National Institute for Health and Care Excellence

Final

Venous thromboembolic diseases: diagnosis, management and thrombophilia testing

[F] What factors determine the optimum duration of pharmacological treatment for DVT or PE in people with a VTE?

NICE guideline NG158

Evidence reviews underpinning recommendations 1.4.2 to 1.4.6, 1.4.12 and research recommendations in the guideline

March 2020

Final version

This evidence review was developed by the Guideline Updates Team

Disclaimer

Copyright

ISBN: 978-1-4731-3735-6

Contents

Predicting VTE-recurrence and major bleeding

Review question

What factors determine the optimum duration of pharmacological treatment for DVT or PE in people with a VTE?

Introduction

The aim of long term anticoagulant treatment is to prevent VTE recurrence. However, extended duration treatment is accompanied by an increased bleeding risk. There is uncertainty about which factors influence a person's risk of bleeding and of VTE-recurrence, and uncertainty about which factors should be taken into account when considering the duration of treatment. Several risk tools (composites of multiple risk factors) have been developed that aim to provide an estimated risk of recurrence or bleeding in individual patients. These tools could help to identify candidates for longer term anticoagulant treatment (those with a high recurrence risk, weighed against the risk of bleeding).

This review examines the ability of these tools to predict VTE recurrence in people who have been treated for at least 3 months for VTE and have now discontinued treatment, and to predict major bleeding in people currently receiving anticoagulant treatment for VTE.

PICO table

See Table 1 for a summary of the review protocol, which is presented in full in appendix A.

Table 1 PICO table for the prediction of risk of VTE recurrence and major bleeding

Population	Adults (18+ years) who have had VTE For studies predicting VTE recurrence, participants must have previously received a course of anticoagulant treatment, but not be receiving anticoagulant treatment for the duration of follow up. For studies predicting major bleeding participants must be receiving anticoagulant treatment for the duration of follow up.
Intervention	Any risk tool using multiple predictors to predict the risk of VTE recurrence following cessation of anticoagulant treatment. For example: DASH, HERD002, Vienna. Any risk tool using multiple predictors to predict the risk of major bleeding during anticoagulant treatment. For example: HAS-BLED, VTE-BLEED.
Outcomes	• Recurrence of VTE **or** • Major bleeding

Model performance measures	For each outcome, prognostic accuracy measures will be reported where available, for example: • Odds ratios/hazard ratios • Model fit (e.g. r squared) • Sensitivity, specificity, positive and negative predictive values.

Methods and process

This evidence review was developed using the methods and process described in developing NICE guidelines: the manual (2014). Methods specific to this review question are described in the review protocol in appendix B.

Additional methodological issues were as follows:

- Prognostic model studies can be divided into three groups (derivation studies without external validation, derivation studies with external validation, and external validation studies for existing models that may or may not include updating the original model). For this review, data from validation studies was prioritised, where available, over data from derivation studies. However, where derivation studies were the only studies available for a given model these were extracted, and this was noted in the evidence tables and evidence statements.
- For sensitivity/specificity outcome measures, a high risk score was defined as positive and low risk score as negative. For example, a low risk person who goes on to experience an event (recurrence or bleeding) was classified as a false negative, and a person with a high risk score that did not experience an event was classified as false positive. However, many risk tools gave three levels of risk: low, moderate and high. To enable estimation of sensitivity and specificity, three level tools were merged into two levels: high versus moderate/low risk (moderate and low risk merged) or high/moderate versus low risk (moderate and high risk merged). Both of these comparisons were reported in the GRADE tables for each three level tools for which sensitivity and specificity were estimable.
- Studies predicting major bleeding were excluded if the study included major bleeding events that occurred off-treatment. Conversely, studies predicting VTE-recurrence were excluded if the study included recurrences that occurred on-treatment.
- Since the risk tool assessed in Nieto 2013 was not clearly named and was based on data from the RIETE database, we have named the tool RIETE 2. This tool was designed to predict fatal bleeding, but was included in this review because the study also presented results for major bleeding.

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

Protocol deviation

Priority screening was not used for this review. All references returned by the search were screened at title and abstract level.

Clinical evidence

Included studies

This review was conducted as part of a larger update of the 2012 NICE VTE guideline (CG144). A systematic literature search for prognostic studies and systematic reviews (SRs) of prognostic studies was conducted and returned 11,220 references (see appendix C for the literature search strategy). Based on title and abstract, 11,064 references were excluded, and 156 references were ordered for screening on full text. Systematic reviews were used as an additional source of primary studies and were then excluded.

Of the 156 references screened as full texts, 23 references met the inclusion criteria specified in the review protocol for this question (appendix A). The clinical evidence study selection is presented as a diagram in appendix D.

A second set of searches was conducted at the end of the guideline development process for all updated review questions using the original search strategies, to capture papers published whilst the guideline was being developed. These searches returned 6,272 references in total for all the questions included in the update, and these were screened on title and abstract. Seven references were screened at full text and 2 additional relevant references were included for this review question.

Excluded studies

See Appendix I for a list of references for excluded studies, with reasons for exclusion and appendix J for the full references.

Summary of clinical studies included in the evidence review

This review identified a number of trials validating prediction tools that aim to predict VTE-recurrence (following stopping anticoagulant therapy) or major bleeding (whilst receiving treatment). There is a high degree of heterogeneity in the studies included in this review: various different risk tools are used; several tools consider the same risk factors, however there are differences between studies in how these are measured, defined and coded (e.g. some continuous variables such as D-dimer are either left as continuous or transformed into a dichotomous variable).

Tools investigated in this review for their ability to predict VTE-recurrence are listed in Table 2, outlining how the tools are scored and interpreted, and which studies evaluated each tool; details on the accompanying studies are listed in Table 4. Tools investigated for predicting major bleeding can be found in Table 3, with the accompanying studies in Table 5. Derivation studies were only included if there were no corresponding validation studies.

See appendix E for full evidence tables and appendix J for the full references.

Table 2 Tools to predict VTE-recurrence

Tool	Studies evaluation this tool	Score	Interpretation	Key prediction time point	Derivation or validation study
DASH	Tosetto (2017)	Abnormal D-dimer (measured ~1 months after stopping anticoagulation): 2 points Age <50 years: 1 point Male sex: 1 point Hormone use at VTE onset (if female): -2 points	Low risk: 0-1 point High risk: 2+ points	12, 24, 60 months	External validation study
DAMOVES	Franco (2016)	Age, Sex, Obesity (yes vs. no), Abnormal D-dimer (yes, no, not performed), factor VIII count, genetic thrombophilia (yes vs. no), varicose veins (yes vs. no)	**Figure 2 in study** (Nomogram for the risk of recurrence)	Not time-point specific	Derivation and internal validation study
HERDOO2	Rodger (2017)	Male sex: 2 points one of more of: hyperpigmentation, edema or redness of the leg: 1 point VIDAS D-dimer 250 µg/L: 1 point Obesity (body mass index 30 or more): 1 point	Low risk: 0-1 points High risk: 2+ points	12 months	External validation study

Tool	Studies evaluation this tool	Score	Interpretation	Key prediction time point	Derivation or validation study
		Age 65 year or older: 1 point			
VIENNA	Marcucci (2015)	Considers D-dimer (as a continuous variable), age and index VTE location (distal DVT versus proximal DVT/PE)	Online calculator	12, 60 months	External validation study
Dynamic VIENNA	Eichinger (2014)	Same as above but adjusted to account for time since discontinuing anticoagulation (baseline, 3, 9, 12 and 15 months).	Online calculator	12, 60 months	Modified validation study

Table 3 Tools to predict major bleeding

Tool	Included studies	Score	Interpretation	Derivation or validation study
ACCP	Palareti (2018) Poli (2013) Riva (2014)	age 66–75 year: 1 point age > 75 years: 2 points previous major bleeding: 1 point cancer (active): 1 point metastatic cancer: 1 point renal failure (CrCl < 30–60 mL min-1): 1 point liver failure (reported in the history): 1 point thrombocytopenia (< 100 000): 1 point previous stroke/TIA: 1 point diabetes anemia (Hb < 10): 1 point antiplatelet therapy: 1 point poor anticoagulant control (time spent in the therapeutic range < 60%): 1 point comorbidity recent surgery (within 3 months from index event): 1 point frequent falls (≥ 2 in the last year): 1 point alcohol abuse (reported in the history): 1 point	Low risk: 0 Moderate risk: 1 High risk: 2+	External validation studies

Tool	Included studies	Score	Interpretation	Derivation or validation study
		non-steroidal anti-inflammatory drugs use: 1 point		
ATRIA	Klok (2016) Poli (2013) Riva (2014)	Anemia: 3 points severe renal disease (GFR < 30 mL min-1 or dialysis-dependent): 3 points 75 years+: 2 points previous bleed: 1 point hypertension: 1 point	Low risk: 0-3 points Moderate risk: 4 points High risk: 5-10 points	External validation studies
EPIPHANY index	Carmona-Bayonas (2017)	Final model contained 6 variables: the Hestia-like CDR variable (any risk factor present vs none), ECOG-PS (o2 vs X2), oxygen saturation (o90 vs X90%), presence of PE-specific symptoms, previous tumour response evaluation (tumour progression, unknown, or not evaluated vs others), and prior surgical resection of primary tumour	Decision tree model	Derivation and internal validation study
HAS-BLED	Brown (2018) Klok (2016) Kooiman (2015) Kresoja (2019) Poli (2013) Riva (2014)	>160 mmHg systolic: 1 point Abnormal liver function (history of cirrhosis, or bilirubin > 2x the upper limit of normal in association with aspartate aminotransferase/alanine aminotransferase/ alkaline phosphatase levels > 3x the upper limit of normal): 1 point Abnormal renal function (on dialysis, a history of kidney transplantation, or serum creatinine values > 200 µmol/L): 1 point Stroke: 1 point Bleeding history or predisposition: 1 point Liable INR (time within therapeutic range < 60%): 1 point Age 65 years+: 1 point Drug use (platelet inhibitors or NSAIDS): 1 point Alcohol abuse (>8 units/week): 1 point	Low risk: 0 points Moderate risk: 1-2 points High risk 3+ points	External validation studies

Tool	Included studies	Score	Interpretation	Derivation or validation study
HEMORR2AGES	Klok (2016) Poli (2013) Riva (2014)	Hepatic or renal disease: 1 point Ethanol abuse: 1 point Malignancy: 1 point Age 75 years plus: 1 point Reduced platelet count or function: 1 point Prior bleeding: 2 points Hypertension: 1 point Anemia: 1 point Genetic factors (CYP2C9 single-nucleotide polymorphisms): 1 point Excessive fall risk: 1 point Prior stroke/TIA: 1 point	Low risk: 0-1 points Moderate risk: 2-3 points High risk >3 points	External validation studies
Kearon score	Scherz (2013) Seiler (2017) Kearon (2019)	Previous stroke: 1 point Previous peptic ulcer disease: 1 point Previous gastrointestinal bleeding: 1 point Renal impairment: 1 point Anemia: 1 point Thrombocytopenia: 1 point Liver disease: 1 point Diabetes mellitus: 1 point Use of antiplatelet therapy: 1 point Age 65 years+: 1 point	Low: 0–1 points Intermediate: 2–3 points High: 4+	External validation study
Kuijer score	Klok (2016) Kuijer (1999) Sherz (2013) Piovella (2014) Riva (2014) Seiler (2017) Zhang (2019)	Age >60 years: 1.6 points Female sex: 1.3 points Malignancy: 2.2 points	Low risk = 0 points Moderate risk = 1-2 points High risk = 3+ points	External validation studies

Tool	Included studies	Score	Interpretation	Derivation or validation study
Modified Ottawa Score (MOS)	Alatri (2017)	Female sex: 1 point Lung cancer: 1 point Prior VTE: 1 point Localized cancer without metastasis (stages 1 and 2 for solid tumors): -1 point Breast cancer: -1 point	Low: -1 or fewer points Intermediate: 0 points High 1+ points	External validation study
Nieuwenhuis score	Zhang (2019)	WHO grade 1: 1 point WHO grade 2: 1 point History of bleeding diathesis: 2 point Recent trauma (<2 months) or surgery: 1 point Body surfacy area <2m^2: 1 point	Low risk: 0-2 points Moderaterisk:3-4 points High risk: >5	External validation study
Outpatient bleeding risk index (OBRI)	Poli (2013) Scherz (2013) Riva (2014) Piovella (2014) Seiler (2017)	Age 65 years+: 1 point Previous stroke: 1 point GI bleeding in the last 2 weeks: 1 point Recent MI, anemia (hematocrit <30%) or renal insufficiency (creatinine >1.5mg/dl): 1 point Creatinine >1.5mg dL-1: 1 point diabetes mellitus: 1 point	Low risk: 0 points Moderate risk: 1-2 points High risk: 3+ points	External validation studies
RIETE 1	Klok (2016) Poli (2013) Ruiz-Gimenez (2008) Scherz (2013) Riva (2014) Piovella (2014) Seiler (2017) Zhang (2019)	Recent bleeding: 2 points Abnormal renal function (Creatinine levels >1.2 mg/dl): 1.5 points Anemia: 1.5 points Age >75 years: 1 point Active malignancy: 1 point PE diagnosis: 1 point	Low risk: 0 points Moderate risk: 1-4 points High risk >4 points	External validation study

Tool	Included studies	Score	Interpretation	Derivation or validation study
RIETE 2	Nieto (2013)	Age >75 years: 1 point Metastatic cancer: 2 point Immobility ≥4 days: 1 point Recent major bleeding: 1.5 points Abnormal prothrombin time: 1 point CrCl b 30 ml/min: 1 point Platelet Count <100 × 10^9/L: 1 point Anemia: 1 point Distal DVT: -1 point	Low risk: <1.5 points Moderate risk: 1.5 – 4.0 points High risk: >4.0 points	External validation study
Seiler score	Seiler (2017)	Previous major bleeding Active cancer Low physical activity Anemia Thrombocytopenia Antiplatelet drugs or NSAIDs Poor INR control	Low risk: 0-1 point Moderate risk:2-3 points High risk: 4+ points	Derivation and internal validation study
Shireman score	Riva (2014)	Age ≥ 70 years: 0.49 points Female gender: 0.32 points Remote bleeding: 0.58 points Recent bleeding: 0.62 points Alcohol/drug abuse: 0.71 points Diabetes: 0.27 points Anaemia (Ht < 30%): 0.86 points Antiplatelet use: 0.32 points	Low risk: ≤1.07 points Moderate: >1.07 and <2.19 points High: ≥2.19 points	External validation study
VTE-BLEED	Klok (2017) Klok (2018) Kresoja (2019)	Active cancer: 2 points Male with uncontrolled arterial hypertension: 1 points Anaemia: 1.5 points History of bleeding: 1.5 points Age ≥60 years old: 1.5 points Renal dysfunction: 1.5 points	Low risk: 0-1 high risk: 2+	External validation studies

Table 4 Study characteristics (VTE-recurrence)

Author (year)	Sample	Mean age (SD)	Demographic information	Tool(s) investigated	Duration
Eichinger (2014)	553	53.4 years **median age, mean not stated*	• Location: Austria • 39.6% female • 16.1% distal DVT, 38.5% proximal DVT, 45.4% PE with or without DVT • Median duration of AC treatment: 6.7 months	Dynamic VIENNA (internal validation only)	Median 68 months.
Franco (2016)	398	61.0 years **median, mean not stated*	• Location: Spain • 45.5% female • 67.8% PE. 20.8% proximal DVT. 11.3% distal DVT • Median duration of AC treatment: 7.0 months	DAMOVES (internal validation only)	Median 21.3 months
Marcucci (2015)	904	68 years **median age, mean not stated.*	• Location: Austria • 39.5% female • 32.2% PE with or without DVT, 66.5% proximal DVT, 1.3% distal DVT 13.6 % recurrent VTE	VIENNA tool	Median 22 months (25th, 75th percentiles: 14 months, 29 months)
Rodger (2017)	934	54.4 (SD 16.7) years.	• Location: 7 countries • 44% female • 40% index DVT only, 39% index PE only, 21% both • 48.9% D-dimer over 250 UG/L 37.2% obesity (over 30 BMI)	HERDOO2 tool	1-year

Tosetto (2017)	827	55.3 (SD17.5) years	• Location: Italy • 47.8% female • Mean duration of prior AC treatment 13 months Hormone therapy at time of index VTE: 43.0%	DASH	Median 25.2 months

Table 5 Study characteristics (major bleeding)

Author (year)	Sample	Mean age (SD)	Demographic information	Tool(s) investigated	Duration
Alatri (2017)	383	68 (14) years	**All participants had active cancer and VTE.** • Location: Switzerland • 44.9% metastatic disease • 60.8% curative treatment plan • 12.5% lung cancer • 11.2% breast cancer • 55.1% localised cancer • 18.8% had a prior VTE • 13.6% renal failure • 25.6% chemotherapy 9.1% radiotherapy	MOS	90 days (on-treatment only)
Brown (2018)	132,280	63.7% aged 18-64 12.7% aged 65-74 23.6% aged 75 years or older	• Location: USA • 56% female • 44.5 % hypertension • 7.7% renal disease • 5.7 % liver disease • 11.7% had bleeding history	HAS-BLED	Outcomes reported at 30, 60, 90 and 180 days (on-treatment only)

Carmona-Bayonas (2017)	1,075	64 (12) years	**Participants had acute PE and active cancer** • Location: Spain • 54.2% female • 11.7% breast cancer • 25.3% lung cancer • 11.9% COPD • 5.2% chronic CVD	EPIPHANY index (internal validation)	15 days
Klok (2016)	448	64 (17) years	**Participants had acute PE** • Location: Germany • 55% female • 51% unprovoked • 30% had a prior VTE • 59% hypertension • 6.7% known renal insufficiency • 2.5% known liver disease • 1.6% bleeding history • 18% active cancer • 18 % chronic heart failure	• HAS-BLED • RIETE score 1 • Kuijer score • HEMORR2HAGES • ATRIA score	30 days
Klok (2017)	8,240	55.8 (16) years	**All participants were part of the HOKUSAI-VTE trial** • Location: 37 countries • 40% PE and DVT, 60% DVT only • 26% arterial hypertension • 11% renal dysfunction • 1% history of bleeding • 2.5% active cancer	VTE-BLEED	12 months (on-treatment only)
Klok (2018)	3653	60 (17) years	**Participants had DVT** • Location: 21 countries • 90% DVT only; 9.8% both DVT and PE	VTE-BLEED	At least 12 months or until they died. (on-treatment only)

			• 46% female • 11% active cancer • Mean systolic blood pressure 137 mm HG (SD19), 2179 participant had missing data. • first eGFR reading <30 ml/min: 1.4% 30-50 ml/min: 5.0% 50 ml/min or greater 58% 1601 participants missing		
Kooiman (2015)	537	PE patients: 60.8 years DVT: 57.8 years	• Location: The Netherlands • 41.5% PE, 58.5% DVT • 31% hypertension • 3.5% bleeding history • 11.2% NSAIDS/ antiplatelet	HAS-BLED	6 months (on-treatment only)
Kresoja (2019)	522	69 (IQR 56-78)	**Study predicts in-hospital major bleeding in people with PE.** • 53% female • 62% unprovoked PE • 12% prior VTE • 3.7% liver disease • 4.7% bleeding history • 17% active cancer • 16% chronic heart failure	VTE-BLEED HAS-BLED	During in hospital treatment
Kuijer (1999)	780	60 (17) years	**Included derivation, internal validation and external validation: Only external validation study was extracted as other had <250 participants** • 75% DVT 25% PE • 49% female	Kuijer score	3 months (on-treatment only)
Nieto (2013)	15,206	35% over 75 years of age	• Location: 12 countries • 51% female	RIETE score 2	3 months (on-treatment only)

		*Mean not given	• 51% Bilateral PE; 3.8% bilateral DVT; 11% Distal DVT • 55% anemia • 2.5% low platelet count • 7.4% abnormal propthrombin time • 7.4% CrCl <30 ml/min		
Palareti (2018)	2,263	67 (IQR 51 - 77) years *Median; mean value not given	• Location: Italy • 48.7% female • 56.6% DVT only; 20.9% DVT + PE; 20.3% isolated PE • 68.2% unprovoked VTE 25.7% had CrCl<30-60 mL min-1 • 2.9% had history of liver failure 0.5% NSAID use • 4.0% antiplatelet use 3.5% Anemia 7.7 % active cancer	ACCP	Median 12 months (on-treatment only)
Piovella (2014)	8,576	66 (18) years	• Location: Spain, France, Italy, Israel, Argentina • 52% female	RIETE 1 Kuijer OBRI	90 days
Poli (2013)	1078	84 years (range 80-89 Mean age (SD)) * median, mean not given	**All participants were >80 years of age** • Location: Italy • 62.8% female • 62.9% hypertension • 5.1% NSAIDS/antiplatelet use • 10.1% active cancer	• HAS-BLED • OBRI • HEMORR$_2$HAGES • ATRIA • ACCP • RIETE 1	Mean 1.83 years (on-treatment only)
Riva (2013)	681	median 63 (IQR 46-74) years	• Location: Italy • 43% PE	• HAS-BLED • OBRI • HEMORR$_2$HAGES	1 year or until treatment stopped: Mean 8.8 months (on-treatment only)

			• 76% lower extremity DVT • 5% upper extermity DVT • 1.6% superficial vein thrombosis only. • 58% unprovoked VTE • 2.0% severe renal failure	• ATRIA • ACCP • RIETE 1 • Kuijer score	
Ruiz-Gimenez (2008)	6,572	66 (SD17) years	*Demographic information on validation study only (see appendix E for information on derivation study)* • Location: Spain, France, Italy, Israel, Argentina • 50% female 32% Anemia 20 % active cancer 28% inpatients 12% recent surgery 24% immobility 1.4% recent major bleeding 2.5% with platelet count <100,000/mm3 14% with abnormal creatinine levels	RIETE score 1	90 days (on-treatment only)
Scherz (2013)	663	75 (65-97) years **Median and range, mean not given*	**All participants were ≥ 65 years** • Location: Switzerland • 45.4% female • 32.4% DVT only, 13.7% DVT+PE, 53.8% PE only. • 17.8% had renal impairment 2.0 % liver disease 3.8% previous major bleed 32.3% used platelet inhibitors 38.3% Anemia 14.6 % active cancer	• RIETE score 1 • Kuijer score • Kearon score • OBRI	90 days (on-treatment only)

Seiler (2017)	743	75 years (IQR 70-81) **Median and range, mean not given*	**Participants were ≥65 years plus** **All participants had received 3 months AC at time of assessment.** • Location: Switzerland • 47% female • 18.3% arterial hypertension • 7.3% renal disease • 0.4% liver disease • 3.6% bleeding history • 3.1% active cancer	• RIETE score 1 • Kuijer score • Kearon score • OBRI • Seiler score (internal validation only)	Up to 36 months of extended treatment
Zhang (2019)	578	65 years (IQR 53-74) **Median and range, mean not given*	**PE-only** • 49.5% female • 7.8% bleeding history • 1.7% NSAIDS/ antiplatelet use • 12.4% active cancer	• RIETE score 1 • Kuijer score • Kearon score • Nieuwenhuis score	3 months

Quality assessment of clinical studies included in the evidence review

See evidence tables in appendix E for quality assessment of individual studies, appendix F for forest plots and appendix G for GRADE tables. Please refer to the evidence statement section for an overall summary of the evidence.

