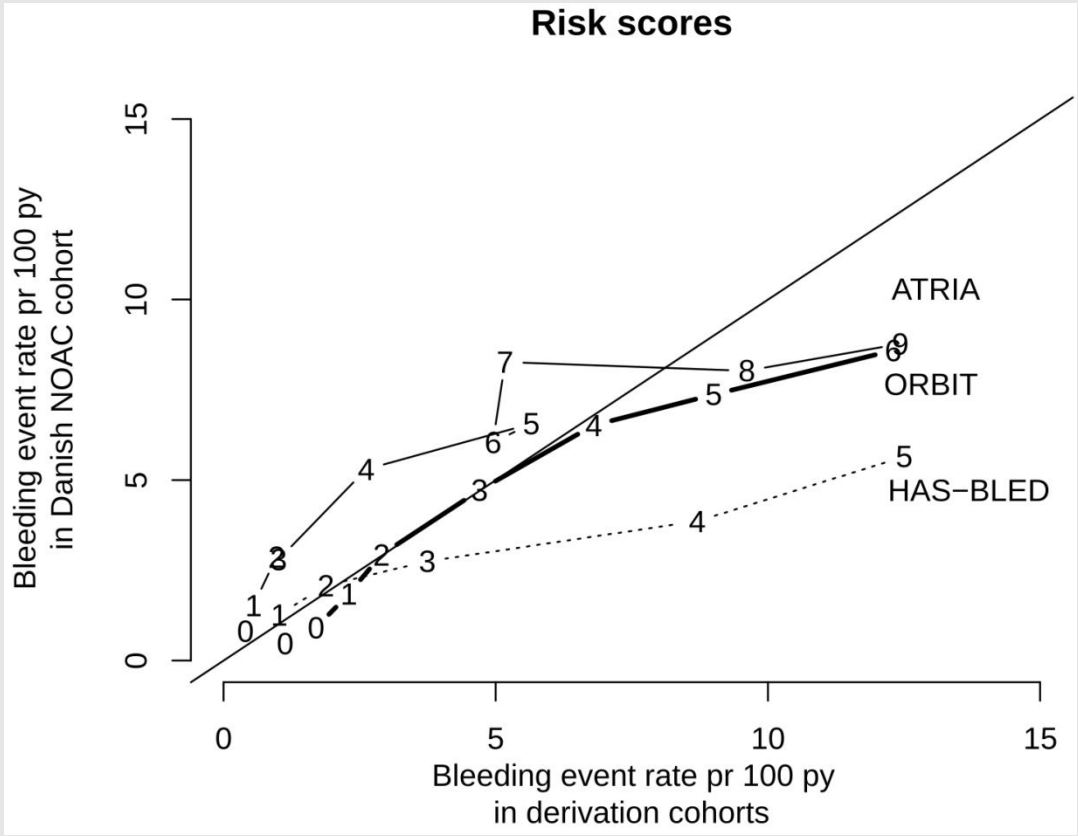
Reference	Lip, 2018[77]
	ORBIT: 22.5 and 91.8 Calibration. Orbit was the best calibrated, especially at the lowest scores 

Table 66. Mori, 2019[88]

Reference	Mori, 2019[88]
Study type	Prospective cohort study
Study sample	2216 patients with NVAF using DOACs; 63.6% male; median age 73 years; median CHADS2 2; hypertension 73.5%; DM 27.9%; Dyslipidaemia 65.2%; eGFR 64.9; CAD 19.8%; PAD 7.1%; HF 23.7%; prior stroke 20.2%; prior bleeding 27.1%; antiplatelets 21.5%
Inclusion criteria	All people with NVAF using dabigatran, rivaroxaban, edoxaban and apixaban
Exclusion criteria	None reported
Anticoagulants used	DOACs
Risk tools used	ORBIT HAS-BLED
Outcome definition	Major bleedingas defined by ISTH
Mean follow up time	315 days
Number of bleeding events	Incidence 4.2% (93)
Results	**C statistics** ORBIT 0.64(0.59-0.70) HAS-BLED 0.62(0.57-0.68) **Calibration** Calibration plots of the ORBIT bleeding score showed a similar predictive performance compared with the HAS-BLED score [slope 0.91(0.4 to 1.43)vs 0.71(-2.35 to 3.76)and intercept 0.24 (-2.13 to 2.61) vs 0.71(-2.35 to 3.76)]

Table 67. Nielsen, 2016[90]

Reference	Nielsen, 2016[90]
Study type	Retrospective cohort study
Study sample	Unknown number of OAC-treated patients from a cohort of 210,299 patients with AF taken from 3 Danish patient registries from 1999 to 2013. Demographic data for the sub-group having OACs is not reported
Inclusion criteria	AF
Exclusion criteria	Bleeding event within 7 days after discharge
Anticoagulants used	Unclear
Risk tools used	HAS-BLED Recalibrated HAS-BLED (2 points for previous haemorrhagic stroke instead of 1 point)
Outcome definition	Major bleeding
Mean follow up time	Unclear
Number of bleeding events	4.73 (per 100 person-years)
Results	NRI Recalibrated HAS-BLED v HAS-BLED: +0.09 (+0.048 to +0.123) C statistics Reported to be similar to C statistics in whole cohort, but data not shown. Data for whole cohort were 0.613 for original HAS-BLED and 0.616 for recalibrated HAS-BLED.

Table 68. O'Brien, 2015[91]

Reference	O'Brien, 2015[91]
Study type	Retrospective cohort study
Study sample	14,264 patients with AF on either rivaroxaban (20mg daily) or Warfarin. This was the external validation cohort, comprising patients from the ROCKET-AF. Demographics of this external validation sample not reported.
Inclusion criteria	Not reported
Exclusion criteria	Not reported
Anticoagulants used	Rivaroxaban and warfarin
Risk tools used	ORBIT HAS-BLED ATRIA-bleeding
Outcome definition	Major bleeds
Mean follow up time	1.9 years
Number of bleeding events	772 major bleeds
Results	C statistics ORBIT (cont): 0.63(0.61-0.65) ORBIT (cat): 0.62(0.60-0.64) HAS-BLED: 0.59(0.57-0.61) ATRIA: 0.60(0.58-0.62) Sensitivity and specificity data contained within a table, but this is for the derivation cohort, which is not presented in this review. Calibration The ORBIT score displayed superior calibration compared with the other 2 scores, followed by HAS-BLED (worst at low risk strata) and ATRIA (not good for most risk groups).

Reference	O'Brien, 2015[91]

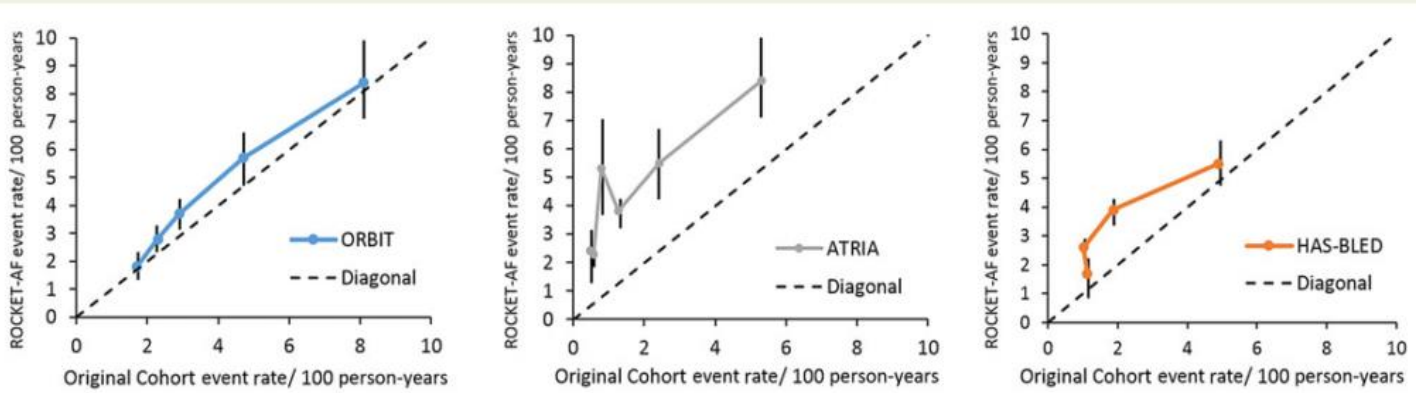


Figure 1 Calibration plot of outcomes registry for better informed treatment, hypertension, abnormal renal/liver function, stroke, bleeding history or predisposition, labile INR, elderly, drugs/alcohol concomitantly, and anticoagulation and risk factors in atrial fibrillation in the rivaroxaban once-daily oral direct factor Xa inhibition compared with vitamin K antagonism for prevention of stroke and embolism trial in atrial fibrillation external validation cohort. This figure displays the major bleeding events rates per 100 patient-years and 95% confidence intervals observed in the external validation rivaroxaban once-daily oral direct factor Xa inhibition compared with vitamin K antagonism for prevention of stroke and embolism trial in atrial fibrillation cohort vs. those previously published from the original derivation cohorts for each discrete score point value. The highest risk categories for each score were combined to promote stable estimates as follows: outcomes registry for better informed treatment (0, 1, 2, 3, ≥4), anticoagulation and risk factors in atrial fibrillation (0, 1, 2, 3, ≥4), and hypertension, abnormal renal/liver function, stroke, bleeding history or predisposition, labile INR, elderly, drugs/alcohol concomitantly (0, 1, 2, ≥3). ORBIT-AF; Outcomes Registry for Better Informed Treatment of Atrial Fibrillation; 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; ATRIA, anticoagulation and risk factors in atrial fibrillation.