Economic evidence

A systematic search was carried out for this review question to identify relevant economic analyses. This search returned 2,535 records, all of which were excluded on title and abstract.

An additional search was conducted at the end of the guideline development process to capture economic evidence published while the guideline was being developed. This was conducted as a single re run search covering all questions in the guideline. This search returned 2,013 records in total, all of which were excluded on title and abstract for this review question.

Therefore, no published cost-effectiveness studies were included in this review and this question was not prioritised for *de novo* economic modelling.

Evidence statements

The format of the evidence statements is explained in appendix B. Evidence statements are presented for each tool in turn and are based on likelihood ratios, c-statistics and hazard or odds ratios.

Tools to predict VTE-recurrence

Evidence from external validation studies

DASH

Low to moderate quality evidence from one study with 827 participants found the DASH tool had **adequate** classification accuracy in predicting VTE recurrence in all participants (C-statistic = 0.65), **poor** classification accuracy in those participants over 65 years of age (C-statistic= 0.54), and **good** classification accuracy in those participants 65 years of age or younger (C-statistic= 0.72).

HERDOO2

Low quality evidence from one study with 914 participants found that a **low risk** classification indicated a **slight** decrease in the probability of a recurrent VTE (LR-= 0.61 [0.42, 0.89]) and a **high risk** classification indicated a **slight increase** in the probability of a recurrent VTE (LR+=1.74 [1.34, 2.27]).

VIENNA

Low to moderate quality evidence from one study with 904 participants found the VIENNA tool had **adequate** classification accuracy in predicting VTE recurrence in all participants (C-statistic = 0.63) and **poor** classification accuracy in those participants 65 years of age and younger (C-statistic = 0.59).

Evidence from internal validation studies (or modified tools)

DAMOVES

Moderate quality evidence from one internal validation study with 398 participants found that a **low risk** classification indicated a **large decrease** in the probability of a recurrent VTE (LR-=0.11 [0.05, 0.25]) and a **high risk** classification indicated a **large increase** in the probability of a recurrent VTE (LR+=5.04 [3.96, 6.43]).

Dynamic VIENNA (time modified VIENNA tool)

Moderate quality evidence from one internal derivation study with 553 participants found the dynamic VIENNA tool had **adequate** classification accuracy in predicting VTE recurrence in all participants, when assessed 3 weeks (C-statistic = 0.63), 3 months (AUC= 0.61) and 9 months (C-statistic = 0.61) after discontinuing anticoagulation, and **poor** classification accuracy in predicting VTE recurrence when assessed 15 months after stopping anticoagulation (C-statistic = 0.58).

Tools to predict major bleeding

Evidence from external validation studies (or models with optimised thresholds)

ACCP

The ACCP tool had **poor** classification accuracy for predicting major bleeding up to 12 months, overall (C-statistic = 0.57 (0.49, 0.64, very-low quality evidence from two studies with 2,994 participants) and specifically in people aged 80 years or older (C-statistic = 0.55 [0.45, 0.64], based on very low quality evidence from one study with 1,078 participants).

ATRIA

At 30 days of treatment, very low to low quality evidence from one study with 448 participants found that:

- a **low risk** classification was associated with a **slight decrease** in the probability of major bleeding (LR-= 0.68 [0.44, 1.06]) and a **low-moderate risk** classification was associated with a **slight decrease** in the probability of a major bleeding (LR-= 0.94 [0.75, 1.18])
- a **moderate-high risk** classification was associated with a **slight increase** in the probability of major bleeding (LR+=1.89 [1.19, 3.02]) and a **high risk** classification was associated with a **slight increase** in the probability of a major bleeding (LR+=1.34 [0.54, 3.31])

The ATRIA tool had **poor** classification accuracy for predicting major bleeding up to 12 months, overall (C-statistic = 0.47 [0.31, 0.63], very low quality evidence from one study with 681 participants) and specifically in people aged 80 years or older (C-statistic = 0.58 [0.48, 0.67], very low quality evidence from one study with 1,078 participants).

Two-level optimised threshold ATRIA

At 30 days of treatment, very low to low quality evidence from one study with 448 participants that used an optimised threshold ATRIA tool found that a **low risk** classification was associated with a **slight decrease** in the probability of a major bleeding (LR-= 0.53 [0.25, 1.13]) and a **high risk** classification was associated with a **slight increase** in the probability of major bleeding (LR+=1.43 [1.09, 1.87]).

HAS-BLED

At 30 days of treatment, very low to low quality evidence from one study with 448 participants found that:

- a **low- moderate risk** classification was associated with a **slight increase in** the probability of major bleeding (LR-= 1.17 [0.72, 1.89]), and a **low risk** classification was associated with a **large decrease** in the probability of major bleeding (LR-=0.14 [0.01, 2.26])
- a **moderate-high risk** classification was associated with a **slight increase** in the probability of a major bleeding (LR+=1.17 [1.12, 1.22]) and a **high risk** classification was associated with a **slight decrease** in the probability of major bleeding (LR+=0.90 [0.62, 1.30])

Up to 6 months of treatment, very low to low quality evidence from one study with 537 participants found that:

- a **low-moderate risk** classification (using the standard 3+ cut-off) was associated with a **slight decrease** in the probability of major bleeding (LR-= 0.52 [0.27, 1.00]) and a **low-moderate** classification (using a modified cut off of 4+ points) was associated with a **slight decrease** in the probability of major bleeding (LR-= 0.84 [0.63, 1.11])
- a **high risk** classification (using the standard 3+ cut-off) was associated with a **moderate increase** in the probability of major bleeding (LR+=4.22 [2.35, 7.56]) and a **high classification** (using a modified cut off of 4+ points) was associated with a **large increase** in the probability of major bleeding (LR+=7.97 [2.02, 31.45])

The HAS-BLED tool had a **poor** classification accuracy for prediction major bleeding during in hospital stay (C-statistic = 0.58 [0.48, 0.69], very low quality evidence from one study with 522 participants), **good** classification accuracy up to 6 months (C-statistic = 0.71 (0.70, 0.72, moderate quality evidence from two studies with 132,817 participants) and **adequate** classification accuracy up to 12 months (C-statistic = 0.60 [0.45, 0.75], very-low quality evidence from one study with 681 participants).

In people with cancer, the HAS-BLED tool had **adequate** classification accuracy for predicting major bleeding (C-statistic = 0.69 [0.67, 0.71], low quality evidence from one study with 24,915 participants) and in people aged 80 years or older, it had **poor** classification accuracy (C-statistic = 0.55 [0.46, 0.64], based on very low quality evidence from one study with 1,078 participants).

Very low quality evidence from one study with 522 participants **could not differentiate** rates of major bleeding during hospital stay in people with low compared to high classification accuracies (OR of 1.10 [0.40, 2.90]).

Moderate quality evidence from one study with 132,280 participants found the HAS-BLED high risk group (category 4) had an **increased risk** of major bleeding (HR of 1.71 [1.44, 2.04]) compared to category 3.

Low quality evidence from one study with 537 participants found the HAS-BLED high risk group had an **increased risk** of major bleeding using a cut-off of 3+ points (HR of 8.7 [2.7, 28.4]) and using a cut-off of 4+ points (HR of 10.8 [2.3, 50.0]) compared to the moderate/low risk group.

Moderate quality evidence from one study with 24,915 people with cancer found the HAS-BLED high risk group (category 4) had an **increased risk** of major bleeding (HR 1.66 [1.26, 2.20]) compared to category 3.

HEMORR$_2$HAGES

At 30 days of treatment, very low quality evidence from one study with 448 participants found that:

- a **low risk** classification was associated with a **slight decrease** in the probability of major bleeding (LR-= 0.75 [0.38, 1.47]) and a **low-moderate risk** classification was associated with a **slight decrease** in the probability of major bleeding (LR-= 0.99 [0.82, 1.19])
- a **moderate-high risk** classification was associated with a **slight increase** in the probability of major bleeding (LR+=1.17 [0.87, 1.58]) and a **high risk** classification was associated with a **slight increase** in the probability of major bleeding (LR+=1.07 [0.37, 3.12])

The HEMORR2HAGES tool had **poor** classification accuracy for predicting major bleeding up to 12 months, overall (C-statistic = 0.51 [0.32, 0.70], very low quality evidence from one study with 681 participants) and **adequate** classification accuracy specifically in people aged 80 years or older (C-statistic = 0.60 [0.50, 0.70], for predicting major bleeding based on very low quality evidence from one study with 1,078 participants).

Two-level optimised threshold HEMORR$_2$HAGES

At 30 days of treatment, very low quality evidence from one study with 448 participants that used an optimised threshold HEMORR$_2$HAGES tool found that a **low risk** classification was associated with a **slight decrease** in the probability of a major bleeding (LR-= 0.60 [0.35, 1.03]) and a **high risk** classification was associated with a **slight increase** in the probability of major bleeding (LR+=1.81 [1.23, 2.65]).

Kearon score

At 3 months of treatment, very low to low quality evidence from one study with 663 participants 65 years of age or older found that a **low-moderate risk** classification was associated with a **slight decrease** in the probability of major bleeding (LR-= 0.95 [0.78, 1.15]) and a **moderate-high risk** classification was associated with a **slight increase** in the probability of major bleeding (LR+=1.26 [0.61, 2.61]).

In participants aged 65 years or older, the Kearon tool had:

- **poor** classification accuracy in predicting major bleeding (C-statistic = 0.59 [0.55, 0.63]), at 3 months based on low quality evidence from one study with 663 participants
- **poor** classification accuracy in predicting major bleeding at 3 months (C-statistic = 0.54 [0.38, 0.69]), 6 months (C-statistic = 0.55 [0.43, 0.67]), 12 months (C-statistic = 0.58 [0.49, 0.67]), 24 months (C-statistic = 0.57 [0.50, 0.64]) and at 36 months (C-statistic = 0.59 [0.52, 0.66]) based on moderate quality evidence from one study with 743 participants who had already received 3 months of anticoagulation

At 3 months of treatment, very low quality evidence from one study with 537 participants found the kearon score to have **good** classification accuracy in predicting major bleeding (C-statistic = 0.75 [0.60, 0.89]).

Kuijer score

At 30 days of treatment:

- a **low-moderate risk** classification was associated with a **slight decrease** in the probability of major bleeding (LR-= 0.94 [0.75, 1.18]) and a **low risk** classification was

associated with a **slight decrease** in the probability of major bleeding (LR-= 0.73 [0.19, 2.76], very low to low quality evidence from one study with 448 participants).
- a **high risk** classification was associated with a **slight increase** in the probability of major bleeding (LR+=1.35 [0.54, 3.31]) and a **moderate-high risk** classification was associated with a **slight increase** in the probability of major bleeding (LR+=1.04 [0.90, 1.21], very low to low quality evidence from one study with 448 participants).

At 3 months of treatment:

- a **low risk** classification was associated with a **large decrease** in the probability of major bleeding (LR- 0.15 [0.04, 0.58], based on very low quality evidence from two studies with 9,497 participants).
- a **low-moderate** risk classification was associated with a **slight decrease** in the probability of major bleeding, overall (LR-= 0.81 [0.47, 1.39], very-low quality evidence from two studies with 9,497 participants) and **a slight increase** specifically in people aged 65 years or older (LR-= 1.04 [0.91, 1.19], low quality evidence from one study with 663 participants).
- a **moderate-high** risk classification was associated with a **slight increase** in the probability of major bleeding (LR+ 1.15 [1.10, 1.20], low quality evidence from two studies with 9,497 participants).
- a **high risk** classification was associated with a **slight increase** in the probability of major bleeding, overall (LR+=1.59 [0.46, 5.55], very low quality evidence from two studies with 9,497 participants) and **a slight decrease** specifically in people aged 65 years or older (LR+=0.76 [0.26, 2.27], very low quality evidence from one study with 663 participants.

The kuijer score had **poor** classification accuracy in predicting major bleeding at 3 months, overall (C-statistic = 0.57 [0.44, 0.68], very low quality evidence from one study with 537 participants) and specifically in people aged 65 years or older (C-statistic = 0.49 [0.45, 0.52], moderate quality evidence from one study with 663 participants), and at 12 months (C-statistic = 0.51 [0.32, 0.70], very low quality evidence from one study with 681 participants).

In people aged 65 years or older who had already received 3 months of anticoagulation, the kuijer score had **adequate** classification accuracy in predicting major bleeding at 3 months (C-statistic = 0.67 [0.54, 0.81]) and 6 months (C-statistic = 0.61 [0.50, 0.72]), and to have **poor** classification accuracy at 12 months (C-statistic = 0.57 [0.48, 0.66]), 24 months (C-statistic = 0.57 [0.50, 0.64]) and at 36 months (C-statistic = 0.57 [0.50, 0.64]). Based on low to moderate quality evidence from one study with 743 participants

Two-level optimised threshold Kuijer score

At 30 days of treatment, very low quality evidence from one study with 448 participants that used an optimised threshold Kuijer tool found that a **low risk** classification was associated with a **slight decrease** in the probability of a major bleeding (LR-= 0.66 [0.36, 1.21]) and a **high risk** classification was associated with a **slight** increase in the probability of major bleeding (LR+=1.38 [0.98, 1.93]).

Modified Ottawa Score (MOS)

In people with cancer-associated VTE at 3 months of treatment, a **low risk** classification was associated with a **slight decrease** in the probability of major bleeding (LR-= 1.00 [0.72, 1.40]) and a **high risk** classification was associated with a **slight decrease** in the probability of major bleeding (LR+=0.99 [0.43, 2.31]), based on very low to low quality evidence from one study with 383 participants.

Nieuwenhuis score

At 3 months of treatment, very low quality evidence from one study with 537 participants found the Nieuwenhuis score to have **poor** classification accuracy in predicting major bleeding (C-statistic = 0.59 [0.41, 0.74]).

Outpatient bleeding risk index (OBRI)

At 3 months of treatment, moderate quality evidence from one study with 8,717 participants found that:

- a **low-moderate risk** classification was associated with a **slight decrease** in the probability of major bleeding (LR-= 0.96 [0.91, 1.02]) and a **low risk** classification was associated with a **moderate decrease** in the probability of major bleeding (LR-= 0.36 [0.19, 0.69])
- a **high risk** classification was associated with a **slight increase** in the probability of major bleeding (LR+=1.77 [1.00, 3.43]) and a **moderate-high risk** classification was associated with a **slight increase** in the probability of major bleeding (LR+=1.24 [1.15, 1.33]).

At 3 months of treatment in people of 65 years of age or older, very low to low quality evidence from one study with 663 participants found that a **low-moderate** risk classification was associated with a **slight decrease** in the probability of major bleeding (LR-= 1.00 [0.90, 1.11]) and a **high risk** classification was associated with a **slight increase** in the probability of major bleeding (LR+=1.03 [0.26, 4.04]).

At 12 months of treatment, very low quality evidence from one study with 681 participants found the OBRI tool had **poor** classification accuracy in predicting major bleeding (C-statistic = 0.59 [0.42, 0.76]).

At 3 months of treatment in people of 65 years of age or older, moderate quality evidence from one study with 663 participants found the OBRI tool had **poor** classification accuracy in predicting major bleeding (C-statistic = 0.54 [0.50, 0.58]).

In people aged 65 years or older and who have already received 3 months of anticoagulation, moderate to high quality evidence from one study with 743 participants found the OBRI tool had **poor** classification accuracy in predicting major bleeding at 3 months (C-statistic = 0.54 [0.43, 0.65]), 6 months (C-statistic = 0.51 [0.42, 0.60]), 12 months (C-statistic = 0.52 [0.44, 0.60]), 24 months (C-statistic = 0.52 [0.46, 0.59]) and at 36 months (C-statistic = 0.53 [0.47, 0.59])

In people aged 80 years or older, very low quality evidence from one study with 1,078 participants found the OBRI had **poor** classification accuracy in predicting major bleeding (C-statistic = 0.58 [0.49, 0.67]).

RIETE 1

At 30 days of treatment:

- a **low-moderate risk** classification was associated with a **slight decrease** in the probability of major bleeding (LR-= 1.00 [0.86, 1.16], low quality evidence from one study with 448 participants)
- a **high risk** classification was associated with a **slight increase** in the probability of a major bleeding (LR+=1.04 [0.27, 3.98], very low quality evidence from one study with 448 participants

At 3 months of treatment:

- a **low risk** classification was associated with a **large decrease** in the probability of major bleeding based on low quality evidence from two studies with 15,289 participants (LR-= 0.13 [0.01, 1.79]).
- a **low-moderate risk** classification was associate with a **slight decrease** in the probability of major bleeding, overall (LR-= 0.91 [0.86, 0.96], moderate quality evidence from two studies with 15,289 participants) and specifically in people aged 65 years or older (LR-= 0.94 [0.81, 1.10], low quality evidence from one study with 663 participants.
- a **moderate-high risk** classification was associated with a **slight increase** in the probability of a major bleeding based on moderate quality evidence from two studies with 15,289 participants (LR+= 1.15 [0.97, 1.37]).
- a **high risk** classification was associated with a **moderate increase** in the probability of a major bleeding, overall (LR+=2.27 [1.62, 3.17], low quality evidence from two studies with 15,289 participants) and a **slight increase** a slight increase specifically in people aged 65 years or older (LR+=1.59 [0.62, 4.08], very low quality evidence from one study with 663 participants)

The RIETE 1 score had **poor** classification accuracy in predicting major bleeding at 3 months of treatment, overall (C-statistic = 0.56 [0.45, 0.71], very low quality evidence from one study with 537 participants) and **adequate** classification accuracy specifically in people aged 65 years or older (C-statistic = 0.60 [0.56, 0.64], low quality evidence from one study with 663 participants). At 12 months, it had **poor** classification accuracy (C-statistic = 0.54 [0.38, 0.71], very low quality evidence from one study with 681 participants.

In people aged 65 years or older who had already received 3 months of anticoagulation treatment, the RIETE 1 tool had **poor** classification accuracy in predicting major bleeding at 3 months (C-statistic = 0.59 [0.45, 0.72]) and 6 months (C-statistic = 0.59 [0.48, 0.70]), and **adequate** classification accuracy at 12 months (C-statistic = 0.63 [0.54, 0.72]), 24 months (C-statistic = 0.62 [0.55, 0.69]) and at 36 months (C-statistic = 0.63 [0.56, 0.70]) (low to moderate quality evidence from one study with 743 participants.

In people aged 80 years or older, the RIETE 1 tool had **adequate** classification accuracy in predicting major bleeding (C-statistic = 0.61 [0.51, 0.71], very low quality evidence from one study with 1,078 participants).

Two-level optimised threshold RIETE 1

At 30 days of treatment, a **low risk** classification was associated with a **slight decrease** in the probability of a recurrent VTE (LR-= 0.59 [0.30, 1.16]) and a **high risk** classification associated a **slight increase** in the probability of major bleeding (LR+=1.42 [1.05, 1.92], very low to low quality evidence from one internal validation study with 448 participants).

RIETE 2

At 3 months of treatment:

- a **low risk** classification was associated with a **moderate decrease** in the probability of **fatal bleeding** (LR-= 0.30 [0.17, 0.52], moderate quality evidence from one study with 15,206 participants).
- a **low-moderate risk** classification was associated with a **slight decrease** in the probability of **fatal bleeding** (LR-= 0.92 [0.85, 1.01], moderate quality evidence from one study with 15,206 participants).

- a **moderate-high risk** classification was associated with a **moderate increase** in the probability of **fatal bleeding** (LR+=2.26 [1.98, 2.58], moderate quality evidence from one study with 15,206 participants).
- a **high risk** classification was associated with a **moderate increase** in the probability of **fatal bleeding** (LR+=4.27 [1.85, 9.90], moderate quality evidence from one study with 15,206 participants).

At 3 months of treatment, moderate quality evidence from one study with 15,206 participants found the RIETE 2 tool had **good** classification accuracy in predicting **major bleeding** (C-statistic = 0.72 [0.69, 0.75]) and **fatal bleeding** (C-statistic= 0.78 [0.72, 0.83]).

Shireman score

At 3 months of treatment, Shireman tool had **adequate** classification accuracy in predicting major bleeding (C-statistic = 0.63 [0.45, 0.81], very low quality evidence from one study with 681 participants.

VTE-BLEED

In-hospital treatment for PE:

- a **low risk** classification was associated with a **moderate decrease** in the probability of major bleeding (LR-= 0.26 [0.07, 0.97], very low quality evidence from one study with 522 participants).
- a **high risk** classification was associated with a **slight increase** in the probability of a major bleeding (LR+=1.55 [1.29, 1.86], low quality evidence from one study with 522 participants).

After up to 12 months of treatment in people who had already received 30 days of anticoagulation:

- a **low risk** classification was associated with a **slight decrease** in the probability of major bleeding (LR-= 0.57 [0.45, 0.72], low quality evidence from two studies with 12,249 participants).
- a **high risk** classification was associated with a **slight increase** in the probability of a major bleeding (LR+=1.96 [1.50, 2.56], low quality evidence from two studies with 12,249 participants.

During in-hospital treatment for PE, VTE-BLEED had **good** classification accuracy in predicting major bleeding based on very low quality evidence from one study with 522 participants (C-statistic = 0.69 [0.58, 0.80].

After up to 12 months of treatment in people who had already received 30 days of anticoagulation, VTE-BLEED had **adequate** classification accuracy in predicting major bleeding (C-statistic = 0.67 [0.62, 0.71], low quality evidence from two studies with 12,687 participants).

People scoring ≥2 compared to <2 using VTE-BLEED had an **increased risk** of major bleeding during in-hospital treatment for PE (1.30 [1.00,1.70], Low quality evidence from one study with 522 participants.

People who had already received 30 days of anticoagulant treatment scoring ≥2 on VTE-BLEED had an **increased risk** of major bleeding compared to those with a score of under 2 (HR 2.30 [1.10,4.50]). In these people, there was a per point increase OR 1.40 (1.10, 1.60). When the population was restricted to people with unprovoked VTE, the test **could not detect a difference** in risk of major bleeding between groups (HR 1.70 ([0.69, 4.40]). In

these people, there was a per point increase OR 1.30 (1.00, 1.70) (Based on very low to low quality evidence from one study with 4,447 participants).