Table 69. Olesen, 2011[95]

Reference	Olesen, 2011[95]
Study type	Retrospective cohort study
Study sample	44, 771 patients with AF receiving OACs in Denmark during 1997-2006. Demographic data given as two values as separate data for those with major bleeding / those without. Age 74.6 / 71.2, male 66.8 / 61.2 %, HASBLED score 2.5-2, HF 24.4/19.8%, hypertension 51.6/49.5%, DM 11.4/9.5%, Stroke 22.3/17.4, Renal disease 8.2/4.6%, Vascular disease 18.6/14.8%, Bleeding history 22.6/8.2%, **antiplatelet drugs 33% / 25.5%,**NSAIDs 22.8/19.1%.
Inclusion criteria	On OACS and with NVAF
Exclusion criteria	Death or events within 7 days of any hospitalisation (as medication may be changed after hospitalisation)
Anticoagulants used	44,671 on VKAs and 100 on Heparins
Risk tools used	HAS-BLED HEMORRHAGES
Outcome definition	Hospitalisation or death from major bleeding, including GI bleeding, IC bleeding, bleeding from the
Mean follow up time	1 year
Number of bleeding events	2051 events
Results	C statistics HAS-BLED (cont):0.795(0.759-0.829) HAS-BLED (cat): 0.795 (0.759-0.829) HEMORRHAGES (cont): 0.771(0.733-0.806) HEMORRHAGES (cat): 0.782(0.745-0.816) Derived from Table 2 in paper At threshold of >low risk for HASBLED (≥2) Sen 81.6% Spec 64.43% At threshold of >low risk for HEMORRHAGES (≥2) Sen 71.1% Spec 48.2%

Table 70. Pisters, 2010[103]

Reference	Pisters, 2010[103]
Study type	Retrospective cohort study
Study sample	1956 patients on OACs only with NVAF (validation cohort). Data not given for this validation cohort subset. None on antiplatelets/NSAIDS.
Inclusion criteria	>18 years with a Halter-proven diagnosis of AF, enrolled from the Euro Heart Survey, with data collected between 2003 and 2004.
Exclusion criteria	None reported
Anticoagulants used	OACs (not specified)
Risk tools used	HAS-BLED HEMORRHAGES
Outcome definition	Major bleeding (2005 ISTH)
Mean follow up time	1 year
Number of bleeding events	1.75 bleeds/100 patient-years
Results	C statistics HAS-BLED: 0.69(0.59-0.80) HEMORRHAGES: 0.64(0.53-0.75)

Table 71. Poli, 2017[110]

Reference	Poli, 2017[110]
Study type	Prospective cohort study
Study sample	4579 patients with AF on DOACS (n=1048) or VKAs (n=3531) on START register in Italy. Age 76, 55% men, 15% HF, 80% hypertensive, 20% DM, 18% CAD, 6% PAD, 43% moderate renal impairment (eGFR 30-60 ml/min), 15% previous stroke/TIA, 3.4% history of major bleeding, TTR 67, **concomitant antiplatelet drugs 16.5%, dual antiplatelet therapy 1.3%.**
Inclusion criteria	Not reported

Reference	Poli, 2017[110]
Exclusion criteria	Not reported
Anticoagulants used	Warfarin and DOACS
Risk tools used	HAS-BLED HAS-BED (HAS-BLED but without labile INR score) CHADS2 CHADSVASC
Outcome definition	Major bleeding – as defined by International Society of Thrombosis and Haemostasis
Mean follow up time	1.4 years
Number of bleeding events	115 patients experienced a MB event (13 fatal)
Results	Not sub-grouped to OAC HAS-BLED (cont): 0.61(0.560-0.667) HAS-BED (cont): 0.58(0.530-0.639) CHADS2 (cont): 0.58(0.531-0.638) CHADSVASC (cont): 0.56(0.509-0.618) HAS-BLED (cat): 0.59(0.539-0.643) HAS-BED (cat): 0.52(0.468-0.579) CHADS2 (cat): 0.54 (0.494-0.596) CHADSVASC (cat): 0.51(0.455-0.561) Sensitivity/specificity HAS-BLED ≥3: 0.609/0.408 HAS-BED ≥3: 0.504/0.659 CHADS2 ≥3:0.747/0.074 CHADSVASC ≥3: 0.930/0.0878

Table 72. Prochaska, 2018[113]

Reference	Prochaska, 2018[113]
Study type	Prospective cohort study
Study sample	1089 patients with medical and electrophysiological evidence of AF, and on VKAs, as part of the thrombEVAL cohort. Denmark. The following baseline data is separated into paroxysmal (n=398) and sustained (n=691) sub-groups by the paper: male 63/63%, age 72/75, DM 30/33%, Family history of MI/stroke 44.5/42%, hypertension 83/81.6%, CKD 24/27%, CAD 43.6/46.7%, HF 43.5/55.2%, history of major bleeding 6.8/6.2%, history of stroke/TIA 16.7/18.7%, MI 21.8/20.8%, PAD 16.1/17.5%, **aspirin 18.3/15.1**
Inclusion criteria	Not reported
Exclusion criteria	Not reported
Anticoagulants used	VKA - phenprocoumon
Risk tools used	HAS-BLED HAS-BLED with a point for sustained AF Simplified HAS-BLED
Outcome definition	Clinically relevant bleeding – composite of major bleeding and clinically relevant non-major bleeding.
Mean follow up time	3 years
Number of bleeding events	150people with bleeding events
Results	C statistics HAS-BLED:0.583(0.54-0.63) HAS-BLED with a point for sustained AF: 0.606(0.57-0.65) Simplified HAS-BLED: 0.642(0.60-0.68)

Table 73. Proietti, 2016[116]

Reference	Proietti, 2016[116]

Reference	Proietti, 2016[116]
Study type	Retrospective cohort study
Study sample	3551 patients receiving warfarin in the pooled population dataset from the SPORTIF III and V studies with AF. De-identified datasets with patient-level information for the SPORTIF trials were obtained directly from Astra Zeneca, and all the analyses were performed independent of the company. All patients assigned to the warfarin treatment arms and with available data for the clinical variables used to calculate the four bleeding prediction scores were included in the present analysis. The majority of patients were male (69.5%) and the median [IQR] age was 72 [66–77] years. HAS-BLED score ≥3: 71%. 706/3551 (19.9%) treated concomitantly with aspirin. 20.1% VKA naïve at baseline prior to VKA initiation.
Inclusion criteria	Not reported
Exclusion criteria	Not reported
Anticoagulants used	Warfarin
Risk tools used	HAS-BLED ORBIT ATRIA HEMORRAGES ORBIT with TTR <65% (adding one point to score if <65%) ATRIA with TTR <65% (adding one point to score if <65%) HEMORRAGES with TTR <65% (adding one point to score if <65%)
Outcome definition	'major bleeding' events were defined in two distinct ways, as follows: (i) "investigator level" events (that included the crude number of all the major bleeding events reported by any investigator at every study site); and (ii) "adjudicated events" (corresponding to the final trial adjudicated major bleeding events, after the independent central adjudication committee evaluated all the reported events).
Mean follow up time	1.6 years
Number of bleeding events	162 investigator level events (of which 127 were confirmed as 'adjudicated')
Results	**C statistic** HAS-BLED:0.581 (0.564-0.597) ORBIT: 0.589 (0.573-0.606)

Reference	Proietti, 2016[116]
	ATRIA: 0.590 (0.574-0.606) HEMORR2HAGES: 0.549 (0.532-0.565) ORBIT with TTR <65%: 0.609 ATRIA with TTR <65%: 0.611 HEMORRAGES with TTR <65%: 0.578 Head to head: HEMORRHAGES significantly worse than HAS-BLED (p=0.039), ORBIT (p=0.006) and ATRIA (p=0.003). Other comparisons NS. Sensitivity/specificity (based on somewhat approximate data as calculated from data containing rounded percentages) HAS-BLED ≥1: 0.992/0.007 ≥2:0.968/0.068 ≥3:0.787/0.289 ≥4:0.543/0.5867 ATRIA ≥1: 0.937/0.007 ≥2:0.874/0.615 ≥3:0.700/0.739 ≥4:0.346/0.985 ORBIT ≥1: 0.700/0.432 ≥2:0.417/0.722 ≥3:0.126/0.959 HEMORRHAGES ≥1: 0.953/0.091 ≥2:0.480/0.582 ≥3:0.173/0.912 **NRI**

Reference	Proietti, 2016[116]
	Orbit v HAS-BLED: -0.0077 Atria v HAS-BLED: -0.0883 Haemorrhages v HAS-BLED: -0.1366 Atria v ORBIT: 0.0355 Haemorrhagesv ORBIT: -0.2164 Haemorrhagesv ATRIA: -0.3128 ORBIT with TTR <65% v ORBIT: 0.2508 ATRIA with TTR <65% v ATRIA: 0.250 Haemorrhageswith TTR <65% v haemorrhages: 0.263

Table 74. Proietti, 2018[114]

Reference	Proietti, 2018[114]
Study type	Retrospective cohort study
Study sample	18,113 patients with AF on dabigatran (110 or 150 mg) or warfarin in the RE-LY trial. Multinational cohort. Age 72, 36% female, 79% hypertension, DM 23%, CAD 28%, prev stroke 22%, symptomatic HF 27%, VKA naïve 50%, **anti-platelets 40%**, CHADS2 2. BLINDED ASSESSORS.
Inclusion criteria	Not reported
Exclusion criteria	Not reported
Anticoagulants used	Dabigatran and warfarin
Risk tools used	HAS-BLED ORBIT ATRIA HEMORRHAGES
Outcome definition	Major bleeding (2005 ICTH) Life-threatening bleeding (sub-category of MB comprising fatal bleeding OR symptomatic IC bleedingOR bleeding with decrease in