Evidence from internal validation studies

EPIPHANY index

At 15 days in people with cancer, a **low risk** classification was associated with a **large decrease** in the probability of major bleeding (LR-= 0.19 [0.06, 0.57]) and a **high risk** classification was associated with a **slight increase** in the probability of major bleeding (LR+=1.34 [1.24, 1.45])(very low to low quality evidence from one internal validation study with 1,075 participants)

Seiler score

In people aged 65 years or older who had already received 3 months of anticoagulation treatment the Seiler score had **good** classification accuracy in predicting major bleeding at 3 months (C-statistic = 0.75 [0.61, 0.88]) and **adequate** classification accuracy at 6 months (C-statistic = 0.69 [0.58, 0.79]), 12 months (C-statistic = 0.68 [0.60, 0.76]), 24 months (C-statistic = 0.67 [0.60, 0.73]) and at 36 months (C-statistic = 0.68 [0.61, 0.74]) (very low to low quality evidence from one internal validation study with 743 participants).

Economic evidence statements

No relevant economic evidence was identified for this review question.

The committee's discussion of the evidence

Interpreting the evidence

The outcomes that matter most

The included studies reported data on sensitivity/specificity and likelihood ratios; C-statistics; odds ratios and hazard ratios. The committee agreed that, since this review is looking at the risk of VTE recurrence and of major bleeding, the most useful outcome measure was that of likelihood ratios. These give a measure of the size of increased or decreased probability of having an event in the future following a positive of negative test result and therefore indicate how useful the prediction tools will be for decision making. C-statistics were also thought to be useful as they allow for easier comparison of overall diagnostic accuracy between prediction tools (as they are reported using a single estimate, however these are less useful in the context of this review question as they do not differentiate between FP and FN. As FP and FN results have different clinical outcomes, it is important that these are considered separately. The committee noted that for two level tests (such as VTE-BLEED),likelihood ratios are particularly useful but for tests with 3 (or more) levels (such as HAS-BLED) there is a greater difficulty in interpreting likelihood ratio data as the different levels must be groups into two levels (a negative and a positive group) and this involves grouping the results in ways not intended by the tests. In these cases, it is useful to have additional measures of diagnostic accuracy (such as c-statistics).

Treatment for VTE can be divided into several stages: immediate treatment upon diagnosis followed by initial treatment that lasts several months and then extended treatment. Following initial treatment, the committee agreed that it was important to review the individual's risk of VTE recurrence before deciding to offer them extended treatment. However, there is a trade-off between managing the risk of VTE recurrence if treatment is

discontinued and that of major bleeding if treatment is continued. The committee discussed the different consequences of TP, TN, FP and FN results in this decision-making process.

VTE-Recurrence

VTE-recurrence is a key outcome for people with VTE because it negatively impacts quality of life, is potentially fatal and can lead to other harmful conditions such as post-thrombotic syndrome and chronic thromboembolic pulmonary hypertension.

The committee agreed that:

- TPs will be correctly identified as being at high risk of VTE recurrence and will be offered continued anticoagulant treatment to prevent this happening. They will be at risk of major bleeding due to the anticoagulation treatment.
- TNs will be correctly identified as being at low risk of a VTE recurrence and could potentially come off anticoagulation medication, thus avoiding unnecessary treatment that could lead to major bleeding.
- FPs will be incorrectly identified as being at high risk of VTE recurrence and will be offered unnecessary treatment which exposes them to an increased risk of major bleeding.
- FNs will be incorrectly identified as low risk of VTE recurrence and could potentially come off anticoagulation medication. This could lead to an increase in the numbers of people with VTE recurrence and may have serious consequences for their health (as discussed above). However, the risk of major bleeding would be decreased if the person came off the anticoagulation medication.

The committee agreed that it was most important to avoid FN results as the lack of treatment could lead to VTE recurrence and this is potentially fatal. However, the committee noted that recurrent PE is more immediately dangerous than recurrent DVT due to the risk of death. However, recurrent DVT is still of crucial importance as this leads to morbidities and can lead to a PE. A person with a recurrent VTE event will typically require anticoagulation treatment again. The committee agreed that it was also important to minimise FP results as these people would be exposed to an increased risk of major bleeding (which may cause death or severe morbidity).

Major bleeding

Major bleeding is another key outcome for people with VTE as it can also be fatal or seriously disabling, particularly when it is intracranial. In addition, major bleeding has deleterious effects on quality of life and may result in an individual stopping treatment, exposing them to a risk of VTE-recurrence.

The committee agreed that:

- TPs will be correctly identified as being at high risk of major bleeding and may be able to avoid long term anticoagulant treatment (depending on their risk of VTE recurrence) and thus avoid a major bleed. If they need to remain on anticoagulants due to a high risk of VTE recurrence they could receive additional monitoring for major bleeding.
- TNs will be correctly identified as being at low risk of a major bleeding and can receive extended treatment for VTE if this is indicated based on their risk of VTE recurrence. The extended anticoagulant treatment may reduce their risk of recurrence.
- FPs will be incorrectly identified as being at high risk of major bleeding, which is likely to be stressful for the individual with VTE. If they are also at high risk of VTE recurrence this may result in a difficult decision for the clinician and person with VTE. In addition, if the person with VTE stops taking the anticoagulant treatment due to the perceived high risk of major bleeding this could result in that individual having an increased risk of VTE-

recurrence, which poses a major threat to health. If they remain on treatment they could receive additional, unnecessary monitoring for major bleeding, which as well as being stressful for the person with VTE, could use up resources that could be better used elsewhere. However, if the person with VTE has a low risk of VTE recurrence, the cessation of treatment may not be problematic.

- FNs will be incorrectly identified as low risk of major bleeding and may be offered extended treatment (depending on their risk of VTE recurrence). This could lead them to experience a major bleed, which can lead to death or extreme morbidity.

The committee agreed that FNs and FPs are both problematic, but they agreed that in the case of major bleeding, a FP diagnosis could lead to the worst outcome for the person with VTE.

The quality of the evidence

This review included studies attempting to predict an individual's risk of recurrence (if they were to stop treatment) and risk of major bleeding (if they were to continue treatment) using risk assessment tools. Studies reporting prognostic risk tools can be divided into derivation, derivation with internal validation, external validation and model modification studies). The committee agreed that it was appropriate to exclude derivation studies from this review in cases where an external validation study was available for a given tool because derivation studies have higher risk of bias due to the model being developed based on a particular dataset. As a result, the evidence from external validation studies (or validation data from within derivation studies) was prioritised. However, where validation studies were not available, the committee agreed that looking at derivation studies was potentially informative although they noted that there was more uncertainty around the performance of tools that had not been externally validated (e.g. Seiler score, EPIPHANY index) in a random sample of people with VTE. The committee agreed that due to the high risk of bias, no recommendations could made be supporting the use of tools that had not been externally validated, although these tools could be the subject of research recommendations.

The committee discussed the very low quality of the body of evidence. In particular, they commented on the fact that many of the studies involved retrospective reviews of registry databases and that this poses a risk of bias due to the potential for misreporting / miss-recording information on predictor variables and outcomes.

The committee agreed with the decision to apply GRADE to likelihood ratios (as opposed to sensitivity/specificity outcomes) as they considered these more useful in the context of prediction as outcomes can be easily classified into groups according to strength (how much a positive or negative result changes the likelihood of experiencing the event in the future).

VTE-recurrence

The committee were concerned that only single studies were available for each tool predicting VTE-recurrence. Additionally, they noted that the studies assessing tools that predict VTE-recurrence limited participants to those with unprovoked VTE. These individuals are at an increased risk of recurrence compared to those with provoked VTE as the latter's risk drops once the provoking risk factor is removed.

The quality of the evidence was mainly moderate due to issues with risk of bias of included studies or the lack of a confidence interval to accompany c-statistic results. There was limited data available that could be used to calculate sensitivity and specificity and the bulk of the data was presented as C-statistics, which do not allow differentiation between the numbers of misclassified people (FP and FN). However, the committee agreed that FNs were at greater risk of adverse outcomes (see discussion above). Additionally, the lack of c-statistic

CIs meant that there was uncertainty surrounding these point estimates.

Major bleeding

The committee agreed that a major bleeding risk tool would be used to assess the risk of major bleeding in an individual were they continue anticoagulation treatment (typically after 3 months initial treatment) because at the initial treatment phase the risk of major bleeding would be less than the risk of VTE if no treatment was offered. However, the majority of included studies looked at the risk of major bleeding in treatment naïve individuals (or individuals beginning anticoagulation for an index VTE). Only one study (Seiler, 2017) assessed the prognostic accuracy of tools in participants who have already received at least 3 months anticoagulation. The committee agreed that although this data is not ideal as the risk of major bleeding could decrease over time, it was unlikely that the tool would perform differently in detecting that risk and agreed with the inclusion of these studies in the evidence base.

Bleeding risk was mainly related to treatment on VKA, however some of the later trials had a subgroup of their study population receiving DOACs. For example, the population in the Brown (2018) HAS-BLED validation study consisted of people taking VKA, LMHW or a DOAC. However, even if this was not the case the committee noted that the DOACs were designed to emulate VKAs in the dosing range of INR 2.0 to 3.0. Therefore, it is reasonable to extrapolate the use of tools developed with people taking LMWH or VKA to those taking DOACs. The committee agreed that although the bleeding risk was not the same for VKAs and DOACs, this should not affect the predictive accuracy of the tools, particularly in the long term as the risk of major bleeding stabilises.

The committee noted that many risk tools for predicting major bleeding divided people into low, moderate and high risk groups and that this was problematic when it came to calculating 2x2 tables to determine sensitivity, specificity and likelihood ratios. The committee agreed with the approach taken to convert three level tools into two levels by combining the moderate group with either the low or high risk group and presenting the results for both combinations. The committee were concerned that this approach may violate the original intention of the model, but they agreed that this method was preferable to other options (such as optimised two-level cut-offs) and the results at the extremes of the test (for people in high or low risk groups) could provide information about the risks of stopping or continuing treatment.

The committee agreed that the RIETE 2 paper on fatal bleeds was appropriate to include and discuss, considering fatal bleeds as a sub-category of major bleeds, and taking into account that the study also included a c-statistic for major bleeding.

Benefits and harms

This review aimed to identify the optimum duration of treatment for people with VTE. Initially, the balance of benefits and harms weighs heavily in favour of anticoagulant treatment, due to the high risk of VTE progression. At later time points (after 3 months or more of treatment), the trade-off between benefits and harms becomes more uncertain and the need for secondary prevention must be balanced against the risk of bleeding associated with continued treatment. The committee agreed that it was important to discuss the benefits and harms of treatment with the person with VTE after the initial treatment phase because of the decreased certainty in the benefit of treatment over the risks of major bleeding. They noted that it was important to involve the person with VTE in the decision-making process in general, but especially in cases where the risk of VTE recurrence and major bleeding were similar and patient preference was more likely to be the deciding factor.

In addition to major bleeding, a person undergoing anticoagulation may experience clinically relevant non-major bleeding, which can be distressing and have an impact upon quality of life. This risk, and the attitude of the person with VTE towards risk in general should also be taken into account when making joint decisions about continued treatment. The committee stressed the importance of not presupposing that the default option was continued treatment. The committee wrote a recommendation for this discussion to take place after 3 months of initial anticoagulant treatment (please see evidence review D for more details about the committee's rationale for making this recommendation, including information about the cross-references to other relevant NICE guidelines.) This conversation might involve the use of decision aids to ensure that the person with VTE is able to understand the risks and benefits associated with extended treatment and that they can make an informed decision about continuing or stopping.

The committee noted that people with an unprovoked VTE are at higher risk of recurrence compared to those with a provoking risk factor whose risk of recurrence subsides when the provoking factor is removed. The committee agreed that because of this lower risk of recurrence, people with provoked VTE could come off anticoagulation completely if they had a simple disease course. They wrote a recommendation to reflect this. They also recommended that these people should be provided with written information about what to do and who to contact if they experience important signs/symptoms. The committee agreed that this was important to ensure that people stopping treatment are as confident as possible in their ability to identify and act on any developments and that they have a specific port of call for any concerns they have. In contrast, people with unprovoked VTE have a higher risk of recurrence and are more likely to continue treatment, but this decision requires consideration of bleeding risk and individual preferences as well as the risk of recurrence.

The committee were aware of a systematic review with a meta-analysis suggesting a 10% risk of VTE-recurrence in the first year after stopping treatment, increasing to 16% at 2 years, 25% at 5 years and 36% at 10 years in people with unprovoked VTE (Khan, 2017). This study also identified that around 4% of recurrent events resulted in death. This work provided additional support for their decision to recommend that there is a discussion around continuing treatment for people with unprovoked VTE beyond 3 months due to their increased risk of recurrence. In addition, they agreed that if the person with unprovoked VTE had a low risk of bleeding then it was advisable for them to remain on treatment.

The committee discussed how the risk of VTE recurrence and major bleeding could be determined for the individual. The committee noted that, with the exception of DASH in people aged 65 years or younger, all of the tools for VTE recurrence had poor to adequate classification accuracy, poor discrimination based on likelihood ratios or lacked external validation. The committee were concerned with the quality of the evidence for current tools aimed at predicting VTE-recurrence, specifically c-statistics lacking confidence intervals and the limited numbers of studies. Additionally, the committee noted that use of HERDOO2 in clinical practice would be limited to women as men are automatically classified as high risk. The committee noted that many of the major bleeding prediction tools had poor classification accuracy or poor likelihood ratios (indicating slight changes in probability of major bleeding) and were able to rule out the use of many tools as a result. Finally, for both types of prognostic tool it was unclear whether the risk tools performed better than clinician judgement.

The committee agreed that there was a potential for harm associated with recommending the use of prediction tools that are not sufficiently validated or prognostically accurate. Taking these issues into account, the committee included a recommendation to make it clear that the tools should not be used in isolation, but rather as part of a discussion that involved clinician judgement of risk of recurrence and major bleeding if they are to be used at all.

The committee noted that the diagnostic accuracy evidence for DASH specifically in people who are 65 years of age or younger had good classification accuracy. However, the committee were again concerned with the quality of this evidence because it was from a single retrospective study which did not report confidence intervals for the C-statistic estimate, leaving uncertainty surrounding the precision of this estimate. Based on these issues the committee felt that they could not recommend a prediction tool to help assess the risk of VTE recurrence.

The committee also noted that various individual factors are important for predicting VTE recurrence, including whether the index event was provoked or unprovoked, and age and sex. However, the committee noted that a much larger number of individual characteristics may be important in the context of predicting the risk of recurrence (or major bleeding) and that these may not be captured well in a prediction tool.

Of all of the tools examined that aim to predict risk of major bleeding, the committee identified HAS-BLED as having potential clinical utility under certain conditions. A score of 4+ has a likelihood ratio indicating a large increase in the probability of a major bleed whereas a score of 3+ was associated with a moderate increase in risk; the tool had good classification accuracy up to 6 months and the risk of bleeding was 8.7 times higher in high versus moderate/low risk groups (HR 8.7 [2.7, 28.4]). The committee also discussed the evidence base for HAS-BLED, noting that it consisted of with one study (Brown 2018) presenting data on over 100,000 participants.

Due to the risk of VTE recurrence associated with stopping anticoagulation, the committee felt that a tool should only be recommended if it could identify those at a particularly high or low risk of bleeding accurately and that recommending a HAS-BLED score of 4+ would provide enough certainty to warrant stopping anticoagulation. They noted that if this risk was modifiable (for example, if the person was frail and at risk of falls) then it would be important to address these issues to try to remove this additional risk before reassessing overall risk of bleeding.

For an explanation of the rationale behind the recommendations for the choice of anticoagulant to use for extended treatment, including the use of aspirin, see the benefits and harms section of the evidence review for pharmacological treatment of VTE (chapter D).

The committee agreed that it was also important to revisit the discussion about bleeding risk and any decision to continue treatment in the future, especially if there were any changes in health state and wrote a recommendation reflecting this point.

The committee noted that studies predicting VTE-recurrence were conducted in separate populations to those predicting major bleeding. The committee agreed that in practice it would be helpful to have a tool capable of predicting the risk of both events in a given individual. Additionally, the committee noted that the risk of major bleeding was typically assessed at the point of beginning initial therapy, however for the purpose of this review and the decision faced by clinicians and people with VTE, risk after 3 months treatment is more directly applicable. Based on these discussions, the committee made a research recommendation for a combined tool to be designed that could predict both an individual's risk of recurrence were they to stop treatment **and** risk of bleeding were they to continue. This tool would ideally provide a score that corresponds to a clinical decision (continue or stop anticoagulation) allowing the individual with VTE to be treated accordingly. They agreed that tool should be validated externally to ensure that it is able to predict risk accurately in a separate group of people to the derivation cohort. The committee also recommended that the tool should be tested after 3 months and in the longer term to ensure that it had utility for

decision making at both of these times (see research recommendation 1 in appendix K for more details.

Cost effectiveness and resource use

No economic evidence was identified to inform the cost effectiveness of using different tools to predict the risk of VTE-recurrence or bleeding. The committee noted that in current practice, clinicians are likely to consider factors such as age and sex when assessing a person's risk of VTE-recurrence. The use of the DASH score would require a D-dimer test approximately 1 month after stopping anticoagulation, which may be associated with an additional cost if not already part of routine follow-up. The HAS-BLED prognostic score is based on information that is likely to already be available as part of a person's clinical history or routine care and therefore the use of this tool to assess bleeding risk is unlikely to be associated with additional costs.

Other than the time and information requirements associated with the administration of risk prediction tools, the main economic implications of using them depend on the accuracy of the tools for informing optimal treatment decisions. For VTE-recurrence, correctly identifying TNs would potentially lead to cost savings if people with a low risk of recurrence are able to safely stop taking anticoagulation medication whereas people who are incorrectly identified as FNs and stop anticoagulation medication will incur costs in the event of a recurrence. People who are incorrectly identified as FPs and continue to be treated with anticoagulation medication not only incur the costs of long-term treatment but are also at an increased risk of bleeding. The use of the HAS-BLED score to identify people who are at a high risk of major bleeding could lead to additional costs associated with more frequent monitoring or treatments to address modifiable risk factors (such as prescribing proton pump inhibitors to reduce the risk of gastrointestinal bleeds) but these may be offset by cost savings from reductions in major bleeds.

Other factors the committee took into account

The committee noted that existing tools for major bleeding and VTE recurrence had poor classification accuracy in older populations. For example, various major bleeding tools were validated in populations of people 65 years or older and typically had poor classification accuracy. This would limit the use of existing tools in this age group.

The committee agreed that it was difficult to assess the usefulness of existing prediction tools in clinical practice because none of the trials identified during this review also assessed clinician judgement. They noted that in future updates of this review it would be useful to include clinician judgment as a comparator and specifically look for information concerning this because they reasoned that although existing tools (and any new tools) may not have very high prognostic accuracy, they are still likely to be useful in the decision-making process if they have greater prognostic accuracy then clinician judgement alone. Taking this into account, the committee made a second research recommendation to test the newly developed combined tool for assessing VTE recurrence and major bleeding (see research recommendation 1) against clinician judgement. It was intended that this recommendation would only be conducted after the new tool had been externally validated. The committee envisioned that the new tool could be tested against clinician judgment using a prospective cohort study or test and treat RCT (see research recommendation 2 in appendix K for more details).

Appendices

Appendix A – Review protocols

Review protocol: What factors determine the optimum duration of pharmacological treatment for DVT or PE in people with an VTE?

Field (based on PRISMA-P)	Content
Review question	What factors determine the optimum duration of pharmacological treatment for DVT or PE in people with VTE?
Type of review question	Prognostic
Objective of the review	It was identified during scoping that there is varying practice as to how long a person who has had a VTE should receive anticoagulant medication. The aim of this review is to determine the factors that should be weighed up when considering whether to extend anticoagulation treatment.
Eligibility criteria – population	Adults (18+ years) who have had VTE For studies predicting VTE recurrence, participants must have previously received a course of anticoagulant treatment, but not be receiving anticoagulant treatment for the duration of follow up. For studies predicting major bleeding, participants must be receiving anticoagulant treatment for the duration of follow up. Analysis will be stratified according to first/recurrent VTE if data is available.
Eligibility criteria – prognostic factor(s)	**Prognostic factors:** • Any risk tool using multiple predictors to predict the risk of VTE recurrence following

	cessation of anticoagulant treatment. For example: DASH, HERD002, Vienna. • Any risk tool using multiple predictors to predict the risk of major bleeding during anticoagulant treatment. For example: HAS-BLED, VTE-BLEED
Outcomes of interest	• Recurrence of VTE • Major bleeding
Outcome measures	For each outcome, prognostic accuracy measures will be reported where available, for example: • Odds ratios/hazard ratios • Model fit (e.g. r squared) • Sensitivity, specificity, positive and negative predictive values.
Eligibility criteria – study design	Prospective or retrospective observational cohorts will be included for evaluation of risk prediction tools. If a review of individual risk factors is undertaken, only prospective studies will be included. Studies must have a minimum sample size of 250.
Other inclusion exclusion criteria	• English language only. • Case-control studies will be excluded. • If a review of individual risk factors is carried out, studies reporting univariate analysis only will be excluded. • Retrospective studies will be excluded if a review of individual risk factors is undertaken. • Studies with a sample size of fewer than 250 will be excluded.
Proposed sensitivity/sub-group analysis, or meta-regression	None

Selection process – duplicate screening/selection/analysis	10% of the abstracts were reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. If meaningful disagreements were found between the different reviewers, a further 10% of the abstracts were reviewed by two reviewers, with this process continued until agreement is achieved between the two reviewers. From this point, the remaining abstracts will be screened by a single reviewer. This review made use of the priority screening functionality with the EPPI-reviewer systematic reviewing software. See Appendix B for more details.
Data management (software)	See Appendix B

Information sources – databases and dates	• Sources to be searched
	  ○ Clinical searches - Medline, Medline in Process, PubMed, Embase, Cochrane CDSR, CENTRAL, DARE (legacy records) and HTA. MHRA Drug Alerts
	  ○ Economic searches - Medline, Medline in Process, PubMed, Embase, NHS EED (legacy records) and HTA, with economic evaluations and quality of life filters applied.
	• Supplementary search techniques
	  ○ None identified
	• Limits
	  ○ Studies reported in English
	  ○ Study design RCT, SR and Observational filter will be applied (as agreed)
	  ○ Animal studies will be excluded from the search results
	  ○ Conference abstract / letters will be excluded from the search results

Identify if an update	This is a new question for the update of CG144, therefore no date limit to be applied to searches.