Reference	Proietti, 2018[114]
	Hb of at least 50 g/L, or bleeding requiring transfusion of at least 4 units of blood/inotropic agents/surgery. IC bleeding All centrally adjudicated
Mean follow up time	Median 2 years
Number of bleeding events	1182 major bleeding events (including 555 life-threatening bleeds, which also included 157 IC bleeds)
Results	C statistics major bleeding ALL HAS-BLED:0.62(0.60-0.63) ORBIT:0.66(0.65-0.68) ATRIA:0.64(0.62-0.65) HEMORRHAGES:0.62(0.61-0.64) C statistics major bleeding dabigatran 110mg HAS-BLED:0.61(0.58-0.64) ORBIT:0.68(0.65-0.71) ATRIA:0.64(0.61-0.67) HEMORRHAGES:0.61(0.58-0.64) C statistics major bleeding dabigatran 150mg HAS-BLED:0.64(0.62-0.67) ORBIT:0.70(0.68-0.73) ATRIA:0.67(0.65-0.70) HEMORRHAGES:0.66(0.64-0.69) C statistics major bleeding warfarin HAS-BLED:0.59(0.57-0.62) ORBIT:0.62(0.59-0.64) ATRIA:0.59(0.57-0.62) HEMORRHAGES:0.59(0.56-0.62)

Reference	Proietti, 2018[114]
	C statistics life-threatening bleeding ALL HAS-BLED:0.61(0.59-0.64) ORBIT:0.66(0.64-0.68) ATRIA:0.63(0.61-0.66) HEMORRHAGES:0.62(0.60-0.64) C statistics life-threatening bleeding dabigatran 110mg HAS-BLED:0.60(0.56-0.64) ORBIT:0.67(0.63-0.71) ATRIA:0.63(0.58-0.67) HEMORRHAGES:0.61(0.57-0.66) C statistics life-threatening bleeding dabigatran 150mg HAS-BLED:0.65(0.61-0.69) ORBIT:0.71(0.68-0.75) ATRIA:0.68(0.64-0.72) HEMORRHAGES:0.66(0.63-0.70) C statistics life-threatening bleeding warfarin HAS-BLED:0.59(0.55-0.63) ORBIT:0.62(0.58-0.65) ATRIA:0.59(0.56-0.63) HEMORRHAGES:0.59(0.56-0.62) C statistics intracranial bleeding ALL HAS-BLED:0.56(0.52-0.61) ORBIT:0.62(0.57-0.66) ATRIA:0.58(0.54-0.63) HEMORRHAGES:0.59(0.55-0.64) C statistics intracranial bleeding dabigatran 110mg

Reference	Proietti, 2018[114]
	HAS-BLED:0.52(0.42-0.63) ORBIT:0.63(0.55-0.72) ATRIA:0.59(0.50-0.69) HEMORRHAGES:0.54(0.44-0.65) C statistics intracranial bleeding dabigatran 150mg HAS-BLED:0.56(0.48-0.64) ORBIT:0.60(0.50-0.69) ATRIA:0.59(0.50-0.68) HEMORRHAGES:0.61(0.52-0.70) C statistics intracranial bleeding warfarin HAS-BLED:0.57(0.52-0.63) ORBIT:0.62(0.57-0.67) ATRIA:0.58(0.52-0.63) HEMORRHAGES:0.60(0.55-0.66) Head to head ORBIT was significantly better than HAS-BLED in terms of C statistic for MB, LTB and IH. ATRIA was better than HAS-BLED for MB. No other sig differences with HAS-BLED. Sensitivity/specificity for MB (ALL, across OACs) HAS-BLED ≥2:0.298/0.819 ORBIT ≥3: 0.403/0.798 ATRIA ≥4:0.172/0.932 HEMORRHAGES ≥2: 0.446/0.932

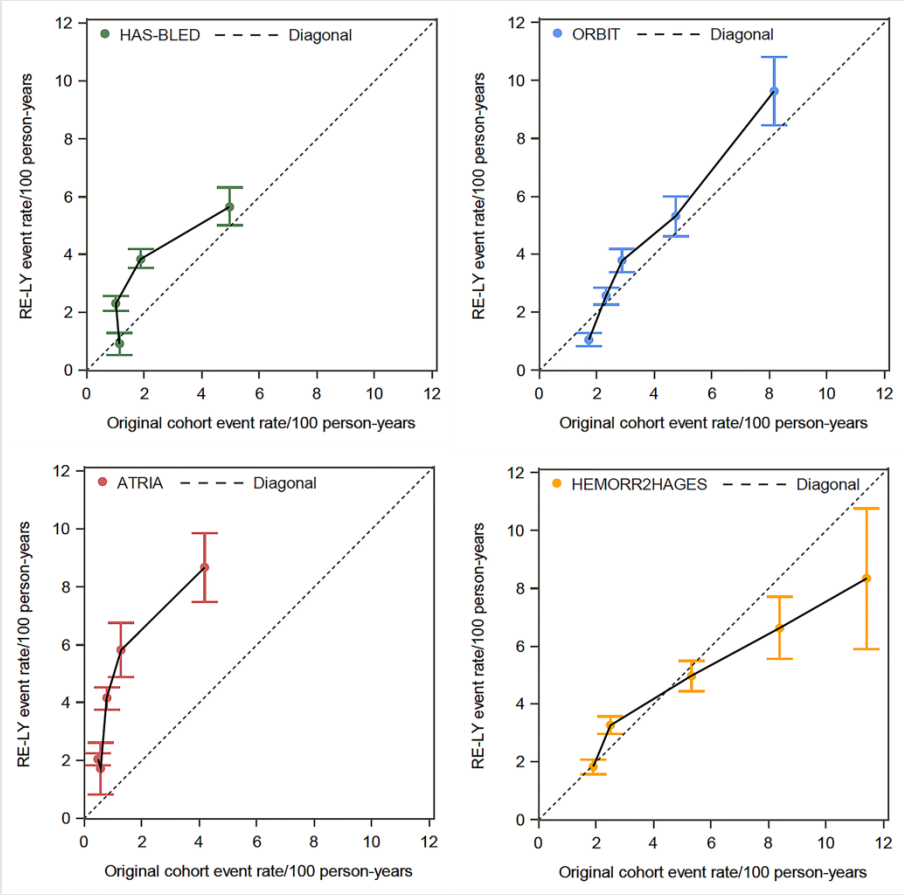
Reference	Proietti, 2018[114]
	Calibration (ALL) ORBIT score had best agreement between predicted and observed risks. ATRIA had worst agreement. ATRIA and HAS-BLED tended to overestimate the risk of bleeding. HEMORRHAGES tended to underestimate bleeding risk. 

Table 75. Proietti, 2018[115]

Reference	Proietti, 2018[115]
Study type	Retrospective cohort study
Study sample	3550 AF patients enrolled on the SPORTIF III trial who were on Warfarin. Age 72, 30.5% female, 76.7% hypertension, 23.5% DM, 44.3% CAD, 20.6% stroke/TIA, 37.3% HF, 5.6% previous bleeding, 25.9% CKD, **19.9% aspirin use**. TTR 68.1. HAS-BLED: 3. 804 patients interrupted Warfarin during the follow up period. BLINDED ASSESSORS.
Inclusion criteria	Not reported
Exclusion criteria	Not reported
Anticoagulants used	Warfarin
Risk tools used	HAS-BLED GARFIELD
Outcome definition	Major bleeding (2005 ICTH) with blinded adjudication by a committee Major/CRNM bleeding Any bleeding
Mean follow up time	1.56 years
Number of bleeding events	127 major bleeds, 168 major/CRNM bleeds, 1450 any bleeds
Results	C statistics Major bleeding HAS-BLED: 0.58(0.56-0.60) GARFIELD: 0.56(0.54-0.57) Major/CRNM bleeding HAS-BLED: 0.56(0.54-0.58) GARFIELD: 0.57(0.55-0.58) Any bleeding HAS-BLED: 0.55(0.53-0.57)

Reference	Proietti, 2018[115]
	GARFIELD: 0.51(0.49-0.53) Head to head GARFIELD significantly better than HAS-BLED for ANY BLEEDING, but NS difference for MB and Major/CRNM bleeding NRI (GARFIELD v HAS-BLED) Major bleeding: -0.042(-0.189 to 0.087) Major/CRNM bleeding: +0.033(-0.094 to 0.129) Any bleeding: -0.087 (-0.131 to -0.056) For those completing Warfarin treatment throughout follow up (n=2746) Major bleeding HAS-BLED: 0.60(0.53-0.68) GARFIELD: 0.55(0.47-0.63) Major/CRNM bleeding HAS-BLED: 0.59(0.53-0.66) GARFIELD: 0.57(0.50-0.65) Any bleeding HAS-BLED: 0.56(0.54-0.58) GARFIELD: 0.50(0.48-0.53) Head to head: again, for ANY BLEEDING, Garfield was sig better.