Author contacts	https://www.nice.org.uk/guidance/indevelopment/gid-ng10087
Highlight if amendment to previous protocol	For details please see section 4.5 of Developing NICE guidelines: the manual
Search strategy – for one database	For details please see appendix C of the evidence review
Data collection process – forms/duplicate	A standardised evidence table format will be used, and published as appendix E (clinical evidence tables) or I (economic evidence tables, where relevant) of the evidence review.
Data items – define all variables to be collected	For details please see evidence tables in appendix E (clinical evidence tables) or I (economic evidence tables, where relevant) of the evidence review.
Methods for assessing bias at outcome/study level	See appendix B
Criteria for quantitative synthesis (where suitable)	See appendix B
Methods for analysis – combining studies and exploring (in)consistency	See appendix B
Meta-bias assessment – publication bias, selective reporting bias	See appendix B
Assessment of confidence in cumulative evidence	See appendix B
Rationale/context – Current management	For details please see the introduction to the evidence review.
Describe contributions of authors and guarantor	A multidisciplinary committee developed the guideline. The committee was convened by the NICE Guidelines Updates Team and chaired by Susan Bewley in line with section 3 of Developing NICE guidelines: the manual. Staff from the NICE Guidelines Updates Team undertook systematic literature searches, appraised the evidence, conducted meta-analysis

	and cost-effectiveness analysis where appropriate, and drafted the guideline in collaboration with the committee. For details please see the methods section of the evidence review.
Sources of funding/support	The NICE Guideline Updates Team is an internal team within NICE.
Name of sponsor	The NICE Guideline Updates Team is an internal team within NICE.
Roles of sponsor	The NICE Guideline Updates Team is an internal team within NICE.

Appendix B – Methods

Priority screening

The reviews undertaken for this guideline all made use of the priority screening functionality with the EPPI-reviewer systematic reviewing software. This uses a machine learning algorithm (specifically, an SGD classifier) to take information on features (1, 2 and 3 word blocks) in the titles and abstract of papers marked as being 'includes' or 'excludes' during the title and abstract screening process, and re-orders the remaining records from most likely to least likely to be an include, based on that algorithm. This re-ordering of the remaining records occurs every time 25 additional records have been screened.

Research is currently ongoing as to what are the appropriate thresholds where reviewing of abstract can be stopped, assuming a defined threshold for the proportion of relevant papers it is acceptable to miss on primary screening. As a conservative approach until that research has been completed, the following rules were adopted during the production of this guideline:

- In every review, at least 50% of the identified abstract (or 1,000 records, if that is a greater number) were always screened.
- After this point, screening was only terminated if a pre-specified threshold was met for a number of abstracts being screened without a single new include being identified. This threshold was set according to the expected proportion of includes in the review (with reviews with a lower proportion of includes needing a higher number of papers without an identified study to justify termination), and was always a minimum of 250.
- A random 10% sample of the studies remaining in the database were additionally screened, to check if a substantial number of relevant studies were not being correctly classified by the algorithm, with the full database being screened if concerns were identified.

As an additional check to ensure this approach did not miss relevant studies, the included studies lists of included systematic reviews were searched to identify any papers not identified through the primary search.

Incorporating published systematic reviews

For all review questions where a literature search was undertaken looking for a particular study design, systematic reviews containing studies of that design were also included. All included studies from those systematic reviews were screened to identify any additional relevant primary studies not found as part of the initial search.

Quality assessment

Individual systematic reviews were quality assessed using the ROBIS tool, with each classified into one of the following three groups:

- High quality – It is unlikely that additional relevant and important data would be identified from primary studies compared to that reported in the review, and unlikely that any relevant and important studies have been missed by the review.
- Moderate quality – It is possible that additional relevant and important data would be identified from primary studies compared to that reported in the review, but unlikely that any relevant and important studies have been missed by the review.

- Low quality – It is possible that relevant and important studies have been missed by the review.

Each individual systematic review was also classified into one of three groups for its applicability as a source of data, based on how closely the review matches the specified review protocol in the guideline. Studies were rated as follows:

- Fully applicable – The identified review fully covers the review protocol in the guideline.
- Partially applicable – The identified review fully covers a discrete subsection of the review protocol in the guideline.
- Not applicable – The identified review, despite including studies relevant to the review question, does not fully cover any discrete subsection of the review protocol in the guideline.

Using systematic reviews as a source of data

If systematic reviews were identified as being sufficiently applicable and high quality, and were identified sufficiently early in the review process, they were used as the primary source of data, rather than extracting information from primary studies. The extent to which this was done depended on the quality and applicability of the review, as defined in Table 6. When systematic reviews were used as a source of primary data, any unpublished or additional data included in the review which is not in the primary studies was also included. Data from these systematic reviews was then quality assessed and presented in GRADE tables as described below, in the same way as if data had been extracted from primary studies. In questions where data was extracted from both systematic reviews and primary studies, these were cross-referenced to ensure none of the data had been double counted through this process.

Table 6 Criteria for using systematic reviews as a source of data

Quality	Applicability	Use of systematic review
High	Fully applicable	Data from the published systematic review were used instead of undertaking a new literature search or data analysis. Searches were only done to cover the period of time since the search date of the review.
High	Partially applicable	Data from the published systematic review were used instead of undertaking a new literature search and data analysis for the relevant subsection of the protocol. For this section, searches were only done to cover the period of time since the search date of the review. For other sections not covered by the systematic review, searches were undertaken as normal.
Moderate	Fully applicable	Details of included studies were used instead of undertaking a new literature search. Full-text papers of included studies were still retrieved for the purposes of data analysis. Searches were only done to cover the period of time since the search date of the review.
Moderate	Partially applicable	Details of included studies were used instead of undertaking a new literature search for the relevant subsection of the protocol. For this section, searches were only done to cover the period of time since the search date of the review. For other sections not covered by the systematic review, searches were undertaken as normal.

Prognostic test accuracy evidence

In this guideline, prognostic test accuracy data are classified as any data in which a feature – be it a symptom, a risk factor, a test result or the output of some algorithm that combines many such features – is observed in some people who go on to develop the condition of interest and some people who do not. Such data either explicitly provide, or can be manipulated to generate, a 2x2 classification of true positives and false negatives (in people who, according to the reference standard, truly develop the condition) and false positives and true negatives (in people who, according to the reference standard, do not). This category would include studies classed as prediction models under the TRIPOD statement, provided the data were reported a 2x2 classification data.

The 'raw' 2x2 data can be summarised in a variety of ways. Those that were used for decision making in this guideline are as follows:

- **Positive likelihood ratios** describe how many times more likely positive features are in people who develop the condition compared to people who do not. Values greater than 1 indicate that a positive result makes the condition more likely.
  - LR^+ = (TP/[TP+FN])/(FP/[FP+TN])
- **Negative likelihood ratios** describe how many times less likely negative features are in people who develop the condition compared to people who do not. Values less than 1 indicate that a negative result makes the condition less likely.
  - LR^- = (FN/[TP+FN])/(TN/[FP+TN])
- **Sensitivity** is the probability that the feature will be positive in a person who goes on to develop the condition.
  - sensitivity = TP/(TP+FN)
- **Specificity** is the probability that the feature will be negative in a person who does not go on to develop the condition.
  - specificity = TN/(FP+TN)

Interpretation of prognostic accuracy measures

Clinical decision thresholds were chosen by the committee to correspond to the likelihood ratio above (for positive likelihood ratios) or below (for negative likelihood ratios) which a diagnostic test was accurate enough to be recommended. The following schema, adapted from the suggestions of Jaeschke et al. (1994), was used inform these discussions.

Table 7: Interpretation of likelihood ratios

Value of likelihood ratio	Interpretation
LR ≤ 0.1	**Very large** decrease in probability of disease
0.1 < LR ≤ 0.2	**Large** decrease in probability of disease
0.2 < LR ≤ 0.5	**Moderate** decrease in probability of disease
0.5 < LR ≤ 1.0	**Slight** decrease in probability of disease
1.0 < LR < 2.0	**Slight** increase in probability of disease
2.0 ≤ LR < 5.0	**Moderate** increase in probability of disease
5.0 ≤ LR < 10.0	**Large** increase in probability of disease
LR ≥ 10.0	**Very large** increase in probability of disease

The schema above has the effect of setting a minimal important difference for positive likelihoods ratio at 2, and a corresponding minimal important difference for negative

likelihood ratios at 0.5. Likelihood ratios (whether positive or negative) falling between these thresholds were judged to indicate no meaningful change in the probability of disease.

Quality assessment

Individual studies were quality assessed using the PROBAST tool, which contains five domains: participant selection, predictors, outcome, sample size and participant flow, analysis. Each individual study was classified into one of the following two groups:

- Low risk of bias – Evidence of non-serious bias in zero or one domain.
- Moderate risk of bias – Evidence of non-serious bias in two domains only, or serious bias in one domain only.
- High risk of bias – Evidence of bias in at least three domains, or of serious bias in at least two domains.

Each individual study was also classified into one of three groups for directness, based on if there were concerns about the population, predictive features and/or reference standard in the study and how directly these variables could address the specified review question. Studies were rated as follows:

- Direct – No important deviations from the protocol in population, predictive feature and/or reference standard.
- Partially indirect – Important deviations from the protocol in one of the population, predictive feature and/or reference standard.
- Indirect – Important deviations from the protocol in at least two of the population, predictive feature and/or reference standard.

Methods for combining prognostic test accuracy evidence

Where applicable, prognostic test accuracy syntheses were stratified by:

- Presenting symptomatology (features shared by all participants in the study, but not all people in the full relevant clinical population).
- The length of time between the measurement of the predictive feature and the final outcome.
- The reference standard used for categorising true positives.

Where five or more studies were available for all included strata, a bivariate model was fitted using the `mada` package in R v3.4.0, which accounts for the correlations between positive and negative likelihood ratios, and between sensitivities and specificities. Where sufficient data were not available (2-4 studies), separate independent pooling was performed for positive likelihood ratios, negative likelihood ratios, sensitivity and specificity, using Microsoft Excel. This approach is likely to somewhat underestimate test accuracy (see Deeks 2001).

Random-effects models (der Simonian and Laird) were fitted for all syntheses, due to the expected level of between study heterogeneity in prognostic reviews.

In any meta-analyses where some (but not all) of the data came from studies at high risk of bias, a sensitivity analysis was conducted, excluding those studies from the analysis. Results from both the full and restricted meta-analyses are reported. Similarly, in any meta-analyses where some (but not all) of the data came from indirect studies, a sensitivity analysis was conducted, excluding those studies from the analysis.

Modified GRADE for prognostic test accuracy evidence

GRADE has not been developed for use with prognostic test accuracy studies; therefore a modified approach was applied using the GRADE framework. GRADE assessments were only undertaken for positive and negative likelihood ratios, as the MIDs used to assess imprecision were based on these outcomes.

Cross-sectional and cohort studies were initially rated as high-quality evidence if well conducted, and then downgraded according to the standard GRADE criteria (risk of bias, inconsistency, imprecision and indirectness) as detailed in Table 8 below.

Table 8: Rationale for downgrading quality of evidence for prognostic questions

GRADE criteria	Reasons for downgrading quality
Risk of bias	Not serious: If less than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from studies at high risk of bias, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies at high and low risk of bias.
Indirectness	Not serious: If less than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from indirect studies, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between direct and indirect studies.
Inconsistency	Concerns about inconsistency of effects across studies, occurring when there is unexplained variability in the treatment effect demonstrated across studies (heterogeneity), after appropriate pre-specified subgroup analyses have been conducted. This was assessed using the I^2 statistic. N/A: Inconsistency was marked as not applicable if data on the outcome was only available from one study. Not serious: If the I^2 was less than 33.3%, the outcome was not downgraded. Serious: If the I^2 was between 33.3% and 66.7%, the outcome was downgraded one level. Very serious: If the I^2 was greater than 66.7%, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies with the smallest and largest effect sizes.
Imprecision	If the 95% confidence interval for positive or negative likelihood ratios crossed the decision threshold for recommending a test the outcome was downgraded 1 level. If the 95% confidence interval crossed 1 (the likelihood ratio corresponding to no diagnostic utility), the outcome was downgraded 2 levels as suffering from very serious imprecision. For information on how decision

GRADE criteria	Reasons for downgrading quality
	thresholds were determined, see the section on interpretation of prognostic accuracy measures. Outcomes meeting the criteria for downgrading above were not downgraded if the confidence interval was sufficiently narrow that the upper and lower bounds would correspond to clinically equivalent scenarios.

The quality of evidence for each outcome was upgraded if either of the following conditions were met:

- Data showing an effect size sufficiently large that it cannot be explained by confounding alone.
- Data where all plausible residual confounding is likely to increase our confidence in the effect estimate.

Publication bias

Where 10 or more studies were included as part of a single meta-analysis, a funnel plot was produced to graphically assess the potential for publication bias.

Evidence statements for likelihood ratios

Evidence statements were written for positive and negative likelihood ratios and indicate the magnitude of effect on the probability of experiencing an recurrent event or a major bleed (based on the categories in Table 7) associated with a positive test result or a negative test result with a quality rating for each finding.

Other prognostic evidence

Other prognostic studies were also included if they reported outcomes of c-statistics, hazard ratios or model fit statistics. These studies were also quality assessed using the PROBAST checklist, as in the prognostic test accuracy section above.

Methods for combining hazard ratios or odds ratios

Where appropriate, hazard ratios were pooled using the inverse-variance method. Adjusted hazard ratios from multivariate models were only pooled if the same set of predictor variables were used across multiple studies and they were on the same scale. For hazard ratios, a range of (0.8, 1.25) was used to assess imprecision in the absence of a more clinically meaningful MID. Where meta-analysis was possible risk of bias and indirectness were assessed as detailed in Table 8 for other prognostic evidence.

In the absence of hazard ratio data that could be meta-analysed, data was pooled to obtain single GRADE ratings per index using the following decision rules:

1. Risk of bias and indirectness were assessed at the individual study level because there were only single studies for each outcome and test. If multiple studies had been included in each outcome then these would have been calculated using the % weight based on sample size.
2. Imprecision:
    a. In cases where a single or multiple per point increase hazard ratios are presented, the level of imprecision is calculated for each study using the MID of 0.8, 1.25.

   b. In cases where several hazard are presented compared to a reference category then the most extreme category is assessed using the MID and a single pooled estimate is determined as in 2a. If the reference categories are in opposite directions then the high reference category data is reversed (1/value) and then included in the analysis as before.
   c. In cases where there is a mix of data then the imprecision is calculated for each study and then merged based on population weight as in 2a.
3. Inconsistency:
   a. For a single study this is judged to be not applicable (N/A).
   b. For multiple studies with single HRs this is judged using I^2 calculated using Review Manager v5.3 and assessed following the rules in Table 8.
   c. For multiple studies with the same reference category, inconsistency was assessing the I^2 value for the pooled studies and then following the rules in Table 8.
   d. If hazard ratio data for a single index was reported in several ways (per point increase, with reference to high and/or low categories) then inconsistency for this outcome was determined to be serious as the results were not comparable.

In the absence of odds ratio data that could be meta-analysed, data was pooled to obtain single GRADE ratings per index using the following decision rules:

1. Risk of bias and indirectness were assessed at the individual study level because there were only single studies for each outcome and test.
2. Inconsistency was not applicable as there was only one study per line of the GRADE table.
3. Imprecision was assessed by determining whether the 95% CI crossed the line of no effect. If this was the case the outcome was downgraded once. If the study had less than 500 participants then the outcome was downgraded another level.

Evidence statements for hazard ratios and odds ratios

For hazard ratio data, evidence statements were divided into 4 groups as follows:

- Situations where the data are only consistent, at a 95% confidence level, with an effect in one direction (i.e. one that is 'statistically significant'), and the magnitude of that effect is most likely to meet or exceed the MID (i.e. the point estimate is not in the zone of equivalence). In such cases, we state that the evidence showed that there is an effect.
- Situations where the data are only consistent, at a 95% confidence level, with an effect in one direction (i.e. one that is 'statistically significant'), but the magnitude of that effect is most likely to be less than the MID (i.e. the point estimate is in the zone of equivalence). In such cases, we state that the evidence showed there is an effect, but it is less than the defined MID.
- Situations where the confidence limits are smaller than the MIDs in both directions. In such cases, we state that the evidence demonstrates that there is no meaningful difference.
- In all other cases, we state that the evidence could not differentiate between the comparators.

For odds ratio data, evidence statements are divided into 2 groups as follows:

- We state that the evidence showed that there is an effect if the 95% CI does not cross the line of no effect.
- The evidence could not differentiate between comparators if the 95% CI crosses the line of

no effect.

Methods for combining c-statistics

C-statistics were assessed in a similar manner to likelihood ratios using the categories in Table 9 below.

Table 9 Interpretation of c-statistics

Value of c-statistic	Interpretation
c-statistic <0.6	**Poor classification accuracy**
0.6 ≤ c-statistic <0.7	**Adequate classification accuracy**
0.7 ≤ c-statistic <0.8	**Good classification accuracy**
0.8 ≤ c-statistic <0.9	**Excellent classification accuracy**
0.9 ≤ c-statistic < 1.0	**Outstanding classification accuracy**

Where there were 2 or more studies reporting c-statistics with 95% CIs for the same tests, meta-analyses were carried out using the metamisc package in R v3.4.0, which confines the analysis results to between 0 and 1 matching the limited range of values that c-statistics can take. i-squared value were calculated from a univariate meta-analysis using the metafor package in R v3.4.0. Random effects meta-analysis was used when the I^2 was 50% or greater.

A modified version of GRADE was carried out to assess the quality of the meta-analysed c-statistics as follows:

- Imprecision - the 95% CI boundaries were examined and if they crossed 2 categories of test classification accuracy then the study was downgraded once (imprecision rated as serious); if the boundaries crossed 3 categories then the study was downgraded twice (very serious imprecision).
- Risk of bias, inconsistency and indirectness were assessed as detailed in Table 8 for other prognostic evidence.

Evidence statements for c-statistics

Evidence statements were written for c-statistics to indicate the prognostic classification accuracy of each test (in according the categories in Table 9), the quality of the evidence and the number of studies and participants.

Appendix C – Literature search strategies

Searches were carried out in Medline, Medline in Process, Medline epub ahead of print, Embase, (all via the Ovid platform) DARE, CDSR and CENTRAL (via the Wiley platform) on 22nd and 23rd October 2018 and re run on 4th April 2019.

The Medline strategy is presented below,

1 Venous Thrombosis/
2 (phlegmasia adj2 dolens).tw.
3 (thrombo* adj2 (vein* or venous)).tw.
4 (venous adj stasis).tw.
5 (dvt or vte).tw.
6 Venous Thromboembolism/ or Embolism, paradoxical/
7 exp pulmonary embolism/
8 ((pulmonary or lung) adj4 (embol* or thromboembo* or microembol*)).tw.
9 (pulmonary adj infarction).tw.
10 or/1-9
11 ((risk* or predict* or decision*) adj2 (tool* or score* or model* or calculat* or rule* or algorithm*)).tw.
12 predictive value of tests/
13 algorithms/
14 decision support techniques/
15 risk assessment/mt
16 statistical models/
17 Decision Making, Computer - Assisted/
18 Proportional Hazards models/
19 or/12-18
20 Recurrence/
21 (recur* or reoccur* or repeat* or reappear* or relapse* or return* or repetit*).tw.
22 anticoagulants/
23 (anticoagulant* or anti coagulant*).tw.
24 hemorrhage/
25 (hemorrhag* or haemorrhag* or bleed*).tw.
26 or/20-25
27 19 and 26
28 10 and (11 or 27)
29 (HAS-BLED or VTE-BLEED or DASH or HERDOO2 or Vienna).tw.
30 10 and (28 or 29)
31 animals/ not humans/
32 30 not 31
33 limit 32 to english language/

Searches to identify economic evidence were run on 30th October 2018 in Medline, Medline in Process, Embase, Econlit (Ovid platform) and NHS EED and the Health Technology Database (Wiley platform). Economic evaluation and Quality of Life filters were attached to the above strategy for Medline and Embase databases. A single search to identify economic evidence across all questions was re run on 9th April 2019. The Medline version of the filters is displayed below.

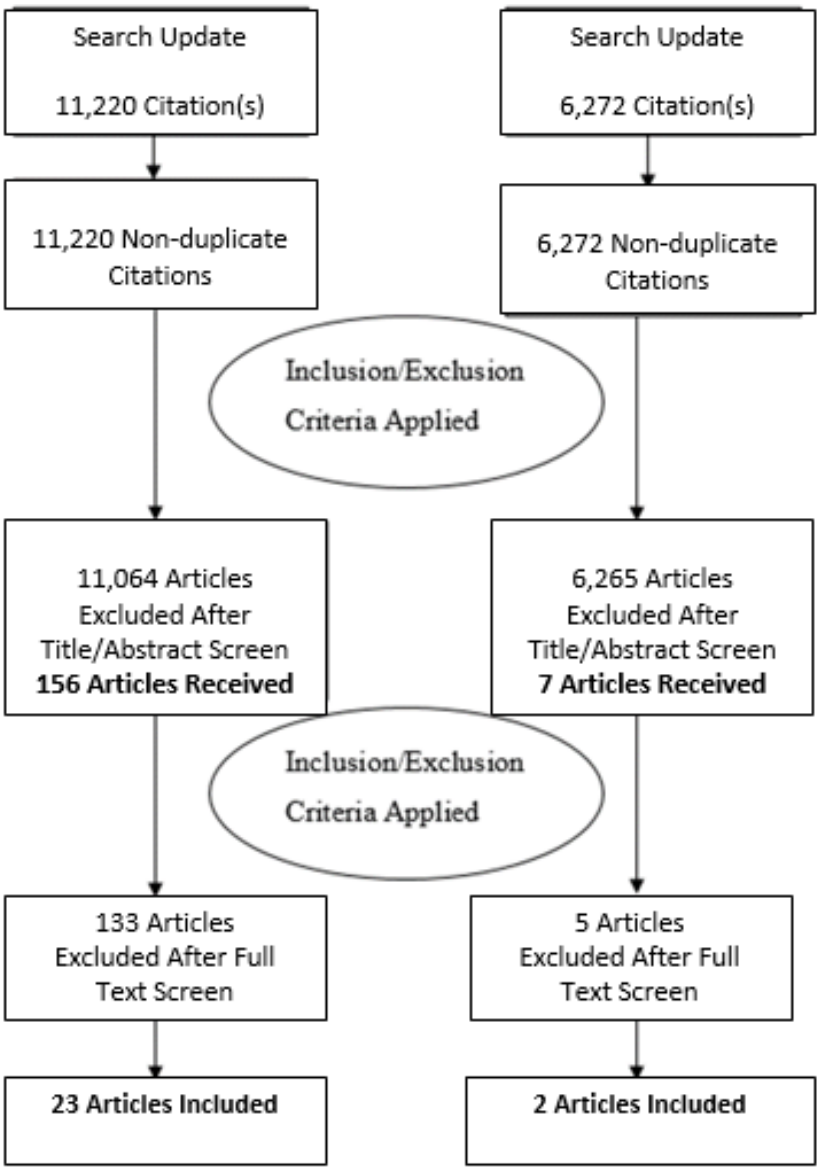
Economic evaluations

1 Economics/
2 exp "Costs and Cost Analysis"/
3 Economics, Dental/
4 exp Economics, Hospital/
5 exp Economics, Medical/
6 Economics, Nursing/
7 Economics, Pharmaceutical/
8 Budgets/
9 exp Models, Economic/
10 Markov Chains/
11 Monte Carlo Method/
12 Decision Trees/
13 econom$.tw.
14 cba.tw.
15 cea.tw.
16 cua.tw.
17 markov$.tw.
18 (monte adj carlo).tw.
19 (decision adj3 (tree$ or analys$)).tw.
20 (cost or costs or costing$ or costly or costed).tw.
21 (price$ or pricing$).tw.
22 budget$.tw.
23 expenditure$.tw.
24 (value adj3 (money or monetary)).tw.
25 (pharmacoeconomic$ or (pharmaco adj economic$)).tw.
26 or/1-25

Quality of Life

1 "Quality of Life"/
2 quality of life.tw.
3 "Value of Life"/
4 Quality-Adjusted Life Years/
5 quality adjusted life.tw.
6 (qaly$ or qald$ or qale$ or qtime$).tw.
7 disability adjusted life.tw.
8 daly$.tw.
9 Health Status Indicators/ (22343)
10 (sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).tw.
11 (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw.
12 (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw.
13 (sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortform sixteen or short form sixteen).tw.
14 (sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).tw.
15 (euroqol or euro qol or eq5d or eq 5d).tw.
16 (qol or hql or hqol or hrqol).tw.
17 (hye or hyes).tw.
18 health$ year$ equivalent$.tw.
19 utilit$.tw.
20 (hui or hui1 or hui2 or hui3).tw.
21 disutili$.tw.
22 rosser.tw.
23 quality of wellbeing.tw.
24 quality of well-being.tw.
25 qwb.tw.
26 willingness to pay.tw.
27 standard gamble$.tw.
28 time trade off.tw.
29 time tradeoff.tw.
30 tto.tw.
31 or/1-30

Appendix D – Clinical evidence study selection



Appendix E – Clinical evidence tables

Author (year)	Title	Study details	Quality assessment
Alatri (2017)	The Modified Ottawa Score and Clinical Events in Hospitalized Patients with Cancer-Associated Thrombosis from the Swiss VTE Registry	Study type • Retrospective analysis of a prospective cohort Study details • Study location Switzerland • Study setting SWIVTER registry - containing data on VTE patients in 11 acute care hospitals in Switzerland. • Study dates November 2012 - February 2015 • Duration of follow-up 90 days • Sources of funding study was funded by the International Society on Thrombosis and Haemostasis (ISTH) 2007 Presidential Fund, Geneva, Sanofi-Aventis (Suisse) SA, Vernier; Bayer (Schweiz) AG, Zurich; Pfizer AG, Zurich; and Bristol-Myers Squibb AG, Cham, Switzerland. Inclusion criteria • VTE • active cancer Sample characteristics • Sample size 383 participants • % female 46.7% female • Mean age (SD) 68 (SD14) years • VTE characteristics 65% received initial therapy with LMWH 33% received inital	Participant selection • Low risk of bias Predictors • Low risk of bias Identified retrospectively using database records, however predictors used in model were unlikely to have been recorded erroneously/not reported. Outcome • Low risk of bias Sample size and participant flow • High risk of bias Low number of participants with the event poses a risk of biased estimations of the model's predictive utility. Analysis • High risk of bias Adjustments performed for VTE-recurrence, but not major bleeding (with only the latter being extracted in this review due to the former not meeting the criteria for inclusion). Only the % experiencing the outcome was reported with no mention of model fit. Overall risk of bias • Moderate VTE-recurrence outcomes were adjusted for potentially confounding variables however major-bleeding outcomes were likely not adjusted. Additionally there were a low number of participants experiencing the event.

Author (year)	Title	Study details	Quality assessment
		therapy with UFH 2% received initial therapy with DOAC • % recurrent VTE 18.8% had a prior VTE • % hypertension 40.7% • % renal disease 13.6% renal failure • % active cancer 100%; 44.9% metastatic disease; 60.8% curative treatment plan; 12.5% had lung cancer; 11.2% had breast cancer; 55.1% active cancer • % chronic lung disease 15.7% • % chronic heart failure 18.5%; 7.8% had history of stroke/TIA • % life expectancy <6 months 26.9% • % chemotherapy/radiotherapy 25.6% chemotherapy 9.1% radiohtherapy Prognostic factor(s) • MOS score Female sex: 1 point Lung cancer: 1 point Prior VTE: 1 point Localized cancer without metastasis (stages 1 and 2 for solid tumors): -1 point Breast cancer: -1 point Low: -1 or fewer points Intermediate: 0 points High 1+ points Outcome of interest • Major bleeding "Major bleeding was defined in accordance with the	Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		International Society on Thrombosis and Haemostasis (ISTH) criteria as fatal bleeding, and/or symptomatic bleeding in a critical area or organ, such as intracranial, intraspinal, intraocular, retroperitoneal, intra-articular or pericardial, or intramuscular with compartment syndrome, and/or bleeding causing a fall in hemoglobin level of 2 g/dL or more, or leading to transfusion of two or more units of whole blood or red cells" Outcome measures • Sensitivity/specificity/Likelihood ratios 2x2 table possible	
Brown (2018)	Risk Stratification for Bleeding Complications in Patients With Venous Thromboembolism: Application of the HAS-BLED Bleeding Score During the First 6 Months of Anticoagulant Treatment	Study type • Retrospective cohort study Validation study using retrospective claims database Study details • Study location USA • Study setting Truven Health MarketScan Commercial Claims and Medicare Supplemental Databases containing data on around 40million people from over 160 large employers in the USA • Study dates database searched between 2010 to 2014; patients had to be diagnosed between January 1, 2010 and November 31, 2013 *Likely a typo and actually December • Duration of follow-up up to 180 days Inclusion criteria • VTE The date of the first qualifying diagnosis of VTE was defined as the index date, requiring that the diagnosis was in the primary position on an inpatient hospital record. VTE was	Participant selection • High risk of bias Participants were identified using a claims database and therefore there is variability in how the index VTE or the major bleeding event was diagnosed and recorded. Predictors • High risk of bias Study used a claims database and identified the presence of HAS-BLED predictors using ICD-9 codes, which therefore has the potential for poor reporting / recording. Outcome • Low risk of bias Sample size and participant flow • Low risk of bias Analysis • Low risk of bias

Author (year)	Title	Study details	Quality assessment
		identified by International Classification of Disease, 9th revision (ICD-9) codes based on previously validated coding algorithms Exclusion criteria • <18 years • >12 months pre-index and 1 month post-index enrolment with a medical + outpatient pharmacy info included in the database • no-treatment in the period up VTE diagnosis Sample characteristics • Sample size 132,280 • %female 56% female • Mean age (SD) 63.7% aged 18-64 12.7% aged 65-74 23.6% aged 75 years or older • % hypertension 44.5% • % renal disease 7.7% • % liver disease 5.7% • % bleeding history 11.7% • % NSAIDS/ antiplatelet use Prognostic factor(s) • HAS-BLED >160 mmHg systolic blood pressure: 1 point Abnormal liver function (history of cirrhosis, or bilirubin > 2x the upper limit of normal in association with aspartate aminotransferase/alanine aminotransferase/ alkaline phosphatase levels > 3x the upper limit of normal): 1 point	Overall risk of bias • Moderate This was a retrospective review using a claims database with potential for poor reporting/recording of outcome event and predictor variables. Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		Abnormal renal function (on dialysis, a history of kidney transplantation, or serum creatinine values > 200 µmol/L): 1 point Stroke: 1 point Bleeding history or predisposition: 1 point Liable INR (time within therapeutic range < 60%): 1 point Age 65 years+: 1 point Drug use (platelet inhibitors or NSAIDS): 1 point Alcohol abuse (>8 units/week): 1 point Low risk: 0 points Moderate risk: 1-2 points High risk 3+ points *Patients with malignant neoplasms and/or metastatic disease during the baseline period were also identified using ICD-9 codes 140.x to 209.x. Anticoagulant treatment was identified as the first observed LMWH, non-vitamin K antagonist oral anticoagulant, or warfarin, allowing for bridge therapy from LMWH to warfarin. All categories for the HAS-BLED score were identified by ICD-9 codes except NSAIDs or antiplatelet medication use, which were identified from pharmacy records.* Outcome of interest • major bleeding Bleeding events were classified as “all” if they met the coding algorithm used or “major” if they occurred during an inpatient stay, were associated with a critical site, resulted in need for transfusion, or lead to death while in the hospital Bleeding events were tracked up until 180 days following diagnosis. The database followed patients until they had a major bleed, died or dropped out of the database.	

Author (year)	Title	Study details	Quality assessment
		Outcome measures • C-statistics C-indices (95% CI) for Cox PH and competing risks models, • HR Incremental Cause-Specific and Sub-Distribution Hazard Ratios Comparing HAS-BLED Scores Risk of Bleeding Events	
Carmona-Bayonas (2017)	Predicting serious complications in patients with cancer and pulmonary embolism using decision tree modelling: the EPIPHANY Index	Study type • Retrospective analysis of a prospective cohort Derivation study Study details • Study location Spain • Study setting 14 Spanish hospital taking part in the EPIPHANY registry • Study dates 2004 - 2015 Inclusion criteria • Acute PE cancer-associated PE, with PE diagnosed by objective imaging (CT angiography scans, high probability scintigraphy, or CT scheduled to assess tumour response or for other reasons). In case of multiple events, only one was considered to be the index PE, defined as the evaluable PE closest to the time of recruitment. The remaining PEs in the same patient were considered 'previous history' if they took place prior to the index PE, or 'recurrence', if subsequent to it. • active cancer withdrawn from the study if the PE had occurred more than 1 month prior to the diagnosis of cancer, or if more than 1 month had elapsed since completing adjuvant chemotherapy. Patients were also excluded if they had not received anticoagulant therapy without justification according to	Participant selection • Low risk of bias Predictors • High risk of bias Derivation study and therefore predictors were selected based upon relationship with outcome. Outcome • Low risk of bias Sample size and participant flow • High risk of bias <10 events per predictor originally entered into the model. Analysis • Low risk of bias Univariate analyses were (likely) performed however the final model took into consideration the weighting of each of the final variables with the outcome and potential interactions between the variables. Additionally the account for overfitting: "To cope with the overfitting and instability inherent to the decision tree, a 10-fold cross-validation procedure was applied. Thus, the data were randomly divided into 10 equal subsets. Trees were systematically built in 9 of those partitions (training subsets) and then tested in the remaining group (testing subset)."

Author (year)	Title	Study details	Quality assessment
		international clinical practice guidelines Sample characteristics • Sample size 1075 • %female 54.2% • Mean age (SD) 64 (SD12) years • % active cancer 100% 11.7% breast cancer 25.3% lung cancer • % chronic heart failure 5.2% had chronic CVD • % ECOG-PS status 2+ 47.1% • % COPD 11.9% Prognostic factor(s) • EPIPHANY index Final model contained 6 variables: the Hestia-like CDR variable (any risk factor present vs none), ECOG-PS (o2 vs X2), oxygen saturation (o90 vs X90%), presence of PE-specific symptoms, previous tumour response evaluation (tumour progression, unknown, or not evaluated vs others), and prior surgical resection of primary tumour Outcome of interest • major bleeding Episodes in a critical location (intracranial, intraspinal, intraocular, retroperitoneal, or pericardial) associated with death; bleeding with hemoglobin levels of >2 g/dL, or bleeding requiring two units of packed red blood cells. Outcome measures • sensitivity/specificity/Likelihood ratios	Overall risk of bias • Moderate Derivation study retrospectively assessed using prospective data. A low event-per-predictor ratio suggests a risk of a biased estimation of the models predictive utility. Directness • Partially directly applicable Only included participants with cancer-associated PE and participants were excluded if PE occurred >1 months prior to cancer diagnosis or if >1 month had elapsed since completing adjuvant chemotherapy.

Author (year)	Title	Study details	Quality assessment
Eichinger (2014)	D-dimer levels over time and the risk of recurrent venous thromboembolism: an update of the Vienna prediction model	Study type • Prospective cohort study analysis of a prospective study in which the original VIENNA model is tested again and expanded using measurements taken at 3 months after discontinuation of anticoagulation treatment to give a risk score at this point in time as well as the initial risk score given immediately following AC cessation. Study details • Study location Austria • Study setting 4 thrombosis centres in Vienna • Study dates participants were enrolled between January 2000 and August 2008 • Duration of follow-up followed for a median of 68 months. Venous blood was collected into 1/10 volume of trisodium citrate 0.11 mmol/L and was immediately centrifuged for 20 minutes at 2000g. The plasma was stored at -80 degrees C. baseline, that is, 3 weeks after discontinuation of vitamin K antagonists at the time when the prothrombin time had normalized, determination of antithrombin, protein C, and protein S, diagnosis of a lupus anticoagulant, and screening for factor V Leiden and for prothrombin G20210A were performed according to standard protocols. D-Dimer was determined by ELISA (Asserachrom D-Dimer, Boehringer Mannheim, Germany). Patients were seen 3 weeks (baseline) and 3, 9, 15, and 24 months after discontinuation of vitamin K antagonists. At these times blood for D-dimer measurement was obtained. • Sources of funding This study was supported by the Oesterreichische Nationalbank (Jubil€aumsfonds) and the Medizinisch-	Participant selection • High risk of bias Unclear whether participants were selected from the same pool as Eichinger 2010 (the original derivation study), therefore sample may be the same or have considerable overlap with original derivation study. Predictors • Low risk of bias Outcome • Low risk of bias Although technically a derivation study, the base model (VIENNA) was derived in a previous study (EICHINGER, 2010) and was tested here at different time points following AC discontinuation. Sample size and participant flow • Low risk of bias Analysis • Low risk of bias Overall risk of bias • Low Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		Wissenschaftlicher Fonds des B€urgermeisters der Bundeshauptstadt Wien. The sponsors had no role in the design and conduct of the study; in the collection, management, analysis, or interpretation of the data; or in the preparation, review, or approval of the manuscript. Inclusion criteria • VTE Patients with VTE provoked by surgery, trauma, pregnancy, or female hormone intake, with a natural inhibitor deficiency, the lupus anticoagulant, or cancer were excluded. Diagnosis of DVT was established by a positive finding on venography or colour duplex sonography. The diagnosis of PE was confirmed by spiral computed tomography or ventilation–perfusion scanning. Patients with both symptomatic PE and DVT were classified as having a PE. Also excluded patients with VTE provoked by a transient risk factor, as their recurrence risk is low and extended anticoagulation is not justified. • received prior anticoagulation > 3 months Sample characteristics • Sample size 553 participants • %female 39.6% • Mean age (SD) 53.4 years old *median age, mean not given • VTE characteristics distal DVT:16.1% proximal DVT: 38.5% PE with or without DVT in 45.4%. • % recurrent VTE 26.8% had recurrent VTE • average duration of AC treatment	

Author (year)	Title	Study details	Quality assessment
		median 6.7 months Prognostic factor(s) • VIENNA tool modified calculator available at: http://www.meduniwien.ac.at/user/georg.heinze/dvpm/ Outcome of interest • recurrent VTE The end point of the study was recurrent symptomatic DVT confirmed by venography or colour duplex sonography (in case of proximal thrombosis of the contralateral leg) or recurrent symptomatic PE, confirmed by ventilation–perfusion scanning and/or spiral computed tomography.	
Franco (2016)	A risk score for prediction of recurrence in patients with unprovoked venous thromboembolism (DAMOVES)	Study type • Prospective cohort study Study details • Study location Spain • Study setting departments of Internal Medicine at two centres in Spain • Study dates March 2004 to August 2013 • Duration of follow-up median 21.3 months Inclusion criteria • VTE "first unprovoked VTE who had been treated with oral anticoagulants for at least 3months were included." "Diagnosis of DVT was established by a positive finding on venography or colour duplex sonography. The diagnosis of PE was confirmed either by ventilation–perfusion scanning or by spiral computed tomography. Patients with symptomatic	Participant selection • Low risk of bias Predictors • High risk of bias Derivation study therefore risk factors were entered into the model based on relationship with the outcome. Outcome • Low risk of bias Sample size and participant flow • Low risk of bias moderate number of events per predictor (9.29) Analysis • Low risk of bias Model fit statistics were not investigated. "To validate the risk model we used an internal validation procedure based on bootstrap cross-validation in the following way: a new sample

Author (year)	Title	Study details	Quality assessment
		PE and DVT were classified as having a PE. All patients received low-molecular-weight heparin at therapeutic doses followed by acenocoumarol (target INR 2–3) for at least 3months. Patients entered the study at the time of discontinuation of oral anticoagulation." Exclusion criteria • <18 years • previous provoked VTE "Patients were excluded if they had VTE provoked by active cancer, surgery, trauma, immobility, previous hospitalization or pregnancy and the puerperium. VTE that occurred in association with hormonal therapy or a thrombophilic blood abnormality (protein C, protein S or antithrombin deficiency, homozygous factor V Leiden mutation, homozygous prothrombin G20210A mutation, anticardiolipin antibodies or lupus anticoagulant) were also excluded." Sample characteristics • Sample size 398 • %female 45.5% • Mean age (SD) 61.0 years *median, mean not given • VTE characteristics 67.8% PE 20.8% proximal DVT 11.3% distal DVT 7.0 months median AC therapy. Prognostic factor(s) • DAMOVES Age, Sex, Obesity (yes vs. no), Abnormal D-dimer (yes, no, not performed), factor VIII count, genetic thrombosphilia (yes vs. no), varicose veins (yes vs. no) *Some variables are entered as continuous data and	of 398 subjects was created by randomly drawing (with replacement) a subject from those of the original cohort and the recurrence rate is estimated in the new dataset (thus, although statistically very unlikely, it is theoretically possible that a new sample formed by 1000 replication of the same subject could be created). Overall risk of bias • Moderate Derivation study. Model fit statistics were not investigated however internal validation using bootstrapping was conducted. Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		therefore an online tool is required for points. Outcome of interest • recurrent VTE The end point of the study was recurrent symptomatic DVT by venography or colour duplex sonography or recurrent symptomatic PE confirmed by ventilation–perfusion scanning and/or spiral computed tomography. Outcome measures • OR • sensitivity/specificity/Likelihood ratios • AUC	
Kresoja (2019)	Prediction and prognostic importance of in-hospital major bleeding in a real-world cohort of patients with pulmonary embolism	Study type • Prospective cohort study Study details • Study location Germany • Study setting University of Gottingen Heart Centre • Study dates September 2008 to November 2016 Inclusion criteria • Acute PE confirmed by CTPA or VQ scan and to be treated with AC Exclusion criteria • <18 years • incomplete follow-up data • received thrombolysis treatment or was included in the PEITHO study, AMPLIFY study. • Included more than once in the study database (initial PE only was taken for this study)	Participant selection • Low risk of bias prospective cohort study Predictors • Low risk of bias Patients were recruited prospective however calculation of risk scores was done retrospectively, missing data was considered to be normal, and it is unclear how many participants had missing data. Outcome • Low risk of bias Sample size and participant flow • High risk of bias Low number of participants with the event. Analysis • Low risk of bias

Author (year)	Title	Study details	Quality assessment
		• Treatment with surgical embolectomy or interventional approaches • subsegmental PE and other acute cardiac, respiratory or inflammatory disease responsible for symptoms and hemodynamic status on admission. Sample characteristics • Sample size 522 • %female 53% • Mean age (SD) median 69 (IQR 56-78) years • VTE characteristics 62% unprovoked PE • % recurrent VTE 12% had a prior VTE • % liver disease 3.7% known liver disease • % bleeding history 4.7% • % active cancer 17% • % chronic heart failure 16% Prognostic factor(s) • HAS-BLED >160 mmHg systolic blood pressure: 1 point Abnormal liver function (history of cirrhosis, or bilirubin > 2x the upper limit of normal in association with aspartate aminotransferase/alanine aminotransferase/ alkaline phosphatase levels > 3x the upper limit of normal): 1 point	Overall risk of bias • Moderate Low number of participants experienced the event and scores were calculated retrospectively. Directness • Partially directly applicable Prediction is limited to short term, in-hospital major bleeding.