Table 76. Quinn, 2016[117]

Reference	Quinn, 2016[117]

Reference	Quinn, 2016[117]
Study type	Retrospective cohort study
Study sample	13,559 patients with AF who were on and off warfarin. No demographic data provided.
Inclusion criteria	Serial outpatient diagnoses of AF.
Exclusion criteria	None reported
Anticoagulants used	Warfarin
Risk tools used	CHADS2 CHADSVASC ATRIA HAS-BLED
Outcome definition	Major haemorrhage (ICTH 2005)
Mean follow up time	Unclear
Number of bleeding events	Unclear
Results	C statistics (3 category score) CHADS: 0.63(0.61-0.65) CHADSVASC 0.56(0.55-0.57) ATRIA bleeding: 0.68(0.66-0.71) HAS-BLED: 0.61(0.59-0.63) C statistics (continuous score) CHADS: 0.65(0.62-0.67) CHADSVASC 0.65(0.62-0.67) ATRIA bleeding: 0.74(0.72-0.76) HAS-BLED: 0.64(0.61-0.66) NRI (all vs CHADS) CHADSVASC: -0.129 ATRIA bleeding: +0.28 HAS-BLED: +0.004

Table 77. Rivera-Caravaca, 2017[120]

Reference	Rivera-Caravaca, 2017[120]
Study type	Retrospective cohort study
Study sample	1361 patients**– same patients as Roldan 2017[128]**- with AF who were taking VKA OACs (acenocoumarol), in Spain. Age 76, 49% male, 82% hypertensive, 27% DM, 19% previous stroke/TIA, 19% CAD, 31% HF, 7% PAD, 10% renal impairment, 33% hypercholesterolemia, 8% previous bleeding episode, 4% alcohol abuse, 1% hepatic disease, 8% cancer. Median HAS-BLED score of 2
Inclusion criteria	Permanent or paroxysmal AF (not defined) who were taking VKA OACs. All had to have good anticoagulation control with INR between 2 and 3 during the past 6 months of clinic visits
Exclusion criteria	Prosthetic heart valves, rheumatic AF, acute coronary syndrome, stroke, potentially unstable chest pain or any haemodynamic instability, as well as patients who had hospital admission or surgical intervention in past 6 months.
Anticoagulants used	VKAs
Risk tools used	HEMORRHAGES HAS-BLED ATRIA ORBIT
Outcome definition	Major bleeding events – based on the 2005 International Society on Thrombosis and Haemostasis criteria, which were: fatal or symptomatic bleeding in a critical area or organ, such as IC, intraspinal, intraocular, retroperitoneal, intra-articular or pericardial, or IM with compartment syndrome causing a fall in Hb of 20 or more g/L.
Mean follow up time	6.5 years
Number of bleeding events	250 (2.83% per year)
Results	**C statistics for Major Bleeding** HAS-BLED: 0.625 (0.599-0.651) ATRIA 0.545 (0.518-0.572) ORBIT 0.565 (0.538-0.591) HEMORR2HAGES 0.547 (0.520-0.573) ATRIA with TTR <65% 0.751 (0.727-0.774)

Reference	Rivera-Caravaca, 2017[120]
	ORBIT with TTR <65% 0.733 (0.709-0.757) HEMORR2HAGES with TTR <65% 0.729 (0.704-0.752) Sensitivity/specificity HAS-BLED ≥3: 0.652/0.598 ATRIA ≥4: 0.296/0.795 ORBIT ≥3:0.34/0.789 HEMORRHAGES ≥2:0.824/0.269 **NRI** ATRIA with TTR <65% versus ATRIA: +0.1527, p<0.001 ORBIT with TTR <65% versus ORBIT: +0.1097, p<0.001 HEAMORRHAGES with TTR <65% versus HEMORRHAGES: +0.0598, p=0.007

Table 78. Rivera-Caravaca, 2019[119]

Reference	Rivera-Caravaca, 2019[119]
Study type	Prospective cohort study
Study sample	940 patientswho were taking VKA OACs (IRR 2-3), in Spain. Age 76, 50.6% male, 82% hypertensive, 26.2% DM, 18.8% previous stroke/TIA, 19.8% CAD, 30.4% HF, 10.6% renal impairment, 33.3% hypercholesterolemia, Median HAS-BLED score of 2
Inclusion criteria	Permanent or paroxysmal AF (not defined) who were taking VKA OACsfor at least 6 months. All had to have good anticoagulation control with INR between 2 and 3 during the past 6 months of clinic visits
Exclusion criteria	Prosthetic heart valves, rheumatic AF, acute coronary syndrome, stroke, potentially unstable chest pain or any haemodynamic

Reference	Rivera-Caravaca, 2019[119]
	instability, as well as patients who had hospital admission or surgical intervention in past 6 months.
Anticoagulants used	VKAs
Risk tools used	HAS-BLED HAS-BLED + VWF HAS-BLED + VWF + NT-proBNP HAS-BLED + VWF + NT-proBNP + IL-6 HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T + BTP HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T + BTP + soluble fibrin monomer complex
Outcome definition	Major bleeding events – based on the 2005 International Society on Thrombosis and Haemostasis criteria, which were: fatal or symptomatic bleeding in a critical area or organ, such as IC, intraspinal, intraocular, retroperitoneal, intra-articular or pericardial, or IM with compartment syndrome causing a fall in Hb of 20 or more g/L.
Mean follow up time	6.5 years
Number of bleeding events	172 major bleeding
Results	**C statistics** HAS-BLED 0.600: (0.561-0.625) HAS-BLED + VWF: 0.636(0.605-0.667) HAS-BLED + VWF + NT-proBNP: 0.639 (0.607-0.669) HAS-BLED + VWF + NT-proBNP + IL-6: 0.639 (0.607-0.669) HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T: 0.638 (0.606-0.669) HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T + BTP: 0.635 (0.604-0.666) HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T + BTP + soluble fibrin monomer complex: 0.635 (0.604-0.666) **NRI(versus HAS-BLED alone)** HAS-BLED + VWF: 0.226(0.038-0.326) HAS-BLED + VWF + NT-proBNP: 0.201(0.002-0.329) HAS-BLED + VWF + NT-proBNP + IL-6: 0.192(0.014-0.325)

Reference	Rivera-Caravaca, 2019[119]
	HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T: 0.194(0.030-0.337) HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T + BTP: 0.196(0.048-0.327) HAS-BLED + VWF + NT-proBNP + IL-6 + Troponin T + BTP + soluble fibrin monomer complex: 0.203(0.004-0.342)

Table 79. Roldan, 2013[125]

Reference	Roldan, 2013[125]
Study type	Prospective cohort study
Study sample	937 consecutive patients with AF receiving anticoagulant therapy with INR from 2-3. 49% male, mean age 76, 82% hypertension, 25% DM, 37% HF, 19% stroke, 10% renal impairment, 19% CAD, 9% previous bleeding, **17% antiplatelet therapy**. Median HAS-BLED score of 2, median CHADS2 score of 2.
Inclusion criteria	INR between 2-3
Exclusion criteria	Prosthetic heart valves, ACS, stroke, valvular AF, any haemodynamic instability, surgical Rx or hospital admission in last 6 months
Anticoagulants used	Acenocoumarol
Risk tools used	HAS-BLED ATRIA
Outcome definition	Major bleeding – 2005 International Society on Thrombosis and Haemostasis criteria.
Mean follow up time	952 days
Number of bleeding events	79 people with major bleeds (16 ICH)
Results	C statistics for major bleeding ATRIA (cont) 0.68(0.65-0.71) HAS-BLED (cont) 0.71(0.68-0.74)

Reference	Roldan, 2013[125]
	ATRIA (0-4 vs ≥5) 0.59(0.55-0.62) HAS-BLED (0-2 vs ≥3)0.68(0.65-0.71) Head to head: HAS-BLED sig better for both methods above. NRI HAS-BLED v ATRIA (cont): +0.136, p=0.43 (due more to correct reclassification of events than non-events) NRI HAS-BLED v ATRIA (cat): +0.196, p=0.19 (due mostly to correct reclassification of events than non-events)

Table 80. Roldan, 2013[126]

Reference	Roldan, 2013[126]
Study type	Prospective cohort study
Study sample	1370 consecutive patients with AF receiving anticoagulant therapy with INR from 2-3. 49% male, mean age 76, 19% stroke, 10% renal impairment, 18% CAD, 9% previous bleeding, **18% antiplatelet therapy**. Median HAS-BLED score of 2, median CHADS2 score of 2.
Inclusion criteria	INR between 2-3
Exclusion criteria	Prosthetic heart valves, ACS, stroke, valvular AF, any haemodynamic instability, surgical Rx or hospital admission in last 6 months
Anticoagulants used	Acenocoumarol
Risk tools used	HAS-BLED CHADS CHADSVASC
Outcome definition	Major bleeding – 2005 International Society on Thrombosis and Haemostasis criteria.
Mean follow up time	996 days
Number of bleeding events	114 people with major bleeds (16 ICH)
Results	C statistics for major bleeding

Reference	Roldan, 2013[126]
	HAS-BLED:0.69(0.67-0.72) CHADS: 0.59(0.56-0.62) CHADSVASC: 0.58(0.55-0.60) Head to head: HAS-BLED sig better than both CHADS2 and CHADSVASC, NRI HAS-BLED v CHADS: +0.3826, p<0.001 (due more to correct reclassification of events than non-events) NRI HAS-BLED v CHADSVASC: +0.3760, p<0.001 (due mostly to correct reclassification of events than non-events)

Table 81. Roldan, 2018[128]

Reference	Roldan, 2018[128]
Study type	Prospective cohort study
Study sample	1361 consecutive patients with AF who were taking VKA OACs (acenocoumarol), in Spain. Age 76, 49% male, 82% hypertensive, 27% DM, 19% previous stroke/TIA, 19% CAD, 31% HF, 7% PAD, 10% renal impairment, 33% hypercholesterolemia, 8% previous bleeding episode, 4% alcohol abuse, 1% hepatic disease, 8% cancer. **18% antiplatelet therapy**. Median HAS-BLED score of 2
Inclusion criteria	Permanent or paroxysmal AF (not defined) who were taking VKA OACs. All had to have good anticoagulation control with INR between 2 and 3 during the past 6 months of clinic visits
Exclusion criteria	Prosthetic heart valves, rheumatic AF, acute coronary syndrome, stroke, potentially unstable chest pain or any haemodynamic instability, as well as patients who had hospital admission or surgical intervention in past 6 months.
Anticoagulants used	VKAs
Risk tools used	HAS-BLED Modified HAS-BLED (including vWF, high sensitivity troponin T, N-terminal fragment B-type natriuretic peptide, high sensitivity IL-6, time in therapeutic range and modification of diet in renal disease) CHADS-VASC Modified CHADSVASC (as above)
Outcome definition	Major bleeding events – based on the 2005 International Society on Thrombosis and Haemostasis criteria, which were: fatal or symptomatic bleeding in a critical area or organ, such as IC, intraspinal, intraocular, retroperitoneal, intra-articular or pericardial, or

Reference	Roldan, 2018[128]
	IM with compartment syndrome causing a fall in Hb of 20 or more g/L.
Mean follow up time	2375 days`(7.49 years)
Number of bleeding events	250 (2.83% per year)
Results	HAS-BLED for major bleeding 0.60(0.56-0.63) Modified HAS-BLED for major bleeding 0.60(0.56-0.64) CHADSVASC for major bleeding 0.55(0.51-0.58) Modified CHADSVASC for major bleeding 0.56(0.53-0.60) NRI modified HAS-BLED vs HAS-BLED: -0.062 (p=0.133) NRI modified CHADSVASC vs CHADSVASC: -0.0026 (p=0.830)

Table 82. Senoo, 2016[136]

Reference	Senoo, 2016[136]
Study type	Retrospective cohort study
Study sample	2283 patients with AF on non-warfarin OAC. UK. Age 70. No other details of demographics reported.
Inclusion criteria	Patients in AMADEUS trial in the idraparinux arm
Exclusion criteria	None reported
Anticoagulants used	Idraparinux (non-warfarin anticoagulant)

Reference	Senoo, 2016[136]
Risk tools used	HAS-BLED ORBIT
Outcome definition	Major bleeding Clinically relevant bleeding
Mean follow up time	Mean 311 days
Number of bleeding events	74 major bleeding and 346 clinically relevant bleeding events
Results	C index clinically relevant bleeding HAS-BLED: 0.61(0.58-0.64) ORBIT: 0.58(0.55-0.61) C index major bleeding HAS-BLED: 0.59(0.53-0.65) ORBIT: 0.58(0.52-0.64) Sensitivity/specificity major bleeding HAS-BLED ≥1:0.959/0.163 ≥2:0.446/0.662 ≥3:0.108/0.937 ORBIT ≥1:0.743/0.374 ≥2:0.297/0.800 Sensitivity/specificity CR bleeding HAS-BLED ≥1:0.913/0.171 ≥2:0.496/0.686 ≥3:0.127/0.944

Reference	Senoo, 2016[136]
	ORBIT ≥1:0.733/0.388 ≥2:0.281/0.811 NRI clinically important bleeding HAS-BLED v ORBIT: +0.156(+0.043 to +0.27) NRI major bleeding HAS-BLED v ORBIT: -0.037(-0.265 to +0.192)

Table 83. Senoo, 2016[137]

Reference	Senoo, 2016[137]
Study type	Retrospective cohort study
Study sample	2293 patients with AF warfarin OAC. UK. Age 71, 65.5% male, paroxysmal AF 35.5%, persistent AF 9.3%, permanent AF 54.9%, hypertension 77%, HF 24%, DM 20%, CAD 31%, Stroke/TIA 25%, TTR 58%, **Aspirin 16.5%;NSAIDS 5.4%.**CHASVASC of 0-2: 28.8%, HAS-BLED 2.
Inclusion criteria	Patients in AMADEUS trial in the Warfarin arm. ECG evidence of AF, indication for long term anticoagulation.
Exclusion criteria	Contraindications to anticoagulation, renal dysfunction (CrCl <10 mL/min, breastfeeding, pregnancy, recent procedures causing prolonged bleeding.
Anticoagulants used	Warfarin
Risk tools used	HAS-BLED ORBIT ATRIA Also with TTR for NRI analysis of ORBIT and ATRIAS only
Outcome definition	Major bleeding (BLINDED) Clinically relevant bleeding (BLINDED)
Mean follow up time	Unclear but probably <1 year
Number of	39 major bleeding and 251 clinically relevant bleeding events

Reference	Senoo, 2016[137]
bleeding events	
Results	C index clinically relevant bleeding HAS-BLED: 0.59(0.56-0.63) ORBIT: 0.52(0.48-0.56) ATRIA: 0.50(0.46-0.53) Head to head: HAS-BLED significantly better. C index major bleeding HAS-BLED: 0.65(0.56-0.73) ORBIT: 0.61(0.51-0.70) ATRIA: 0.61(0.51-0.70) Head to head: NS NRI clinically important bleeding ATRIA + TTR vs ATRIA: +0.260, p<0.001 ORBIT + TTR vs ORBIT: +0.260, p<0.001 NRI major bleeding ATRIA + TTR vs ATRIA: +0.348, p=0.02 ORBIT + TTR vs ORBIT: +0.348, p=0.02

Table 84. Serna, 2018[138]

Reference	Serna, 2018[138]
Study type	Prospective cohort study
Study sample	652 consecutive ASF patients stable on VKAs (INR 2-3) for 6 months. Spain. Age 76, 48.6% male, 82.8% hypertension, 24.2% DM, 18.7% history of stroke/TIA, 18.4% CAD, 31.9% hypercholesterolemia, 34.5% HF, 9.2% renal impairment, 1.5% hepatic impairment, 8.3% previous bleeding. HAS-BLED score 2. **No data on antiplatelets**.

Reference	Serna, 2018[138]
Inclusion criteria	On Acenocoumarol - stable at INR 2-3 for 6 months
Exclusion criteria	Prosthetic heart vales
Anticoagulants used	Acenocoumarol (VKA)
Risk tools used	HAS-BLED GEN /HAS-BLED (added point if patient carrying VKORC1 allele and CYP2C9*3 polymorphisms)
Outcome definition	Major bleeding (20015 ICTH)
Mean follow up time	7.6 years
Number of bleeding events	106 patients with major bleeding (42 ICH, 44 GI bleeding).
Results	C index major bleeds HAS-BLED: 0.66 (0.622-0.696) GEN/HAS-BLED: 0.645(0.607-0.682) Head to head: HAS-BLED sig better [IDI -0.013 (p<0.001)] NRI GEN/HAS-BLED vs HAS-BLED: -0.044 (p=0.015)

Table 85. Schwartz, 2019[135]

Reference	Schwartz, 2019[135]
Study type	Retrospective cohort study
Study sample	Data from 9819 patients with AF who were on DOACs or VKAs were retrieved from the Northwestern Healthcare system's Enterprise Database Warehouse. The data allowed identification of bleeding outcomes, and calculation of prior HAS-BLED 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; 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

Reference	Schwartz, 2019[135]
Anticoagulants used	61% VKA, 39% DOACs
Risk tools used	Modified HAS-BLED(no stroke/TIA component and no labile INR)
Outcome definition	Major bleeding: ISTH criteria
Mean follow up time	971 days after AF diagnosis (mean)
Number of bleeding events	604
Results	**HAS-BLED** C statistic ('whites'): 0.572 (0.546-0.598) C statistic ('non-whites'): 0.603(0.55-0.66) Accuracy (derived from table 3 in the paper, summating the data in 'whites' and 'non-whites' to produce the overall accuracy figures Threshold of >0, sensitivity 0.9255, spec 0.1504 (TP 559, TN 45, FP 7829, TN 1386). Threshold of >1, sensitivity 0.644, spec 0.5063 (TP 389, TN 215, FP 4549, TN 4666). Threshold of >2, sensitivity 0.311, spec 0.826 (TP 188, TN 416, FP 1600, TN 7615).

Table 86. Siu, 2014[142]

Reference	Siu, 2014[142]
Study type	Retrospective cohort study
Study sample	1912 patients with NVAF (not defined) who received OACs (Warfarin). Mean age 73, 47% female, 55.8% hypertensive, 24% DM, 1.8% renal failure on dialysis, 24% HF, 24% CAD, 6.3% PAD, 29.6% prior stroke/TIA, prior IC haemorrhage 2.1%. Mean

Reference	Siu, 2014[142]
	CHADSVASC 3.3. **No data on antiplatelets**
Inclusion criteria	Non valvular AF
Exclusion criteria	Significant valvular heart disease, previous valvular surgery.
Anticoagulants used	Warfarin
Risk tools used	HAS-BLED
Outcome definition	Intracranial haemorrhage (not defined)
Mean follow up time	3.19 years
Number of bleeding events	30 developed ICH during follow up (annual incidence per year if 0.8%)
Results	C statistics for ICH HAS-BLED: 0.574(0.518-0.629)

Table 87. Steinberg, 2016[146]

Reference	Steinberg, 2016[146]
Study type	Prospective cohort study
Study sample	7420 AF patients on OACs, out of an original cohort of 9715 from the ORBIT-AF trial. USA. Ranges for baseline data given as different data given for people in low, intermediate and high risk categories. Age 73-77, female 40-46%, hypertension 83-87%, diabetes 28-38%, previous GI bleed 5.7-16%, CAD 32-48%, Prior stroke/TIA 14-26%, CHF 30-46%, HAS-Bled 1.61-2.17, CHADS2 2.17-2.81. **No data on antiplatelets**.
Inclusion criteria	Aged 18 or older, electrocardiographically documented AF not due to a reversible cause
Exclusion criteria	Patients without follow-up
Anticoagulants used	6942 Warfarin, 478 dabigatran
Risk tools used	ATRIA HAS-BLED
Outcome definition	Major bleeding (2005 ISTH)