Author (year)	Title	Study details	Quality assessment
		Abnormal renal function (on dialysis, a history of kidney transplantation, or serum creatinine values > 200 µmol/L): 1 point Stroke: 1 point Bleeding history or predisposition: 1 point Liable INR (time within therapeutic range < 60%): 1 point Age 65 years+: 1 point Drug use (platelet inhibitors or NSAIDS): 1 point Alcohol abuse (>8 units/week): 1 point Low risk: 0 points Moderate risk: 1-2 points High risk 3+ points • VTE-BLEED Active cancer: 2 points Male with uncontrolled arterial hypertension: 1 points Anaemia: 1.5 points History of bleeding: 1.5 points Age ≥60 years old: 1.5 points Renal dysfunction: 1.5 points Low risk: 0-1 High risk: 2+ Outcome of interest • major bleeding Defined using Control of Anticoagulation Subcommitteee of the international Society on Thrombosis and Hemostasis	

Author (year)	Title	Study details	Quality assessment
		Outcome measures • sensitivity/specificity/Likelihood ratios • AUC	
Klok (2016)	Performance of five different bleeding-prediction scores in patients with acute pulmonary embolism	Study type • Prospective cohort study Study details • Study location Germany • Study setting University of Gottingen Heart Centre • Study dates October 2005 - July 2014 Inclusion criteria • Acute PE confirmed by CTPA or VQ scan and to be treated with VKA Exclusion criteria • <18 years Sample characteristics • Sample size 665; 448 with >30 days data • %female 55% • Mean age (SD) 64 (SD 17) years • VTE characteristics 51% unprovoked VTE • % recurrent VTE 30% had a prior VTE • % hypertension 59% • % renal disease	Participant selection • Low risk of bias prospective cohort study Predictors • Low risk of bias calculated prospectively. However, HEMORR2HAGES outcomes specifically were marked down as checks were not made for the presence of CYP 2C9 single-nucleotide polymorphisms which constitute a part of this tool. Outcome • Low risk of bias primary endpoint of study Sample size and participant flow • High risk of bias Low number of participants with the event. Analysis • Low risk of bias Overall risk of bias • Moderate Low number of participants experienced the event. Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		6.7% known renal insufficiency • % liver disease 2.5% known liver disease • % bleeding history 1.6% • % active cancer 18% • % chronic heart failure 18% Prognostic factor(s) • HAS-BLED >160 mmHg systolic blood pressure: 1 point Abnormal liver function (history of cirrhosis, or bilirubin > 2x the upper limit of normal in association with aspartate aminotransferase/alanine aminotransferase/ alkaline phosphatase levels > 3x the upper limit of normal): 1 point Abnormal renal function (on dialysis, a history of kidney transplantation, or serum creatinine values > 200 µmol/L): 1 point Stroke: 1 point Bleeding history or predisposition: 1 point Liable INR (time within therapeutic range < 60%): 1 point Age 65 years+: 1 point Drug use (platelet inhibitors or NSAIDS): 1 point Alcohol abuse (>8 units/week): 1 point Low risk: 0 points Moderate risk: 1-2 points High risk 3+ points • RIETE score 1	

Author (year)	Title	Study details	Quality assessment
		Recent bleeding: 2 points Abnormal renal function (Creatinine levels >1.2 mg/dl): 1.5 points Anemia: 1.5 points Age >75 years: 1 point Active malignancy: 1 point PE diagnosis: 1 point Low risk: 0 points Moderate risk: 1-4 points High risk >4 points • Kuijer score Age >60 years: 1.6 points Female sex: 1.3 points Malignancy: 2.2 points Low risk = 0 points Moderate risk = 1-2 points High risk = 3+ points • HEMORR2HAGES Hepatic or renal disease: 1 point Ethanol abuse: 1 point Malignancy: 1 point Age 75 years plus: 1 point Reduced platelet count or function: 1 point Prior bleeding: 2 points Hypertension: 1 point Anemia: 1 point Genetic factors (CYP2C9 single-nucleotide polymorphisms): 1 point Excessive fall risk: 1 point	

Author (year)	Title	Study details	Quality assessment
		Prior stroke/TIA: 1 point Low risk: 0-1 points Moderate risk: 2-3 points High risk >3 points • ATRIA score Anemia: 3 points severe renal disease (GFR < 30 mL min-1 or dialysis-dependent): 3 points 75 years+: 2 points previous bleed: 1 point hypertension: 1 point Low risk: 0-3 points Moderate risk: 4 points High risk: 5-10 points Outcome of interest • major bleeding Defined using Control of Anticoagulation Subcommitteee of the international Society on Thrombosis and Hemostasis Outcome measures • sensitivity/specificity/Likelihood ratios	
Klok (2017)	External validation of the VTE-BLEED score for predicting major bleeding in stable anticoagulated patients with venous thromboembolism	**Study type** • Retrospective analysis of a prospective cohort *using data from the HOKUSAI-VTE* **Study details** • Study location	**Participant selection** • Low risk of bias

Author (year)	Title	Study details	Quality assessment
		37 countries • Study setting *439 centres* • Study dates *The Hokusai-VTE study was between January 28, 2010, and October 31, 2012* • Duration of follow-up *12 months* **Inclusion criteria** • VTE • enrolled in HOKUSAI-VTE trial *Inclusion criteria • ≥18 years • Objectively confirmed symptomatic DVT or PE Exclusion criteria • Contraindication(s) for study drugs • Creatinine clearance <30ml/min • Active cancer with long-term LMWH treatment planned • Other indication requiring long-term anticoagulation • received therapeutic doses of any heparin for >48 hours, prior randomization or had one dose of VKA • Continued to receive aspirin for >100mg daily or received dual platelet therapy* **Sample characteristics** • Sample size *8,240* • Mean age (SD) *55.8 (16) years* • VTE characteristics *40% PE and DVT 60% DVT only* • % hypertension	**Predictors** • Low risk of bias **Outcome** • High risk of bias *only bleeding events occurring after 30 days of treatment were counted. However, it was also noted that only bleeds occurring "on-treatment" were considered* **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Moderate *only included bleeding events after the first 30 days of treatment.* **Directness** • Directly applicable

Author (year)	Title	Study details	Quality assessment
		26% arterial hypertension • % renal disease *11% renal dysfunction* • % bleeding history *1% history of bleeding* • % active cancer *2.5% active cancer* **Prognostic factor(s)** • VTE-BLEED Active cancer: 2 points Male with uncontrolled arterial hypertension: 1 points Anaemia: 1.5 points History of bleeding: 1.5 points Age ≥60 years old: 1.5 points Renal dysfunction: 1.5 points Low risk: 0-1 High risk: 2+ *Cancer: diagnosed within 6 months before diagnosis of venous thromboembolism (VTE) (excluding basal-cell or squamous-cell carcinoma of the skin), recently recurrent or progressive cancer or any cancer that required anti-cancer treatment within 6 months before the VTE was diagnosed.* *Hypertension: defined as systolic blood pressure ≥140 mmHg at baseline.* *Anemia: Haemoglobin <130 g/l in men or <120 g/l in women.* *History of bleeding: Including prior major or non-major clinically relevant bleeding event, rectal bleeding, frequent nose bleeding, or haematuria.*	

Author (year)	Title	Study details	Quality assessment
		Renal dysfunction: An estimated glomerular filtration rate (eGRF) <60 ml/min defined the presence of renal dysfunction: eGRF was calculated at baseline with the Cockcroft-Gault formula, which include serum creatinine, age, and body weight. **Outcome of interest** • Major bleeding *Bleeding was defined as major if it was overt and was associated with a decrease in haemoglobin of 2 g per decilitre or more, or required a transfusion of two or more units of blood, occurred in a critical site (i. e. intracranial, intraspinal, intraocular, pericardial, intra-articular, intramuscular with compartment syndrome, retroperitoneal), or contributed to death, based on the on the criteria of the International Society on Thrombosis and Hemostasis (ISTH) (12). All bleeding events were adjudicated by the independent Clinical Events Committee whose members were unaware of the treatment assignment. Only bleeds occurring on-treatment and after the first 30 days following randomisation were considered. Only bleed* **Outcome measures** • C-statistics • OR • sensitivity/specificity/Likelihood ratios	
Klok (2018)	Predictive value of venous thromboembolism (VTE)-BLEED to predict major bleeding and other adverse events in a practice-based	Study type • Prospective cohort study Study details	Participant selection • High risk of bias participants were excluded if they died, had a VTE recurrence or a major bleed in the first 30 days. There is a particularly high risk of bleeding in the first 30 days and this

Author (year)	Title	Study details	Quality assessment
	cohort of patients with VTE: results of the XALIA study	• Study location: 21 countries • Duration of follow-up Patients were followed for at least 12 months or until they died. • Study dates June 2012 - March 2014 Inclusion criteria • DVT DVT was confirmed according to current diagnostic standards. Exclusion criteria • <18 years • did not use AC treatment beyond first 30 days • death/recurrence/major bleeding in first 30 days • those switching drug early those who received a vitamin K antagonist for 1–14 days or parenteral anticoagulation for 3–14 days before they were switched to rivaroxaban ('early switchers') Sample characteristics • Sample size 4457; 804 received lowered dose after 30 day period and were excluded from sensitivity analysis which included remaining 3653 participants. • %female 46% • VTE characteristics DVT only; 90% DVT + PE: 9.8% active cancer 11% • average duration of AC treatment participants were treated with either rivaroxaban or conventional anticoagulation therapy, starting with a course of parenteral anticoagulants, usually followed by a vitamin K antagonist. The decision to treat a patient with rivaroxaban or conventional anticoagulants was at the attending physician's discretion.	censoring may bias predictive estimates. Predictors • Low risk of bias Outcome • Low risk of bias All outcome events were adjudicated by the independent Clinical Events Committee whose members were unaware of the treatment assignment. Sample size and participant flow • Low risk of bias Analysis • High risk of bias over 30% of participants had missing data for haemoglobin, systolic blood pressure and eGFR, each. Furthermore there were only 39 major bleeding events reported which poses a risk of wrongly estimating the models predictive ability Overall risk of bias • High participants with events in the first 30 days were excluded and there was a large amount of missing data for predictor variables of the model. Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		• % hypertension Mean systolic blood pressure 137 mmHG (SD19), 2179 participant had missing data. • haemoglobin mean 140 g/l. 1731 had missing data. • first eGFR reading <30 ml/min: 1.4% 30-50 ml/min: 5.0% 50 ml/min or greater 58% 1601 participants missing Prognostic factor(s) • VTE-BLEED Active cancer: 2 points Male with uncontrolled arterial hypertension: 1 points Anaemia: 1.5 points History of bleeding: 1.5 points Age ≥60 years old: 1.5 points Renal dysfunction: 1.5 points Low risk: 0-1 High risk: 2+ *Cancer: diagnosed within 6 months before diagnosis of venous thromboembolism (VTE) (excluding basal-cell or squamous-cell carcinoma of the skin), recently recurrent or progressive cancer or any cancer that required anti-cancer treatment within 6 months before the VTE was diagnosed.* *Hypertension: defined as systolic blood pressure ≥140 mmHg at baseline.* *Anemia: Haemoglobin <130 g/l in men or <120 g/l in women.* *History of bleeding: Including prior major or non-major clinically relevant bleeding event, rectal bleeding, frequent nose bleeding, or haematuria.* *Renal dysfunction: An estimated glomerular filtration rate (eGRF) <60 ml/min defined the presence of renal*	

Author (year)	Title	Study details	Quality assessment
		dysfunction: eGRF was calculated at baseline with the Cockcroft-Gault formula, which include serum creatinine, age, and body weight. Outcome of interest • major bleeding Major bleeding was defined as overt bleeding associated with a decrease in haemoglobin of 20 g/l or more, or requiring a transfusion of two or more units of blood (or red blood cell concentrates), occurred in a critical site (i.e. intracranial, intraspinal, intraocular, pericardial, intra-articular, intramuscular with compartment syndrome, retroperitoneal), or contributed to death, based on the on the criteria of the International Society on Thrombosis and Haemostasis (ISTH) (Schulman & Kearon, 2005). Only treatment-emergent events, i.e. events occurring up to 2 days after terminating active anticoagulant treatment, were included in this post-hoc analysis. Outcome measures • C-statistics • HR crude and adjusted HR for score of 2+ • OR Crude and adjusted OR for 1-point score increase • sensitivity/specificity/Likelihood ratios 2x2 table possible	
Kooiman (2015)	The HAS-BLED Score Identifies Patients with Acute Venous Thromboembolism at High Risk of Major Bleeding Complications during the First Six Months of Anticoagulant Treatment	Study type • Retrospective cohort study Study details • Study location The Netherlands • Study setting Medical records from two sources (i.e. the three hospitals	Participant selection • High risk of bias retrospective chart review Predictors • High risk of bias The medical records of 388 of 537 patients lacked information on one or more items of the HAS-BLED score,

Author (year)	Title	Study details	Quality assessment
		and the Leiden anticoagulation clinic) were searched for information on patient characteristics at baseline, INR-values, major bleeding complications, and items on the HAS-BLED score. • Study dates starting VKA treatment between 2006 and 2007 • Duration of follow-up 180 days (6 months); median follow-up 179 days Inclusion criteria • VTE VTE diagnosis was objectified by computed tomographypulmonary angiography or ultrasound. Sample characteristics • Sample size 537 • %female • Mean age (SD) PE patients: 60.8 years DVT: 57.8 years • VTE characteristics 41.5% PE 58.5% DVT • % hypertension 31% • % renal disease PE patients: 3.6% DVT patients: 1.0% • % liver disease 1.0% • % bleeding history 3.5% • % NSAIDS/ antiplatelet use 11.2% Prognostic factor(s) • HAS-BLED >160 mmHg systolic blood pressure: 1 point	most frequently on alcohol use (331/537 patients). Outcome • Low risk of bias Participants were censored if they discontinued treatment. Sample size and participant flow • Low risk of bias Analysis • High risk of bias Low event rate suggests that the study was underpowered to estimate predictive validity of risk tool Overall risk of bias • High Retrospective study with a low event rate and missing data for one or more predictors for many participants Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		Abnormal liver function (history of cirrhosis, or bilirubin > 2x the upper limit of normal in association with aspartate aminotransferase/alanine aminotransferase/ alkaline phosphatase levels > 3x the upper limit of normal): 1 point Abnormal renal function (on dialysis, a history of kidney transplantation, or serum creatinine values > 200 µmol/L): 1 point Stroke: 1 point Bleeding history or predisposition: 1 point Liable INR (time within therapeutic range < 60%): 1 point Age 65 years+: 1 point Drug use (platelet inhibitors or NSAIDS): 1 point Alcohol abuse (>8 units/week): 1 point Low risk: 0 points Moderate risk: 1-2 points High risk 3+ points *Items were assessed at time of diagnosis of acute VTE, except for labile INR.* *Laboratory measurements on renal and liver function were recorded up to six months prior to diagnosis of acute VTE, with a preference for the values closest to the day of VTE diagnosis. Missing variables on the HAS-BLED score were scored as normal.* Outcome of interest • major bleeding The primary outcome was the occurrence of major bleeding events, defined by the International Society of Thrombosis and Hemostasis (ISTH) criteria; i.e. fatal bleeding; bleeding causing a drop in hemoglobin of at least 1.24 mmol/L; or requiring transfusion of at least 2 units of whole blood or red cells; or symptomatic bleeding in a critical organ or area (i.e.	

Author (year)	Title	Study details	Quality assessment
		intracranial, intraocular, intraspinal, retroperitoneal, intra-articular, pericardial or intramuscular with accompanying compartment syndrome). Outcome measures • HR • sensitivity/specificity/Likelihood ratios	
Kuijer (1999)	Prediction of the risk of bleeding during anticoagulant treatment for venous thromboembolism	Study type • Retrospective analysis of a prospective cohort This study included a derviation, internal validation and validation study, only the validation study (Performed on a separate participant group) was extracted as the former did not meet the PICO (sample size <250 participants), risk of bias was adjusted for this underpowering of initial model. Study details • Study setting Data for the present analysis were derived from the database of the Columbus Investigators study • Study dates The recruitment of patients began in November 1994 and ended in October 1995. The follow-up of the patients was completed in February 1996 • Follow-up 3 months Inclusion criteria • VTE in the Columbus investigators study !consecutive patients with objectively confirmed VTE were randomly allocated to receive an initial treatment with either subcutaneous low-molecular-weight heparin, reviparin sodium (Clivarin; Knoll AG, Ludwigshafen, Germany), 175 anti-factor Xa U/kg twice daily, or continuous intravenous unfractionated heparin with dose adjustments according to activated partial	Participant selection • High risk of bias taken from an RCT comparing different types of anticoagulants for which bleeding risk is not comparable Predictors • Low risk of bias Outcome • Low risk of bias All outcomes were assessed by an independent adjudication committee that was unaware of treatment allocation. Sample size and participant flow • High risk of bias both the derivation (<10 events per predictor) and validation cohorts were very underpowered (<100 events) Analysis • High risk of bias AC used was not controlled for and model fit was not assessed. Overall risk of bias • High Participants were recruited from an RCT comparing RCT drugs for bleeding risk, drug used was not controlled for in

Author (year)	Title	Study details	Quality assessment
		thromboplastin time. Both treatment groups received oral anticoagulant therapy (warfarin sodium or coumarin), which started on the first or second day of treatment, for at least 3 months, with a targeted therapeutic INR of 2.0 to 3.0. Study outcomes evaluated during a 3-month follow-up period were the incidence of symptomatic recurrent VTE, which was confirmed by objective investigations, and the incidence of bleeding complications." Sample characteristics • Sample size Validation study: 780 participants • %female 49% female • Mean age (SD) 60 (SD 17) years • VTE characteristics 75% DVT 25% PE Prognostic factor(s) • Kuijer score Age >60 years: 1.6 points Female sex: 1.3 points Malignancy: 2.2 points Low risk = 0 points Moderate risk = 1-2 points High risk = 3+ points Outcome of interest • major bleeding Bleeding was defined as major if it was clinically overt and associated with a decline in hemoglobin concentration of at least 20 g/L, if there was a need for transfusion of 2 U or more of red blood cells, if it was retroperitoneal or	the risk model. Model and validation were very underpowered to detect events. Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		intracranial, or if it warranted permanent discontinuation of treatment Outcome measures • sensitivity/specificity/Likelihood ratios 2x2 table possible	
Marcucci (2015)	Risk of recurrence after a first unprovoked venous thromboembolism: external validation of the Vienna Prediction Model with pooled individual patient data	Study type • Prospective cohort study validation model using data from 5 prospective studies (independent of the original VIENNA derivation cohort) Study details • Study location * secondary analysis of 5 prospective cohort studies • Study dates studies published between 2003 and 2013 • Duration of follow-up The overall median post anticoagulation follow-up was 22 months (25th, 75th percentiles: 14 months, 29 months) Inclusion criteria • VTE treated for at least 3 months Exclusion criteria • patient specific exclusion criteria missing or insufficient data for VIENNA model to be applied; • study specific exclusion criteria original VIENNA studies; Not a prospective cohort study; Sample characteristics • Sample size 904 • %female 39.5% female	Participant selection • High risk of bias secondary analysis of 5 studies varying in publication time from 2003 to 2013. Unclear variance / heterogeneity of prediction tool between the different cohorts. Predictors • Low risk of bias Outcome • Low risk of bias Sample size and participant flow • Low risk of bias Overall risk of bias • Low risk of bias Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		• Mean age (SD) 68 years *median age, mean not given. 59% aged over 65 years. • VTE characteristics 32.2% PE with or without DVT 66.5% proximal DVT 1.3% distal DVT • % recurrent VTE 13.6% Prognostic factor(s) • VIENNA tool *Considers D-dimer (as a continuous variable), age and index VTE location (distal DVT versus proximal DVT/PE)* Outcome of interest • recurrent VTE All outcome events were objectively confirmed by compression ultrasound, lung scanning, or computed tomography.	
Nieto (2013)	Validation of a score for predicting fatal bleeding in patients receiving anticoagulation for venous thromboembolism	Study type • Retrospective analysis of a prospective cohort analysis using the RIETE registry (observational registry of consecutive VTE patients) Study details • Study location 12 countries (Spain, Italy, France, Israel, Portugal, Germany, Switzerland, Czech Republic, Macedonia, United States, Brazil and Ecuador) • Study setting analysis of RIETE registry • Study dates from database's inception (2001) to December 2011 • Duration of follow-up After discharge, all patients were followed-up for up to 3	Participant selection • Low risk of bias retrospective analysis however the data of RIETE was collected prospectively. Predictors • Low risk of bias Outcome • Low risk of bias Sample size and participant flow • Low risk of bias

Author (year)	Title	Study details	Quality assessment
		months in the outpatient clinic. During each visit, any signs or symptoms suggesting either DVT or PE recurrences or bleeding complications were noted. Most outcomes were classified as reported by the clinical centers. Inclusion criteria • VTE RIETE included consecutive patients with symptomatic, acute deep vein thrombosis (DVT) or pulmonary embolism (PE), confirmed by objective tests (contrast venography or ultrasonography for suspected DVT; pulmonary angiography, lung scintigraphy, or helical computed tomography scan for suspected PE) Exclusion criteria • Did not receive AC therapy • participating in a trial with a blinded therapy • included in prior derivation study Sample characteristics • Sample size 15,206 • %female 51% female • Mean age (SD) 35% over 75 years of age *mean age not given • VTE characteristics PE: 51% Bilateral DVT: 3.8% Distal DVT 11% • Anemia 55% • Low platelet count 2.5% • abnormal prothrombin time 7.4% • Creatinine clearance levels <30 ml/min	Analysis • Low risk of bias however the event rate for fatal bleeding specifically is quite low and runs the risk of a biased estimate of predicitve ability. Overall risk of bias • Low Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		7.4% Prognostic factor(s) • RIETE score 2 Immobility ≥4 days: defined as non-surgical patients who were confined to bed with bathroom privileges for ≥4 days in the 2-months prior to VTE diagnosis Recent major bleeding: major bleeding less than 30 days before VTE diagnosis Anemia: defined as hemoglobin b13 g/dL in men or b12 g/dL in women. Outcome of interest • major bleeding Major bleeding was defined as an overt bleed that required a transfusion of 2 or more units of blood, was retroperitoneal, spinal or intracranial, or was fatal. The causes of death were assigned by their attending physicians. • Fatal bleeding any death occurring within 7 days of a major bleeding episode, in the absence of an alternative cause of death. Outcome measures • C-statistics for both fatal and all major bleeding • sensitivity/specificity/Likelihood ratios for fatal bleeding alone	
Palareti (2018)	The American College of Chest Physician score to assess the risk of bleeding during anticoagulation in patients with venous thromboembolism	Study type • Retrospective analysis of a prospective cohort The START2 Register is an observational, multicentre, dynamic cohort study that enrolled adults (aged at least 18 years) who started anticoagulation therapy, whatever the	Participant selection • Low risk of bias Predictors • Unclear risk of bias unclear whether weighting was same as original model,

Author (year)	Title	Study details	Quality assessment
		drug and dosage used. Study details • Study location Italy • Study setting START2 register • Study dates database inception until July 2017 • Duration of follow-up database inception until July 2017; median follow-up 12 months Inclusion criteria • VTE first DVT episode of the lower limb and/or PE, taking anticoagulation therapy but for no more than 30 days at time of enrolment. A small number (n = 49) of the patients were treated for recurrent superficial vein thrombosis, a potential indication for extended anticoagulation according to the Italian Federation of Anticoagulation Clinics, and they were also included in the study. • expected follow-up of at least 12 months Exclusion criteria • <18 years Sample characteristics • Sample size 2,263 • %female 48.7% female • Mean age (SD) 67 (IQR 51 - 77) years *Median; mean value not given 29.3% >75 years • VTE characteristics	limited reporting of missing data. Outcome • Low risk of bias All documented bleeding events were reviewed centrally by an independent committee blinded to patients' characteristics and to other variables. Sample size and participant flow • High risk of bias low event rate for major bleeding suggests a risk of the model's predictive utility being wrongly estimated. Analysis • Low risk of bias Overall risk of bias • Moderate unclear reporting on missing data for predictor variable and low event rate. Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		56.6% DVT only 20.9% DVT + PE 20.3% isolated PE 68.2% unprovoked VTE • % renal disease 25.7% had CrCl<30-60 mL min-1 • % liver disease 2.9% had history of liver failure • % NSAIDS/ antiplatelet use 0.5% NSAID use 4.0% antiplatelet use • Anemia 3.5% • % active cancer 7.7% Prognostic factor(s) • ACCP age 66–75 year: 1 point age > 75 years: 2 points previous major bleeding: 1 point cancer (active) : 1 point metastatic cancer: 1 point renal failure (CrCl < 30–60 mL min-1) : 1 point liver failure (reported in the history) : 1 point thrombocytopenia (< 100 000) : 1 point previous stroke/TIA: 1 point diabetes anemia (Hb < 10) : 1 point antiplatelet therapy: 1 point poor anticoagulant control (time spent in the therapeutic range < 60%): 1 point comorbidity recent surgery (within 3 months from index event) : 1 point frequent falls (≥ 2 in the last year) : 1 point alcohol abuse (reported in the history) : 1 point non-steroidal anti-inflammatory drugs use : 1 point	

Author (year)	Title	Study details	Quality assessment
		Low risk: 0 Moderate risk: 1 High risk: 2+ Outcome of interest • major bleeding MB was defined according to the International Society on Thrombosis and Haemostasis criteria [14]) as: fatal, symptomatic in a critical organ (intracranial, intraspinal, intraocular, retroperitoneal, intraarticular, intramuscular with compartment syndrome, pericardial), or causing a fall in the hemoglobin level of at least 2 g/dl or leading to transfusion of at least two units of whole blood or red cells. Outcome measures • sensitivity/specificity/Likelihood ratios 2x2 tables available. • C-statistics Extracted for overall predictive classification and not for low, moderate and high risk individually.	
Piovella (2014)	Comparison of four scores to predict major bleeding in patients receiving anticoagulation for venous thromboembolism: findings from the RIETE registry	Study type • Retrospective analysis of a prospective cohort Study details • Study Location Spain, France, Italy, Israel, Argentina	Participant selection • Low risk of bias Predictors • Low risk of bias *High risk for RIETE 1 tool as there is potential that it is being validated using the same same as it was derived*

Author (year)	Title	Study details	Quality assessment
		• Study setting *Data from the RIETE registry.* • Duration of follow-up *90 days* Inclusion criteria • VTE • At least 3 months follow-up • receiving treatment for VTE • data available for all variables of tools assessed Sample characteristics • Sample size *8,576* • %female *52%* • Mean age (SD) *66 (SD 18) years* Prognostic factor(s) • RIETE score 1 • Kuijer score • Outpatient bleeding risk index Outcome of interest • major bleeding *Bleeding was defined as major if it was clinically overt and*	*from.* Outcome • Low risk of bias Sample size and participant flow • Low risk of bias Analysis • Low risk of bias Overall risk of bias • Low Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		associated with a need for transfusion of two units or more of red blood cells, if it was retroperitoneal or intracranial, if it warranted permanent discontinuation of treatment, or it was fatal. Final classification was based on the full consensus of this committee. Outcome measures • C-statistics • sensitivity/specificity/Likelihood ratios **Note that sensitivity/specificity and likelihood ratios were calculated from the raw data rather than using the accuracy data provided in the paper.*	
Poli (2013)	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	Study type • Retrospective analysis of a prospective cohort Study details • Study location Italy • Study setting Twenty-seven centres affiliated to the Italian Federation of Anticoagulation Clinics participated in the EPICA Study, which prospectively followed 4093 very elderly patients who started vitamin K antagonist (VKA) treatment after the age of 80 years for either atrial fibrillation (AF) or VTE. Inclusion criteria • VTE • over 80 years of age • receiving treatment for VTE unclear whether participants are receiving AC prophylaxis or treatment.	Participant selection • High risk of bias Participants were selected from a VTE and/or atrial fibrillation trial, with only the VTE participants retained. It is unclear how long these participants have been on treatment at the point of enrolment. Predictors • Unclear risk of bias unclear whether any data were missing Outcome • Low risk of bias Sample size and participant flow • High risk of bias low number of events suggests that study risks biased estimated of tools' predictive values.