Reference	Steinberg, 2016[146]
Mean follow up time	Unclear
Number of bleeding events	632
Results	C statistics for major bleeding (not differentiated between OACs) ATRIA: 0.629(0.608-0.65) HAS-BLED: 0.605(0.586-0.624) Sensitivity/specificity ATRIA ≥'intermediate risk': 0.547/0.685 ≥'high risk': 0.402/0.796 HAS-BLED ≥'intermediate risk': 0.98/0.079 ≥'high risk': 0.371/0.803

Table 88. Suzuki, 2014[147]

Reference	Suzuki, 2014[147]
Study type	Prospective cohort study
Study sample	231 NVAF patients on warfarin for at least 1 year. Demographics given as ranges as only reported for sub-groups of eGFR: age 68-74, 63.1-80% male, hypertension 53.2 to 64.4%, CAD 14.4 to 16.7%, CHF: 20 to 25.2%, dyslipidaemia 28.8 to 36.7%, eGFR 12.7 to 74.3 mL/min/1.73m^2) **antiplatelet drugs 36.9 to 50%.**TTR 56.9 to 65.1%.
Inclusion criteria	NVAF
Exclusion criteria	HF, cardiomyopathy, congenital heart disease, permanent pacemaker, uncontrolled pulmonary disease, thyroid dysfunction, malignant disease.
Anticoagulants used	Warfarin
Risk tools used	HAS-BLED

Reference	Suzuki, 2014[147]
	Modified HAS-BLED (renal dysfunction defined by eGFR <60, with exclusion of the 'elderly' factor because eGFR is calculated based on patient age)
Outcome definition	Major haemorrhage event (2005 ICTH)
Mean follow up time	7.1 years
Number of bleeding events	44
Results	C statistics HAS-BLED: 0.64(0.55-0.72) Modified HAS-BLED: 0.67(0.57-0.75) Head to head: NSD NRI Modified HAS-BLED v HAS-BLED +0.50 (p=0.002) IDI 0.033 (p=0.043)

Table 89. Wang, 2016[154]

Reference	Wang, 2016[154]
Study type	Retrospective cohort study
Study sample	21,934 adults with AF who were starting dabigatran (30%) or Warfarin. Patients were on a healthcare claims database in USA. Demographic data given for those on Warfarin (n=15418): Age 65, female 34%, 27% CHF, 31% DM, 93% hypertensive, 20% prior stroke, 22% PVD. 43% with HAS-BLED score of 3 or more. 32% with CHADS2 score of 3 or more.
Inclusion criteria	Aged >18 years; at least one recorded diagnosis of AF according to ICD-9 classification.
Exclusion criteria	None reported

Reference	Wang, 2016[154]
Anticoagulants used	Dabigatran and Warfarin
Risk tools used	HAS-BLED
Outcome definition	Major bleeding – including the ICD codes for haemorrhagic stroke, GI, urogenital or other bleeds.
Mean follow up time	5 months
Number of bleeding events	Annual event rates were 4.6 for major bleeding
Results	C statistics (Dabigatran) HAS-BLED: 0.60 (0.54-0.67) C statistics (Warfarin) HAS-BLED: 0.62 (0.59-0.66) Calibration (goodness of fit statistic) Dabigatran: 6.30, p=0.04 Warfarin: 36.97, p=0.00

Table 90. Yao, 2017[158]

Reference	Yao, 2017[158]
Study type	Retrospective cohort study
Study sample	39, 539 patients with NVAF from USA insurance database (OptumsLabs Data Warehouse) who had started DOACs between 2010 and 2015. Age 71, 42% female, 20% non-white, 28% HF, 86% hypertension, 34% DM, 14% previous strokes/TIA, 48% vascular disease, 7% stage II or IV CKD, 4% abnormal liver function, 9% previous major bleeding, **7% using antiplatelets, 5% using NSAIDs**, 28% had had previous warfarin exposure. HAS-BLED: 2
Inclusion criteria	>18 with NVAF; started apixaban, rivaroxaban, edoxaban or dabigatran between 2010 to 2015
Exclusion criteria	Not reported
Anticoagulants used	Apixaban, rivaroxaban, edoxaban or dabigatran

Reference	Yao, 2017[158]
Risk tools used	CHADSVASC CHADS HAS-BLED ORBIT ATRIA
Outcome definition	Major bleeding
Mean follow up time	0.6 years
Number of bleeding events	665 people with major bleeding (including 74 ICHs)
Results	C statistics Major bleeding (continuous) CHADSVASC: 0.68(0.66 to 0.70) CHADS: 0.65(0.63 to 0.67) HAS-BLED: 0.66(0.64 to 0.67) ORBIT: 0.66(0.64 to 0.68) ATRIA: 0.67(0.65 to 0.69) Major bleeding (categorical) CHADSVASC: 0.65(0.63 to 0.66) CHADS: 0.64(0.62 to 0.65) HAS-BLED: 0.64(0.62 to 0.66) ORBIT: 0.60(0.58 to 0.62) ATRIA: 0.60(0.58 to 0.62) NRI major bleeding (all vs CHADSVASC) CHADS: -0.04 HASBLED: 0.02 ORBIT: 0.01 ATRIA: 0.05

Reference	Yao, 2017[158]
	ICH (continuous) CHADSVASC: 0.65(0.59 to 0.71) CHADS: 0.66(0.60 to 0.72) HAS-BLED: 0.64(0.58 to 070) ORBIT: 0.60(0.54 to 0.66) ATRIA: 0.63(0.57 to 0.68) ICH (categorical) CHADSVASC: 0.61(0.57 to 0.66) CHADS: 0.66(0.60 to 0.72) HAS-BLED: 0.63(0.58 to 0.69) ORBIT: 0.55(0.50 to 0.61) ATRIA: 0.56(0.50 to 0.61) NRI ICH (all vs CHADSVASC) CHADS: 0.09 HASBLED: 0.07 ORBIT: -0.06 ATRIA:- 0.04 Sensitivity/specificity CHADSVASC Major bleeding ≥2: 0.983/0.128 ≥4: 0.669/0.458 ICH ≥2:0.973/0.127 ≥4:0.756/0.454 CHADS2 Major bleeding

Reference	Yao, 2017[158]
	≥2:0.865/0.341 ≥4:0.288/0.856 ICH ≥2:0.865/0.338 ≥4:0.365/0.854 HAS-BLED Major bleeding ≥2:0.915/0.268 ≥3: 0.583/0.642 ICH ≥2: 0.878/0.266 ≥3:0.594/0.638 ORBIT Major bleeding ≥3:0.364/0.831 ≥4:0.185/0.936 ICH ≥3:0.283/0.828 ≥4:0.095/0.936 ATRIA Major bleeding ≥4:0.409/0.772 ≥5:0.313/0.866 ICH ≥4:0.338/0.769

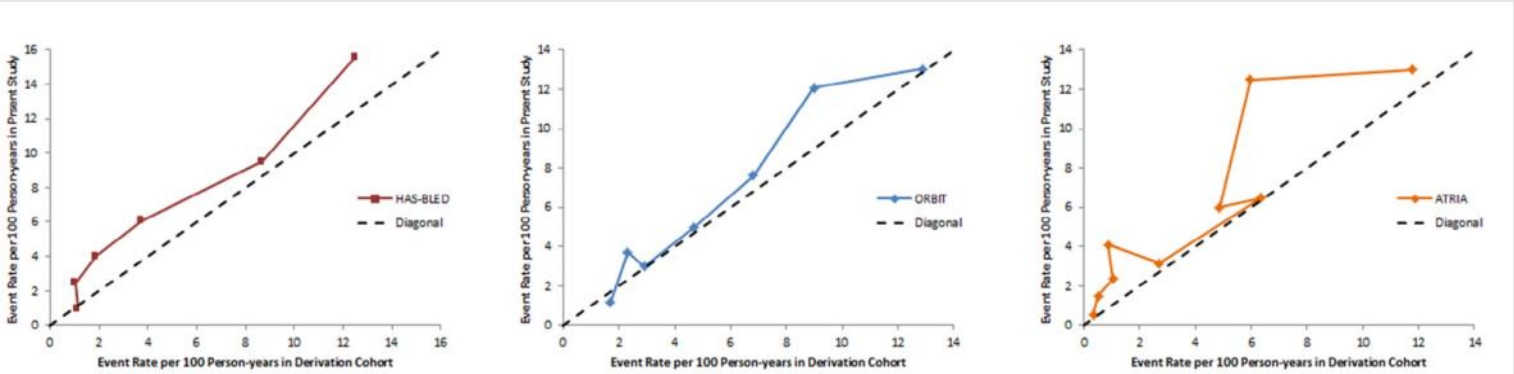
Reference	Yao, 2017[158]
	≥5:0.230/0.861 Calibration ORBIT and HAS-BLED were reported to have better calibration than ATRIA, but no data given. Calibration plots are given below:  Figure 3. Calibration plots for bleeding risk scores.