Author (year)	Title	Study details	Quality assessment
		Sample characteristics • Sample size 1,078 • %female 62.8% female • Mean age (SD) 84 years (range 80-89) * median, mean not given • % hypertension 62.9% • % NSAIDS/ antiplatelet use 5.1% • % active cancer 10.1% Prognostic factor(s) • HAS-BLED >160 mmHg systolic blood pressure: 1 point Abnormal liver function (history of cirrhosis, or bilirubin > 2x the upper limit of normal in association with aspartate aminotransferase/alanine aminotransferase/ alkaline phosphatase levels > 3x the upper limit of normal): 1 point Abnormal renal function (on dialysis, a history of kidney transplantation, or serum creatinine values > 200 µmol/L): 1 point Stroke: 1 point Bleeding history or predisposition: 1 point Liable INR (time within therapeutic range < 60%): 1 point Age 65 years+: 1 point Drug use (platelet inhibitors or NSAIDS): 1 point Alcohol abuse (>8 units/week): 1 point Low risk: 0 points	Analysis • Low risk of bias Overall risk of bias • High Unclear length of AC time at point of enrolment and low event rate poses risk of wrongly estimating the predictive utility of models Directness • Partially directly applicable study suggests that participants were receiving secondary prevention for VTE, it is likely that participants had already received treatment at point of enrolment.

Author (year)	Title	Study details	Quality assessment
		Moderate risk: 1-2 points High risk 3+ points • ACCP age 66–75 year: 1 point age > 75 years: 2 points previous major bleeding: 1 point cancer (active) : 1 point metastatic cancer: 1 point renal failure (CrCl < 30–60 mL min-1) : 1 point liver failure (reported in the history) : 1 point thrombocytopenia (< 100 000) : 1 point previous stroke/TIA: 1 point diabetes anemia (Hb < 10) : 1 point antiplatelet therapy: 1 point poor anticoagulant control (time spent in the therapeutic range < 60%): 1 point comorbidity recent surgery (within 3 months from index event) : 1 point frequent falls (≥ 2 in the last year) : 1 point alcohol abuse (reported in the history) : 1 point non-steroidal anti-inflammatory drugs use : 1 point Low risk: 0 Moderate risk: 1 High risk: 2+ • RIETE score 1 Recent bleeding: 2 points Abnormal renal function (Creatinine levels >1.2 mg/dl): 1.5 points Anemia: 1.5 points	

Author (year)	Title	Study details	Quality assessment
		Age >75 years: 1 point Active malignancy: 1 point PE diagnosis: 1 point Low risk: 0 points Moderate risk: 1-4 points High risk >4 points • OBRI Age 65 years+: 1 point Previous stroke: 1 point GI bleeding in the last 2 weeks: 1 point Recent MI, anemia (hematocrit <30%) or renal insufficiency (creatinine >1.5mg/dl): 1 point Creatinine >1.5mg dL-1: 1 point diabetes mellitus: 1 point Low risk: 0 points Moderate risk: 1-2 points High risk: 3+ points • HEMORR2HAGES Hepatic or renal disease: 1 point Ethanol abuse: 1 point Malignancy: 1 point Age 75 years plus: 1 point Reduced platelet count or function: 1 point Prior bleeding: 2 points Hypertension: 1 point Anemia: 1 point Genetic factors (CYP2C9 single-nucleotide polymorphisms): 1 point	

Author (year)	Title	Study details	Quality assessment
		Excessive fall risk: 1 point Prior stroke/TIA: 1 point Low risk: 0-1 points Moderate risk: 2-3 points High risk >3 points • ATRIA score Anemia: 3 points severe renal disease (GFR < 30 mL min-1 or dialysis-dependent): 3 points 75 years+: 2 points previous bleed: 1 point hypertension: 1 point Low risk: 0-3 points Moderate risk: 4 points High risk: 5-10 points Outcome of interest • major bleeding Outcome measures • C-statistics	
Riva (2014)	Poor predictive value of contemporary bleeding risk scores during long-term treatment of venous thromboembolism. A multicentre retrospective cohort study	Study type • Retrospective cohort study *retrospective chart review of all patients with acute VTE referred to the Anticoagulation Clinics of five Italian hospitals*	Participant selection • Low risk of bias Predictors • High risk of bias *retrospective chart review therefore there is the potential for*

Author (year)	Title	Study details	Quality assessment
		Study details • Study location *Italy* • Study setting *five Italian hospitals* • Study dates *January 2010 to August 2012* • Duration of follow-up *Mean (SD) follow up was 8.82 (3.59) months* Inclusion criteria • VTE *including pulmonary embolism (PE), lower-extremities proximal or distal deep-vein thrombosis (DVT), upper-extremities DVT, or extensive lower limbs superficial vein thrombosis close to the sapheno-femoral junction.* • receiving treatment for VTE *on a course of treatment of at least 3 months* Exclusion criteria • <18 years • Receiving treatment at time of index event Sample characteristics • Sample size *681* • %female *52%* • Mean age (SD) *median 63 (IQR 46-74) years*	*unclear reporting (or misreporting) of the various variables used in the tools.* Outcome • Low risk of bias Sample size and participant flow • High risk of bias *very low number of participants with major bleeding events (13) means that the study is underpowered to accurately assess the predictive utility of the different tools.* Analysis • High risk of bias *purpose of the paper was to assess the composite of major and clinically relevant non-major bleeding with limited reporting of major bleeding alone therefore key statistical analyses (such as sensitivity and specificity) are missing for major bleeding alone.* Overall risk of bias • High *Study was a retrospective chart review with a low number of participants with major bleeding and limited statistical analyses for major bleeding alone.*

Author (year)	Title	Study details	Quality assessment
		• VTE characteristics *43% PE, 76% lower extremity DVT; 5% upper extremity DVT; 1.6% superficial vein thrombosis only. 58% unprovoked VTE* • % renal disease *2.0% severe renal failure* • Anaemia *23.6%* Prognostic factor(s) • HAS-BLED >160 mmHg systolic blood pressure: 1 point Abnormal liver function (history of cirrhosis, or bilirubin > 2x the upper limit of normal in association with aspartate aminotransferase/alanine aminotransferase/ alkaline phosphatase levels > 3x the upper limit of normal): 1 point Abnormal renal function (on dialysis, a history of kidney transplantation, or serum creatinine values > 200 µmol/L): 1 point Stroke: 1 point Bleeding history or predisposition: 1 point Liable INR (time within therapeutic range < 60%): 1 point Age 65 years+: 1 point Drug use (platelet inhibitors or NSAIDS): 1 point Alcohol abuse (>8 units/week): 1 point Low risk: 0 points Moderate risk: 1-2 points High risk 3+ points	Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		• ACCP age 66–75 year: 1 point age > 75 years: 2 points previous major bleeding: 1 point cancer (active) : 1 point metastatic cancer: 1 point renal failure (CrCl < 30–60 mL min-1) : 1 point liver failure (reported in the history) : 1 point thrombocytopenia (< 100 000) : 1 point previous stroke/TIA: 1 point diabetes anemia (Hb < 10) : 1 point antiplatelet therapy: 1 point poor anticoagulant control (time spent in the therapeutic range < 60%): 1 point comorbidity recent surgery (within 3 months from index event) : 1 point frequent falls (≥ 2 in the last year) : 1 point alcohol abuse (reported in the history) : 1 point non-steroidal anti-inflammatory drugs use : 1 point Low risk: 0 Moderate risk: 1 High risk: 2+ • RIETE 1 Recent bleeding: 2 points Abnormal renal function (Creatinine levels >1.2 mg/dl): 1.5 points Anemia: 1.5 points Age >75 years: 1 point Active malignancy: 1 point	

Author (year)	Title	Study details	Quality assessment
		PE diagnosis: 1 point Low risk: 0 points Moderate risk: 1-4 points High risk >4 points • Kuijer score Age >60 years: 1.6 points Female sex: 1.3 points Malignancy: 2.2 points Low risk = 0 points Moderate risk = 1-2 points High risk = 3+ points • OBRI Age 65 years+: 1 point Previous stroke: 1 point GI bleeding in the last 2 weeks: 1 point Recent MI, anemia (hematocrit <30%) or renal insufficiency (creatinine >1.5mg/dl): 1 point Creatinine >1.5mg dL-1: 1 point diabetes mellitus: 1 point Low risk: 0 points Moderate risk: 1-2 points High risk: 3+ points • HEMORR2HAGES Hepatic or renal disease: 1 point Ethanol abuse: 1 point Malignancy: 1 point	

Author (year)	Title	Study details	Quality assessment
		Age 75 years plus: 1 point Reduced platelet count or function: 1 point Prior bleeding: 2 points Hypertension: 1 point Anemia: 1 point Genetic factors (CYP2C9 single-nucleotide polymorphisms): 1 point Excessive fall risk: 1 point Prior stroke/TIA: 1 point Low risk: 0-1 points Moderate risk: 2-3 points High risk >3 points • ATRIA Anemia: 3 points severe renal disease (GFR < 30 mL min-1 or dialysis-dependent): 3 points 75 years+: 2 points previous bleed: 1 point hypertension: 1 point Low risk: 0-3 points Moderate risk: 4 points High risk: 5-10 points Outcome of interest • Major bleeding *MB was defined according to the International Society on Thrombosis and Haemostasis criteria (13), as:*	

Author (year)	Title	Study details	Quality assessment
		• Symptomatic bleeding in a critical organ (intracranial, intraspinal, intraocular, retroperitoneal, intraarticular, intramuscular with compartment syndrome, pericardial); *• Bleeding causing a fall in the haemoglobin level of at least 2 g/dl or leading to transfusion of at least two units of whole blood or red cells; • Fatal bleeding.* Outcome measures • AUC	
Rodger (2017)	Validating the HERDOO2 rule to guide treatment duration for women with unprovoked venous thrombosis: multinational prospective cohort management study	Study type • Prospective cohort study prospective validation management study (high risk participants could be treated for extended duration) Study details • Study location 7 countries • Study setting 44 secondary or tertiary care centres in seven countries. • Study dates Between November 2008 and February 2015 • Duration of follow-up 1-year Patients were contacted by telephone or seen in person at six months (two weeks either way). High risk patients were also telephoned or seen in person at one year (two weeks either way), low risk patients were seen in person at 1 year (two weeks either way) • Loss to follow-up 6 enrolled participants did not complete predictor factor evaluations Inclusion criteria • VTE	Participant selection • Low risk of bias Predictors • Low risk of bias Low overall risk for this domain however note that measurements were taken prior to anticoagulant discontinuation due to this being a management study Outcome • Low risk of bias Outcome adjudicators were not from the same institution as the study participant and therefore assessed blind to risk categorisation Sample size and participant flow • High risk of bias Management study (looking at how the tool would be used in actual practice and patients treated accordingly) in which a large segment of those deemed at high risk continued treatment and are therefore not relevant to this review (for VTE-recurrence outcome).

Author (year)	Title	Study details	Quality assessment
		first episode of major, symptomatic, objectively proved unprovoked VTE 5-12 months before enrolment in the absence of the following major VTE provoking factors: leg fracture or lower extremity plaster cast, immobilisation for more than three days, major surgery in the three months before the index VTE event, and no diagnosis of a malignancy in the past five years (with the exception of localised skin malignancy). Pregnancy associated VTE were not included. Index VTE associated with minor or weak risk factors such as travel, exogenous oestrogen, minor immobilisation, or minor surgery was considered unprovoked as it is unclear if the risk of recurrence in “weakly provoked” VTE is low enough to discontinue anticoagulants. • received prior anticoagulation Index VTE had to have been managed for 5-12 months with an appropriate anticoagulant treatment, including initial treatment with either unfractionated heparin, low molecular weight heparin, rivaroxaban, or apixaban, followed by 5-12 months of oral anticoagulant treatment with vitamin K antagonists (target international normalised ratio 2-3), dabigatran, rivaroxaban, apixaban, or edoxaban. Exclusion criteria • malignancy malignancy arising after diagnosis of index VTE but before enrolment were not eligible for inclusion. • unable or unwilling to give informed consent • <18 years • already discontinued AC therapy at point of enrolment • required ongoing AC • inaccessible for follow-up • planned exogenous oestrogen use Sample characteristics • Sample size 2785; 934 participants did not receive anticoagulation for	Overall risk of bias • Moderate This was a management study in which some high risk patients continued anticoagulation. Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		duration of study. 1851 received anticoagulation for duration of study • %female 44% female • Mean age (SD) 54.4 (SD 16.7) years. 30.5% over 65 years old • VTE characteristics 40% index DT only 39% index PE only 21% both • % abnormal D-dimer 48.9% over 250 UG/L • obesity (over 30 BMI) 37.2% Prognostic factor(s) • HERDOO2 tool Male sex: 2 points one of more of: hyperpigmentation, edema or redness of the leg: 1 point VIDAS D-dimer 250 µg/L: 1 point Obesity (body mass index 30 or more): 1 point Age 65 year or older: 1 point Low risk: 0-1 points High risk: 2+ points Outcome of interest • recurrent VTE • major bleeding	
Ruiz-Gimenez (2008)	Predictive variables for major bleeding events in patients presenting with documented acute venous thromboembolism. Findings from the RIETE Registry	Study type • Retrospective analysis of a prospective cohort derivation and validation study using prospective data from the RIETE register	Participant selection • Low risk of bias Predictors • Low risk of bias

Author (year)	Title	Study details	Quality assessment
		Study details • Study location Spain, France, Italy, Israel, Argentina • Study setting review of the RIETE register, containing data from multiple centres across multiple countries • Study dates database inception up until June 2007 • Duration of follow-up first 90 days of therapy Inclusion criteria • VTE symptomatic, acute deep vein thrombosis (DVT) or pulmonary embolism (PE), confirmed by objective tests (contrast venography or ultrasonography for suspected DVT; pulmonary angiography, lung scintigraphy, or helical computed tomography [CT] scan for suspected PE) Exclusion criteria • unable or unwilling to give informed consent • inaccessible for follow-up • participating in a trial with a blinded therapy Sample characteristics • Sample size derivation sample: 13,057 validation sample: 6,572 • %female derivation sample: 51% female validation sample: 50% female • Mean age (SD) derivation sample: 66 (SD17) years validation sample: 66 (SD17) years • VTE characteristics • Anemia derivation sample: 33% validation sample: 32%	however, missing data is not clearly reported. Outcome • Low risk of bias Sample size and participant flow • Low risk of bias Over 100 events in the validation cohort and over 10 events per predictor (used in univariate analysis) in the derivation model. Analysis • High risk of bias derivation sample used a univariate analysis to derive the model. Unclear whether the model attempted internal validation to test the model on the derivation data set. However, the original sample size was split into a derivation cohort (which was used to form the prediction model) and a validation cohort (for which the finalized model was applied). Overall risk of bias • Low Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		• % active cancer derivation sample: 21% validation sample: 20% • % inpatients derivation sample: 28% validation sample: 28% • % recent surgery derivation sample: 13% validation sample: 12% • % immobility derivation sample: 25% validation sample: 24% • % recent major bleeding derivation sample: 1.6% validation sample: 1.4% • % with platelet count <100,000/mm3 derivation sample: 2.3% validation sample: 2.5% • % with abnormal creatinine levels derivation sample: 14% validation sample: 14% Prognostic factor(s) • RIETE score 1 • derivation risk factors investigated *patient's baseline characteristics (gender, age >75years, body weight<70kg, inpatient vs. outpatient) VTE risk factors (previous VTE, Cancer, surgery, immobility for 4+ days in the 2 months prior to diagnosis). Underlying diseases (chronic lung disease, Chronic heart failure, recent major bleeding , Antiplatelet therapy, NSAIDs, Corticosteroid therapy). Laboratory tests (creatinine levels >1.2 mg/dl, platelet count <100,000/mm3, anemia) Clinically overt PE? Initial therapy (LMWH yes vs. no, LMWH mean dose, UFH yes vs. no, inferior vena cava filter?) Long-term therapy (LMWH yes vs. no, VKA yes vs. no)* Outcome of interest • Major bleeding Bleeding complications were classified as major if they were overt and required a transfusion of two units of blood or more, or were retroperitoneal, spinal or intracranial, or when	

Author (year)	Title	Study details	Quality assessment
		they were fatal. Outcome measures • sensitivity/specificity/Likelihood ratios	
Seiler (2017)	Derivation and validation of a novel bleeding risk score for elderly patients with venous thromboembolism on extended anticoagulation	Study type • Retrospective analysis of a prospective cohort Study details • Study location *Switzerland* • Study setting *data from the SWITCO65+ registry containing data from 5 Swiss university and 4 high-volume non-university hospitals.* • Study dates *September 2009 to December 2013* • Duration of follow-up *up to 36 months; median 28 months. Mean anticoagulation period of 16 months* Inclusion criteria • VTE *participants were excluded if they had thrombosis at a site other than lower limb, or if they had catheter related thrombosis* • received prior anticoagulation *already received 3 months anticoagulation* • 65 years of age or older	**Participant selection** • Low risk of bias **Predictors** • High risk of bias *High risk for Seiler score alone as this was derived in this study. Low risk for other tools as these were validated here.* **Outcome** • Low risk of bias **Sample size and participant flow** • High risk of bias *Validated tools: Low overall event rate (<100 participants with major bleeds)* *Seiler score: <10 events per variable (EPV) for the original derivation of the tool (3.9 EPV) however internal validation of the tool had an EPV closer to 10.* **Analysis** • Low risk of bias *Seiler score derived using competing risk regression*

Author (year)	Title	Study details	Quality assessment
		Exclusion criteria • unable or unwilling to give informed consent • inaccessible for follow-up • Insufficient spoken German or French Sample characteristics • Sample size *743* • %female *47% female* • Mean age (SD) *Median age 75 years (IQR 70-81)* • % hypertension *18.3% arterial hypertension* • % renal disease *7.3%* • % liver disease *0.4%* • % bleeding history *3.6%* • % active cancer *3.1%* Prognostic factor(s) • RIETE score 1 Recent bleeding: 2 points Abnormal renal function (Creatinine levels >1.2 mg/dl): 1.5 points Anemia: 1.5 points	*considering 17 candidate predictor variables.* **Overall risk of bias** • Moderate *Derivation study. *low risk of bias for the tools validated in this study.* **Directness** • Directly applicable

Author (year)	Title	Study details	Quality assessment
		Age >75 years: 1 point Active malignancy: 1 point PE diagnosis: 1 point Low risk: 0 points Moderate risk: 1-4 points High risk >4 points • Kuijer score Age >60 years: 1.6 points Female sex: 1.3 points Malignancy: 2.2 points Low risk = 0 points Moderate risk = 1-2 points High risk = 3+ points • OBRI Age 65 years+: 1 point Previous stroke: 1 point GI bleeding in the last 2 weeks: 1 point Recent MI, anemia (hematocrit <30%) or renal insufficiency (creatinine >1.5mg/dl): 1 point Creatinine >1.5mg dL-1: 1 point diabetes mellitus: 1 point Low risk: 0 points Moderate risk: 1-2 points High risk: 3+ points • Kearon score	

Author (year)	Title	Study details	Quality assessment
		Previous stroke: 1 point Previous peptic ulcer disease: 1 point Previous gastrointestinal bleeding: 1 point Renal impairment: 1 point Anemia: 1 point Thrombocytopenia: 1 point Liver disease: 1 point Diabetes mellitus: 1 point Use of antiplatelet therapy: 1 point Age 65 years+: 1 point Low: 0–1 points Intermediate: 2–3 points High: 4+ • Sieler score Previous major bleeding Active cancer Low physical activity Anemia Thrombocytopenia Antiplatelet drugs or NSAIDs Poor INR control Low risk: 0-1 point Moderate risk:2-3 points High risk: 4+ points Outcome of interest • Major bleeding *Major bleeding was defined as fatal bleeding, symptomatic bleeding into critical area or organ (intracranial, intraspinal, intraocular, retroperitoneal, intraarticular, pericardial, or*	