Table 91. Elvira-Ruiz, 2020[30]

Reference	Elvira-Ruiz, 2020[30]
Study type	Retrospective predictive study
Study sample	2,880 NVAF patients initiating oral anticoagulants; age 77; 51.1% women; 49.3% permanent AF; hypertension 85.5%; DM 33.9%; CHADSVASC 4; HASBLED 2; ATRIA 3; ORBIT 1.
Inclusion criteria	All non-valvular AF patients initiating oralanticoagulation (VKA or NOAC) for the prevention ofstroke or systemic embolism and with an available echocardiogram at two hospitals
Exclusion criteria	Patients who received oralanticoagulants for other indications or for cardioversionwhen long-term anticoagulation was not indicated. Patientswith hypertrophic cardiomyopathy, moderate to severerheumatic mitral stenosis or mechanical prosthetic valvesand those with a previous history of oral anticoagulanttherapy were also excluded. Patients who underwent aorticvalve replacement were

Reference	Elvira-Ruiz, 2020[30]
	censored at the time of intervention.
Anticoagulants used	Apixaban, rivaroxaban, edoxaban or dabigatranand VKAs
Risk tools used	HAS-BLED ORBIT ATRIA HAS-BLEDwith existence of aortic stenosis (AS) ORBITwith AS ATRIAwith AS
Outcome definition	Major bleeding–ISTH defined
Mean follow up time	18 months
Number of bleeding events	185people with major bleeding
Results	C statistics HAS BLED 0.66(0.64-0.68) HAS-BLED with AS: 0.68(0.66-0.70) ATRIA 0.65(0.64-0.67) ATRIA with AS: 0.67(0.66-0.69) ORBIT 0.67(0.65-0.68) ORBIT with AS: 0.68(0.67-0.70) Sensitivity/specificity(calculated from raw data in supplement) Scores Total MB events HASBLED:0-1 445 7 2 1111 50 3 or more 1241 122 ≥2: sen 0.961, spec 0.167; TP 172, FN 7, FP 2180, TN 438 ≥3: sen 0.682, spec 0.573; TP 122, FN 57, FP 1119, TN 1499 Scores Total MB events

Reference	Elvira-Ruiz, 2020[30]
	ATRIA: 0-3 1975 86 4 202 21 5 or more 702 78 ≥4: sen 0.535, spec 0.701; TP 99, FN 86, FP 805, TN 1889 ≥5: sen 0.422, spec 0.768; TP 78, FN 107, FP 624, TN 2070 Scores Total MB events ORBIT : 0-2 2123 96 3 318 33 4 or more 438 56 ≥3: sen 0.481, spec 0.752; TP 89, FN 96, FP 667, TN 2027 ≥4: sen 0.303, spec 0.858; TP 56, FN 129, FP 382, TN 2312 <u>NRI major bleeding</u> HAS-BLED with AS vs HAS-BLED: +0.0481 (p=0.034); better at detecting events ATRIA with AS vs ATRIA: +0.0645 (p=0.025); better at detecting non-events ORBIT with AS vs ORBIT: +0.0227 (p=0.170); better at detecting events

Appendix H: 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 MB 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
Apostolakis, 2012[4]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N	N	U	U	Y	Y	Y	Y	Very serious
Apostolakis, 2013[3]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N	N	U	U	Y	Y	Y	Y	Very serious
Barnes, 2014[8]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y	N	U	U	Y	Y	Y	Y	Very serious
Berg, 2019[11]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y	N	U	U	Y	Y	Y	Y	Very serious
Beshir, 2018[14]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N	N	N	Y	Y	Y	Y	Y	Very serious
Chang, 2016[19]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	N	N	Y	Y	Y	Y	Y	Y	Very serious
Chao, 2018[21]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y	N	U	U	Y	Y	Y	Y	Very serious
Chao, 2018b[20]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y	N	U	U	Y	Y	Y	Y	Very serious
Claxton, 2018[23]	Y	Y	Y	Y	U	Y	Y	Y	NA	Y	U	Y	N	U	U	Y	Y	Y	Y	Very serious

Overall rating	Very serious	Very serious	Serious	Very serious	Very serious	Very serious	Very serious	Very serious	Very serious
Model recalibrated or likely that calibration not needed?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Relevant performance measures?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Complexities in data accounted for?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Non:binary predictors handled appropriately?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Missing data handled appropriately?	U	U	U	U	U	U	U	U	U
All enrolled included in analysis?	U	U	U	U	U	U	U	U	U
Time interval between baseline and outcome appropriate? (>5 years)	N	N	Y	N	N	N	N	N	N
Reasonable number of MB outcome events? (100)	Y	N	Y	Y	Y	Y	U	U	N
Outcome determined without knowledge of predictor information?	U	U	U	Y	U	U	U	U	U
Outcome defined in same way for all?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Predictors excluded from outcome definition?	NA	NA	NA	NA	NA	NA	NA	NA	NA
Pre-specified outcome used?	Y	Y	Y	Y	Y	Y	Y	Y	Y
All relevant predictors analysed?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Predictors all available at time model meant to be used?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Predictor assessments made without knowledge of outcome data?	U	U	U	U	U	U	U	U	U
Predictors defined/ass' d same for all?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Similar health across participants?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Appropriate inc and exc?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Appropriate data sources?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Study	Dalgaard, 2019[25]	Esteve-Pastor, 2016[31]	Esteve-Pastor, 2017a[5]	Esteve-Pastor, 2017b[32]	Fang, 2011[33]	Fox, 2017[36]	Friberg, 2012[37]	Gage, 2006[38]	Gallego, 2012[39]

Overall rating	Serious	Very serious	Very serious	Very serious	Very serious	Very serious	Very serious	Very serious	Very serious	Very serious
Model recalibrated or likely that calibration not needed?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Relevant performance measures?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Complexities in data accounted for?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Non:binary predictors handled appropriately?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Missing data handled appropriately?	U	U	U	U	U	U	Y	U	U	U
All enrolled included in analysis?	U	U	U	U	U	U	N	U	U	U
Time interval between baseline and outcome appropriate? (>5 years)	Y	N	N	N	N	N	N	N	N	N
Reasonable number of MB outcome events? (100)	Y	Y	Y	Y	Y	Y	N	N	Y	Y
Outcome determined without knowledge of predictor information?	U	Y	Y	U	Y	U	U	U	Y	U
Outcome defined in same way for all?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Predictors excluded from outcome definition?	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Pre-specified outcome used?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
All relevant predictors analysed?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Predictors all available at time model meant to be used?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Predictor assessments made without knowledge of outcome data?	U	U	U	U	U	U	U	U	U	U
Predictors defined/ass'd same for all?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Similar health across participants?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Appropriate inc and exc?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Appropriate data sources?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Study	Garcia-Fernandez, 2017[41]	Hijazi, 2014[56]	Hijazi, 2014a[57]	Hijazi, 2016[54]	Hijazi, 2017[52]	Hilkens, 2017[58]	Jaspers Focks, 2016[63]	Jover, 2012[65]	Lip, 2011[71]	Lip, 2014[74]

Study	Overall rating	Model recalibrated or likely that calibration not needed?	Relevant performance measures?	Complexities in data accounted for?	Non:binary predictors handled appropriately?	Missing data handled appropriately?	All enrolled included in analysis?	Time interval between baseline and outcome appropriate? (>5 years)	Reasonable number of MB outcome events? (100)	Outcome determined without knowledge of predictor information?	Outcome defined in same way for all?	Predictors excluded from outcome definition?	Pre-specified outcome used?	All relevant predictors analysed?	Predictors all available at time model meant to be used?	Predictor assessments made without knowledge of outcome data?	Predictors defined/ass' d same for all?	Similar health across participants?	Appropriate inc and exc?	Appropriate data sources?
Lip, 2018[77]	Very serious	Y	Y	Y	Y	U	U	N	U	U	Y	NA	Y	Y	Y	U	Y	Y	Y	Y
Mori, 2019[88]	Serious	Y	Y	Y	Y	NA	Y	N	N	Y	Y	NA	Y	Y	Y	Y	Y	Y	Y	Y
Nielsen, 2016[90]	Very serious	Y	Y	Y	Y	U	U	U	U	U	Y	NA	Y	Y	Y	U	Y	Y	Y	Y
O'Brien, 2015[91]	Very serious	Y	Y	Y	Y	U	U	N	Y	U	Y	NA	Y	Y	Y	U	Y	Y	Y	Y
Olesen, 2011[95]	Very serious	Y	Y	Y	Y	U	U	N	Y	U	Y	NA	Y	Y	Y	U	Y	Y	Y	Y
Pisters, 2010[103]	Very serious	Y	Y	Y	Y	U	U	N	U	U	Y	NA	Y	Y	Y	U	Y	Y	Y	Y
Poli, 2017[110]	Very serious	Y	Y	Y	Y	U	U	N	Y	U	Y	NA	Y	Y	Y	U	Y	Y	Y	Y
Prochaska, 2018[113]	Very serious	Y	Y	Y	Y	U	U	N	Y	U	Y	NA	Y	Y	Y	U	Y	Y	Y	Y
Proietti, 2016[116]	Very serious	Y	Y	Y	Y	U	U	N	Y	U	Y	NA	Y	Y	Y	U	Y	Y	Y	Y
Proietti, 2018[114]	Very serious	Y	Y	Y	Y	U	U	N	Y	Y	Y	NA	Y	Y	Y	U	Y	Y	Y	Y
Proietti, 2018[115]	Very serious	Y	Y	Y	Y	U	U	N	Y	Y	Y	NA	Y	Y	Y	U	Y	Y	Y	Y

Overall rating	Very serious	Serious	Serious	Very serious	Very serious	Serious	Very serious	Very serious	Very serious
Model recalibrated or likely that calibration not needed?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Relevant performance measures?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Complexities in data accounted for?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Non:binary predictors handled appropriately?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Missing data handled appropriately?	U	U	U	U	U	U	U	U	U
All enrolled included in analysis?	U	U	U	U	U	U	U	U	U
Time interval between baseline and outcome appropriate? (>5 years)	N	Y	Y	N	N	Y	N	N	N
Reasonable number of MB outcome events? (100)	U	Y	Y	N	Y	Y	Y	N	Y
Outcome determined without knowledge of predictor information?	U	U	U	U	U	U	U	U	U
Outcome defined in same way for all?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Predictors excluded from outcome definition?	NA	NA	NA	NA	NA	NA	NA	NA	NA
Pre-specified outcome used?	Y	Y	Y	Y	Y	Y	Y	Y	Y
All relevant predictors analysed?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Predictors all available at time model meant to be used?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Predictor assessments made without knowledge of outcome data?	U	U	Y	U	U	U	U	U	U
Predictors defined/ass'd same for all?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Similar health across participants?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Appropriate inc and exc?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Appropriate data sources?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Study	Quinn, 2016[117]	Rivera-Caravaca, 2017[120]	Rivera-Caravaca, 2019[119]	Roldan, 2013a[125]	Roldan, 2013b[126]	Roldan, 2018[128]	Schwartz, 2019[135]	Senoo, 2016[136]	Senoo, 2016b[137]

Study	Serna, 2018[138]	Siu, 2014[142]	Steinberg, 2016[146]	Suzuki, 2014[147]	Wang, 2016[154]	Yao, 2017[158]	Elvira-Ruiz, 2020[30]
Appropriate data sources?	Y	Y	Y	Y	Y	Y	Y
Appropriate inc and exc?	Y	Y	Y	Y	Y	Y	Y
Similar health across participants?	Y	Y	Y	Y	Y	Y	Y
Predictors defined/ass' d same for all?	Y	Y	Y	Y	Y	Y	Y
Predictor assessments made without knowledge of outcome data?	U	U	U	U	U	U	U
Predictors all available at time model meant to be used?	Y	Y	Y	Y	Y	Y	Y
All relevant predictors analysed?	Y	Y	Y	Y	Y	Y	Y
Pre-specified outcome used?	Y	Y	Y	Y	Y	Y	Y
Predictors excluded from outcome definition?	NA	NA	NA	NA	NA	NA	NA
Outcome defined in same way for all?	Y	Y	Y	Y	Y	Y	Y
Outcome determined without knowledge of predictor information?	U	U	U	U	U	U	U
Reasonable number of MB outcome events? (100)	Y	N	Y	N	U	Y	Y
Time interval between baseline and outcome appropriate? (>5 years)	Y	N	N	Y	N	N	N
All enrolled included in analysis?	U	U	U	U	U	U	Y
Missing data handled appropriately?	U	U	U	U	U	U	NA
Non:binary predictors handled appropriately?	Y	Y	Y	Y	Y	Y	Y
Complexities in data accounted for?	Y	Y	Y	Y	Y	Y	Y
Relevant performance measures?	Y	Y	Y	Y	Y	Y	Y
Model recalibrated or likely that calibration not needed?	Y	Y	Y	Y	Y	Y	Y
Overall rating	Serious	Very serious	Very serious	Very serious	Very serious	Very serious	Serious

Y=yes, N=no, U=unclear, NA=not applicable

Appendix I: Economic evidence tables

None.

Appendix J: Excluded clinical studies

No studies were excluded from the review on effectivess.

Table 92: Studies excluded from the clinical reviewRCT

Study	Exclusion reason
Guo, 2020[46]	Study didn't compare different tools

Table 93: Studies excluded from the clinical reviewaccuracy

Study	Exclusion reason
Abumuaileq, 2014[1]	No bleeding accuracy outcomes
Al-Turaiki, 2016[2]	CS study. No bleeding accuracy outcomes
Atzema, 2018[6]	No bleeding accuracy outcomes
Banerjee, 2014[7]	No pure bleeding accuracy outcomes - composites with IS
Benezet-Mazuecos, 2017[9]	Abstract only
Benito-Gonzalez, 2018[10]	Patients undergoing mitral valve repair
Bernaitis, 2017[13]	No bleeding accuracy outcomes
Bernaitis, 2018[12]	No bleeding accuracy outcomes
Burgess, 2013[15]	Only 78% with AF
Caldeira, 2014[16]	SYSTEMATIC REVIEW - REFERENCES CHECKED
Camelo-Castilo, 2020[17]	no specific predictive accuracy outcomes for bleeding outcomes
Candeias Faria, 2018[18]	Abstract only
Chia, 2016[22]	No bleeding accuracy outcomes
Coleman, 2018[24]	Did not evaluate bleeding risk evaluation tools
Deitelzweig, 2014[26]	No bleeding accuracy outcomes
Diemberger, 2018[27]	No bleeding accuracy outcomes
Donze, 2012[28]	Only 61% with AF
Dukanovic, 2017[29]	No bleeding accuracy outcomes
Fanola, 2017[34]	No bleeding risk outcomes; composite outcome only
Fauchier, 2016[35]	No description if OACs were used
Garcia-Fernandez, 2016[40]	Patients undergoing electrical cardioversion
Geersing, 2012[42]	Reference to a trials registry
Giustozzi, 2018[43]	Abstract only
Gorman, 2016[44]	Case control study. Unclear if the data used to form the risk prediction score were based on previous data or simply on data derived at the same time as the bleed. Thus possibility that the study was cross-sectional.
Guo, 2013[45]	Non-anticoagulated
Guo, 2016[48]	Most not anticoagulated
Guo, 2018[47]	Non-anticoagulated
Hijazi, 2014[51]	Conference abstract

Study	Exclusion reason
Hijazi, 2016[50]	No bleeding accuracy outcomes
Hijazi, 2016[53]	Conference abstract
Hijazi, 2017[49]	No bleeding risk outcomes
Hijazi, 2018[55]	No bleeding risk outcomes
Hippisley-Cox, 2014[59]	Not the protocol population
Hippisley-Cox, 2014[60]	Not the protocol population
Iwasaki, 2018[61]	Abstract only
Jaakkola, 2018[62]	No bleeding accuracy outcomes ; only a proportion on OACS
Jensen, 2018[64]	Abstract only
Kearon, 2019[66]	Commentary on Berg, 2019
Lamberts, 2017[67]	No bleeding accuracy outcomes
Lee, 2018[68]	No bleeding accuracy outcomes
Li Kam Wa, 2018[69]	Abstract only
Lip, 2012[70]	<60% on anticoagulants and no separate analysis
Lip, 2012[73]	Review
Lip, 2013[72]	Not an AF population
Lip, 2013[76]	Composite outcomes, not a specific bleeding outcome
Lip, 2018[75]	Exclusively valvular AF
Lobos-Bejarano, 2016[78]	No bleeding accuracy outcomes
Loewen, 2011[79]	SYSTEMATIC REVIEW - REFERENCES CHECKED
Lv, 2020[80]	Only contained evidence relating to GI bleeding
Maeda, 2020[81]	no predictive accuracy data
Marcucci, 2013[83]	No bleeding accuracy outcomes
Marcucci, 2014[82]	No bleeding accuracy outcomes; some not on OACs
McAlister, 2017[84]	Not anticoagulated
McAlister, 2018[85]	No bleeding accuracy outcome
Methavigul, 2020[86]	Did not estimate predictive accuracy for bleeding
Molnar, 2018[87]	Review
O'Caoimh, 2017[92]	Only 17% on OACs
Okumura, 2014[93]	No bleeding accuracy outcomes
Oldgren, 2016[94]	No bleeding accuracy outcomes
Olesen, 2011[96]	No bleeding accuracy outcomes
Olesen, 2011[97]	Conference abstract
Omran, 2012[98]	Only 81% had AF and no sub-grouping
Pardo Sanz, 2018[99]	Abstract only
Parks, 2017[100]	Review
Peacock, 2017[101]	No bleeding accuracy outcomes
Perez-Copete, 2016[102]	Not in English
Poli, 2007[108]	No bleeding accuracy outcomes
Poli, 2009[107]	Conference abstract
Poli, 2009[105]	Conference abstract

Study	Exclusion reason
Poli, 2009[105]	No bleeding accuracy outcomes
Poli, 2011[106]	Conference abstract
Poli, 2011[112]	No bleeding accuracy outcomes
Poli, 2011[104]	Conference abstract
Poli, 2013[111]	Not an AF population
Poli, 2016[109]	Conference abstract
Rivera Caravaca, 2018[123]	Abstract only
Rivera-Caravaca, 2017[118]	No bleeding accuracy outcomes
Rivera-Caravaca, 2017[121]	No bleeding accuracy outcomes
Rivera-Caravaca, 2018[122]	Use of a composite outcome; bleeding risk accuracy not reported
Rivera-Caravaca, 2018[122]	No predictive analysis for bleeding outcomes
Roldan, 2011[127]	No specific bleeding accuracy outcomes
Roldan, 2012[124]	No bleeding accuracy outcomes
Rutherford, 2018[129]	Abstract only
Sadeghi, 2015[130]	Not in English
Saito, 2020[131]	no specific predictive accuracy outcomes for bleeding outcomes
Salpagarova, 2018[132]	Abstract only
Sanders, 2018[133]	SYSTEMATIC REVIEW - REFERENCES CHECKED
Sani, 2016[134]	letter
Shah, 2017[139]	Non-AF population
Shahid, 2017[140]	Review
Silva, 2017[141]	No bleeding accuracy outcomes; some not on OACs
Sogaard, 2017[143]	No bleeding accuracy outcomes
Somme, 2010[144]	No bleeding accuracy outcomes
Sood, 2013[145]	Hemodyalysis patients; non AF
Tchen, 2020[148]	only 81% had AF with no sub-grouping
Thomas, 2014[149]	Review
Toyoda, 2014[150]	No bleeding accuracy outcomes
van Doorn, 2018[151]	RCT but control group were usual care
Van Mieghem, 2017[152]	Review
Wang, 2016[156]	Dialysis population
Wang, 2017[153]	SYSTEMATIC REVIEW - REFERENCES CHECKED
Wang, 2017[155]	No bleeding accuracy outcomes
Wang, 2017[157]	No bleeding accuracy outcomes
Zeng, 2020[159]	SR - references checked
Zhu, 2015[160]	SYSTEMATIC REVIEW - REFERENCES CHECKED
Ziviello, 2019[161]	Abstract only
Zulkifly, 2017[162]	Review

Appendix K: Excluded economic studies

No studies were excluded from the review on effectivenessof tools to predict bleeding.

No studies were excluded from the review on accuracy of tools to predict bleeding.