Author (year)	Title	Study details	Quality assessment
		intramuscular with compartment syndrome), or bleeding causing a fall in hemoglobin level of at least 20 g/l or leading to transfusion of two or more units of whole blood or red cells. Outcome measures • C-statistics	
Scherz (2013)	Prospective, multicenter validation of prediction scores for major bleeding in elderly patients with venous thromboembolism	Study type • Retrospective analysis of a prospective cohort Study details • Study location Switzerland • Study setting 5 university and 4 non-university hospitals • Study dates September 2009 - June 2011 • Duration of follow-up 90 days Inclusion criteria • VTE "We defined DVT as the acute onset of leg pain or swelling plus incomplete compressibility of a venous segment on ultrasonography or an intraluminal filling defect on contrast venography [10]. Because the iliac vein and the inferior vena cava may be technically difficult to compress, an iliac/caval DVT was defined as abnormal duplex flow patterns compatible with thrombosis or an intraluminal filling defect on contrast computed tomography (CT) or magnetic resonance imaging venography [11]. Given that ultrasonography has a reduced sensitivity and specificity for a distal DVT, patients with a distal DVT were included only if the incompressible distal vein transverse diameter was at least 5 mm [12]. We	Participant selection • Low risk of bias Predictors • High risk of bias Data on predictor variables was obtained retrospectively via chart review and therefore there is a risk of reporting/recording bias. Additionally, the database did not differentiate between stroke and TIA therefore the history of stroke, which is a variable in several tools, was expanded to include TIA. Outcome • Low risk of bias blinded committee adjudicated all bleeding events Sample size and participant flow • High risk of bias Low number of participants with the event therefore there is a risk of biased estimate of perdictive ability Analysis • Low risk of bias Overall risk of bias • Moderate

Author (year)	Title	Study details	Quality assessment
		defined PE as the acute onset of dyspnea, chest pain or syncope coupled with a new high-probability ventilation/perfusion lung scan; a new contrast filling defect on spiral CT or pulmonary angiography; or the new documentation of a proximal DVT either by venous ultrasound or contrast venography [13,14]." • 65 years of age or older Exclusion criteria • unable or unwilling to give informed consent • inaccessible for follow-up • Thrombosis at a different site than the lower limb Or catheter related thrombosis • Insufficient spoken German or French Sample characteristics • Sample size 663 participants • %female 45.4% • Mean age (SD) 75 (65-97) years *Median and range, mean not given • VTE characteristics DVT only 32.4% DVT+PE 13.7% PE only 53.8% • % renal disease 17.8% had renal impairment* • % liver disease 2.0% • % bleeding history 3.8% previous major bleed • % NSAIDS/ antiplatelet use 32.3% used platelet inhibitors • Anemia 38.3% • % active cancer	Retrospective chart review Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		14.6% Prognostic factor(s) • RIETE score 1 Recent bleeding: 2 points Abnormal renal function (Creatinine levels >1.2 mg/dl): 1.5 points Anemia: 1.5 points Age >75 years: 1 point Active malignancy: 1 point PE diagnosis: 1 point Low risk: 0 points Moderate risk: 1-4 points High risk >4 points • Kuijer score Age >60 years: 1.6 points Female sex: 1.3 points Malignancy: 2.2 points Low risk = 0 points Moderate risk = 1-2 points High risk = 3+ points • OBRI Age 65 years+: 1 point Previous stroke: 1 point GI bleeding in the last 2 weeks: 1 point	

Author (year)	Title	Study details	Quality assessment
		Recent MI, anemia (hematocrit <30%) or renal insufficiency (creatinine >1.5mg/dl): 1 point Creatinine >1.5mg dL-1: 1 point diabetes mellitus: 1 point Low risk: 0 points Moderate risk: 1-2 points High risk: 3+ points • Kearon score Previous stroke: 1 point Previous peptic ulcer disease: 1 point Previous gastrointestinal bleeding: 1 point Renal impairment: 1 point Anemia: 1 point Thrombocytopenia: 1 point Liver disease: 1 point Diabetes mellitus: 1 point Use of antiplatelet therapy: 1 point Age 65 years+: 1 point Low: 0–1 points Intermediate: 2–3 points High: 4+ Outcome of interest • major bleeding major bleeding in first 90 days following index event Outcome measures • sensitivity/specificity/Likelihood ratios	

Author (year)	Title	Study details	Quality assessment
		• AUC	
Tosetto (2017)	External validation of the DASH prediction rule: a retrospective cohort study	Study type • Retrospective cohort study validation study Study details • Study location Italy • Study setting Thrombosis research italian partnership (TRIP) study group comprised of several thrombosis centres in italy. • Duration of follow-up All TRIP centers regularly followed up their VTE patients with planned visits or telephone interviews for at least 2 years after their first VTE. time of D-dimer measurement and ended at the date of study closure (1 September 2016), or when the patient completed 2 years of follow-up. Inclusion criteria • VTE unprovoked first VTE (either proximal deep vein thrombosis or pulmonary embolism), i.e. a VTE that occurred in the absence of a transient risk factor such as surgery, trauma, active cancer, immobilization, or pregnancy/puerperium. VTE that occurred in association with hormonal therapy (oral contraceptives or hormone replacement therapy) was also included, as this weak risk factor is a component of the DASH score. • received prior anticoagulation must have received at least 3 months prior AC with a VKA or DOAC and not currently be receiving AC therapy at point of enrolment • D-dimer measurement 20-40 days after stopping AC • At least 3 months follow-up	Participant selection • Low risk of bias study was retrospective however only those participants with an initial D-dimer record were assessed and information on other predictor variable was likely to be assessed correctly retrospectively Predictors • Low risk of bias unclear predictor assessments were made without knowledge of outcome data however as predictors were dichotomous and pre-recorded this is unlikely to have affected results Outcome • Low risk of bias unclear whether outcome was assessed without knowledge of predictor variables of interest to this review however measurement of outcome is objective and unlikely to have been influenced by knowledge of predictor variables. Overall risk of bias • Low Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		Exclusion criteria • Known antithrombin deficiency • known antiphospholipid antibodies • resumed AC <2 weeks following D-dimer measurement Sample characteristics • Sample size 827 • %female 47.8% • Mean age (SD) 55.3 (SD17.5) years • VTE characteristics 100% recurrent VTE average duration of prior AC treatment 13 months • % abnormal D-dimer 21.8% • % recurrent VTE 12.1% • average duration of AC treatment with recurrent VTE event: mean 12.8 months without recurrent VTE event: mean 14.0 months • % thrombophilia with recurrent VTE event: 24% without recurrent VTE event: mean 14.1% • hormone therapy at time of index VTE 43.0% Prognostic factor(s) Abnormal D=dimer (measured ~1 months after stopping anticoagulation): 2 points Age <50 years: 1 point Male sex: 1 point	

Author (year)	Title	Study details	Quality assessment
		Hormone use at VTE onset (if female): -2 points Low risk: 0-1 point, high risk 2+ points Outcome of interest • recurrent VTE Recurrent VTE was diagnosed in the presence of clinical symptoms objectively confirmed by compression ultrasound, pulmonary computed tomography scan, pulmonary angiography, or highprobability perfusion scintigraphy.	
Zhang (2019)	Comparison of prediction value of four bleeding risk scores for pulmonary embolism with anticoagulation: A real-world study in Chinese patients.	Study type • Prospective cohort study Study details • Study location China • Study setting Beijing Chao-Yang hospital • Study dates January 2009 - September 2013 • Duration of follow-up 90 days Inclusion criteria • PE Diagnosed with acute symptomatic PE Exclusion criteria • Did not receive any anticoagulation • inaccessible for follow-up Sample characteristics • Sample size 578 participants • %female	Participant selection • Low risk of bias Predictors • Low risk of bias Outcome • Low risk of bias Sample size and participant flow • High risk of bias Low number of participants with the event therefore there is a risk of biased estimate of predictive ability Analysis • High risk of bias Risk scores used were 3-levels and it is unclear how this was translated into sensitivity/ specificity data (which requires 2 level input) as raw data is not given. Overall risk of bias • Moderate

Author (year)	Title	Study details	Quality assessment
		49.5% • Mean age (SD) 65 (53-74) years *Median and range, mean not given • % bleeding history 7.8% previous major bleed • % NSAIDS/ antiplatelet use 1.7% used platelet inhibitors • % active cancer 12.4% Prognostic factor(s) • RIETE score 1 Recent bleeding: 2 points Abnormal renal function (Creatinine levels >1.2 mg/dl): 1.5 points Anemia: 1.5 points Age >75 years: 1 point Active malignancy: 1 point PE diagnosis: 1 point Low risk: 0 points Moderate risk: 1-4 points High risk >4 points • Kuijer score Age >60 years: 1.6 points Female sex: 1.3 points Malignancy: 2.2 points Low risk = 0 points	Low event rate and unclear reporting. Directness • Directly applicable

Author (year)	Title	Study details	Quality assessment
		Moderate risk = 1-2 points High risk = 3+ points • Nieuwenhuis score WHO grade 1: 1 point WHO grade 2: 1 point History of bleeding diathesis: 2 point Recent trauma (<2 months) or surgery: 1 point Body surfacy area <2m^2: 1 point Low risk: 0-2 points Moderate risk: 3-4 points High risk: >5 points • Kearon score Previous stroke: 1 point Previous peptic ulcer disease: 1 point Previous gastrointestinal bleeding: 1 point Renal impairment: 1 point Anemia: 1 point Thrombocytopenia: 1 point Liver disease: 1 point Diabetes mellitus: 1 point Use of antiplatelet therapy: 1 point Age 65 years+: 1 point Low: 0–1 points Intermediate: 2–3 points High: 4+	

Author (year)	Title	Study details	Quality assessment
		Outcome of interest • major bleeding major bleeding in first 90 days following index event Outcome measures • AUC	

Appendix F – Forest plots

Sensitivity, specificity and LRs

Kuijer score: High vs moderate/low

Figure 1: Sensitivity, Kuijer score: High vs moderate/low

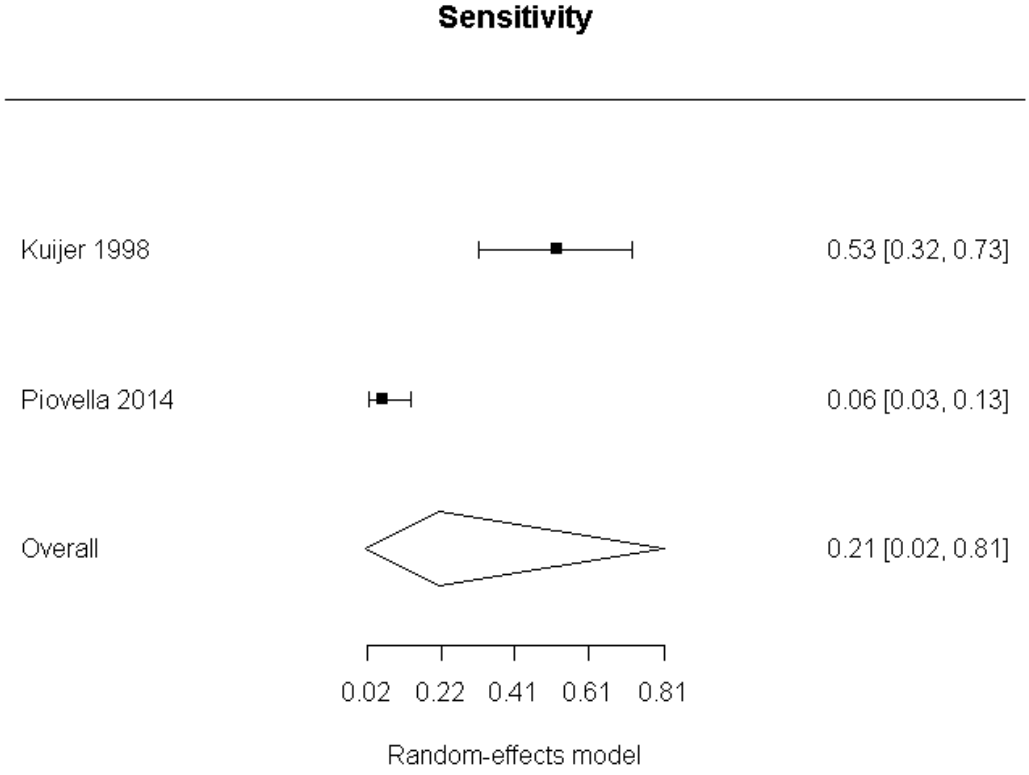


Figure 2: Specificity, Kuijer score: High vs moderate/low

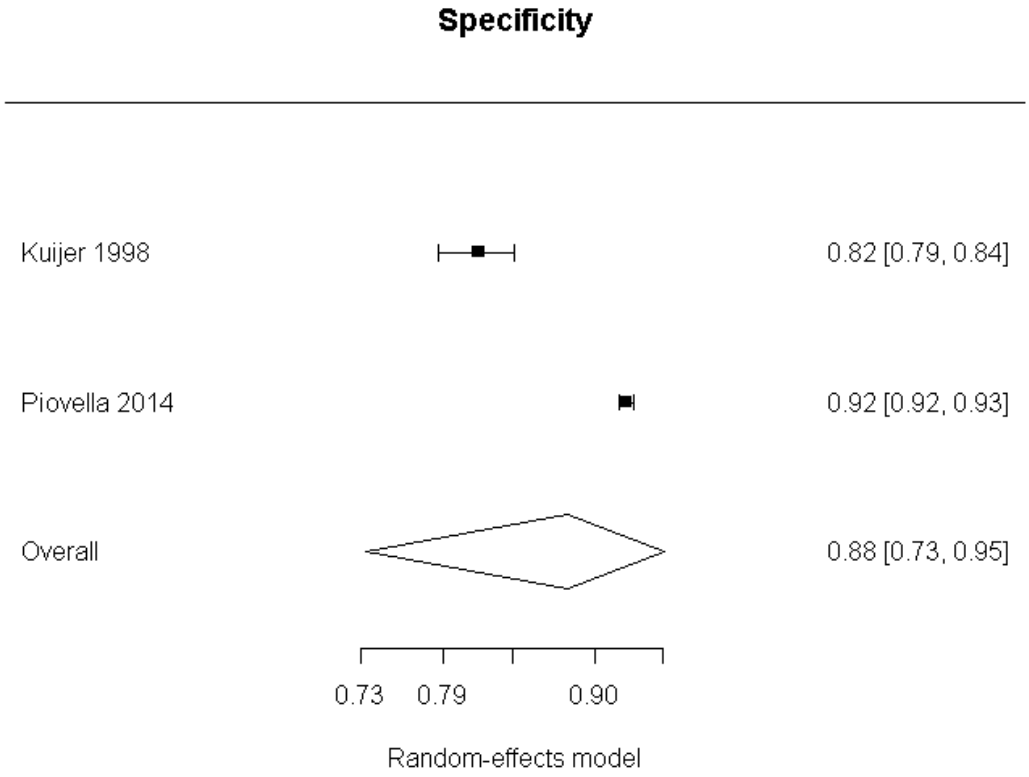


Figure 3: LR+, Kuijer score: High vs moderate/low

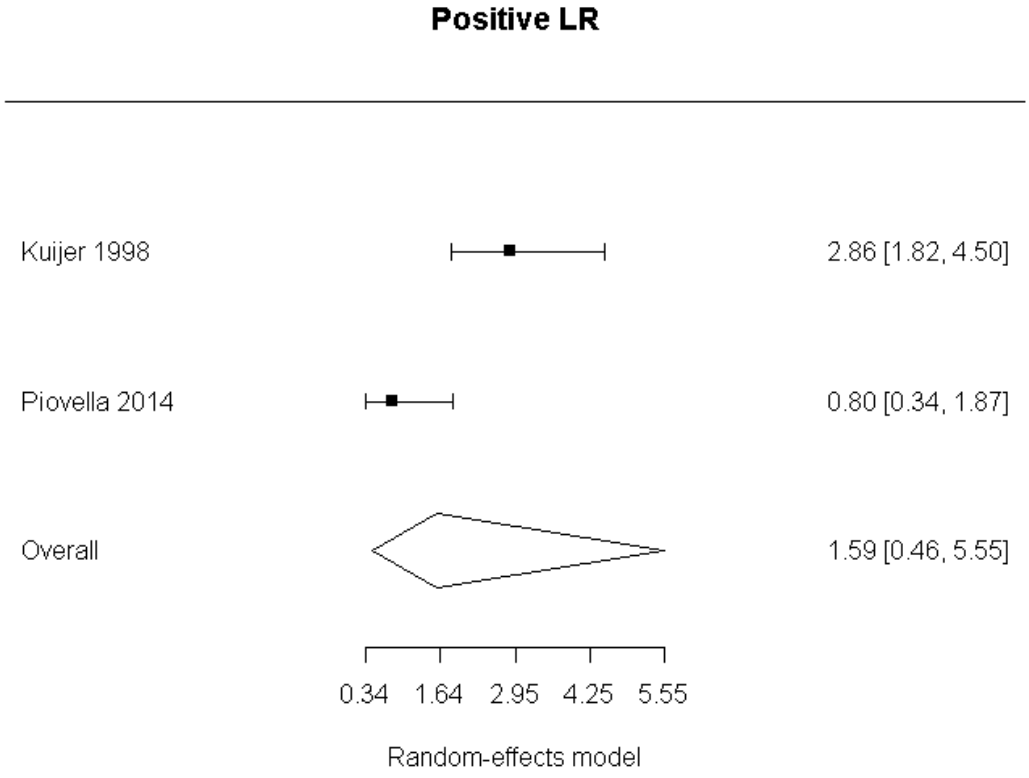
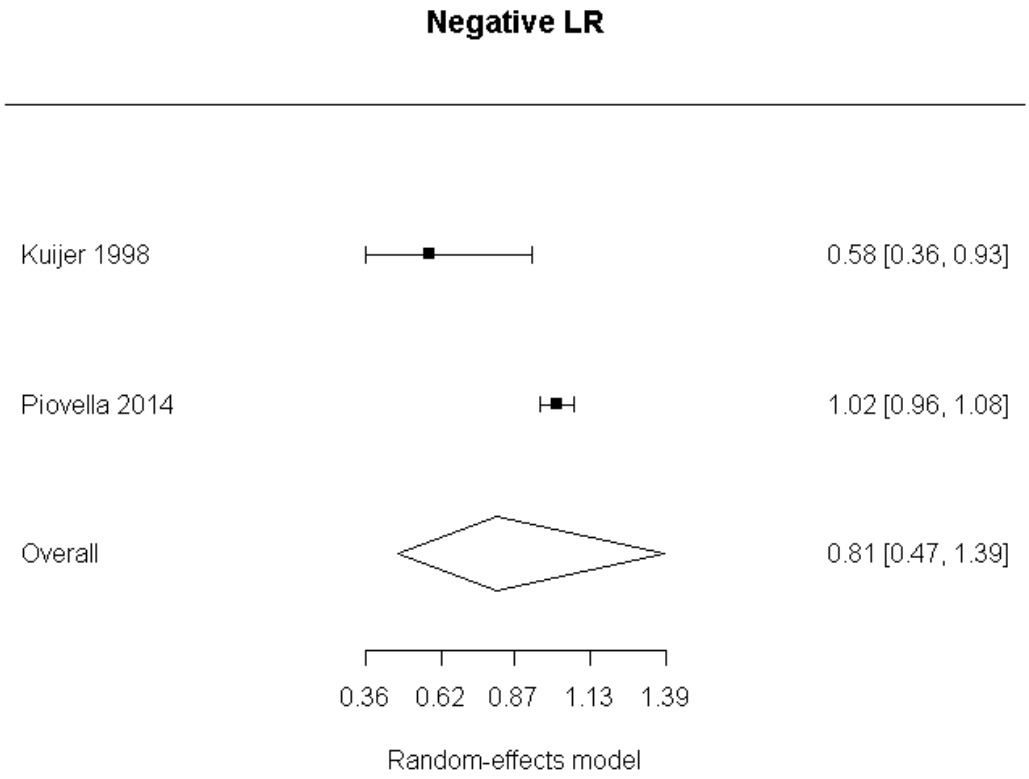


Figure 4: LR-, Kuijer score: High vs moderate/low



Kuijer score High/moderate vs low

Figure 5: Sensitivity, Kuijer score: High/moderate vs low

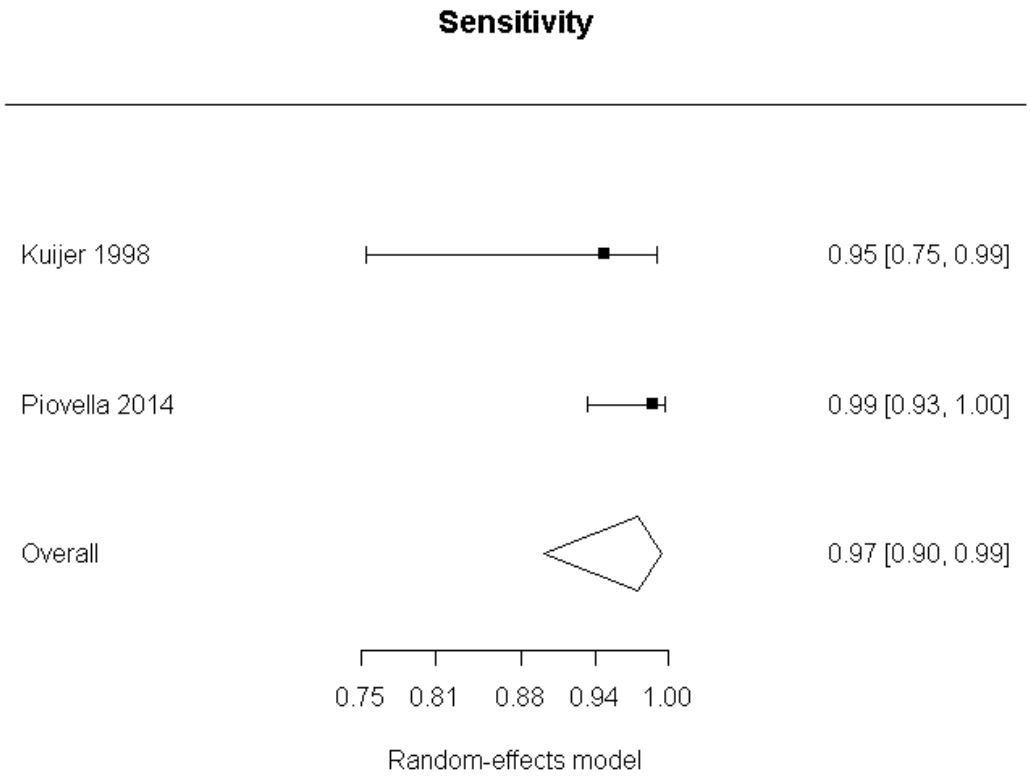


Figure 6: Specificity, Kuijer score: High/moderate vs low

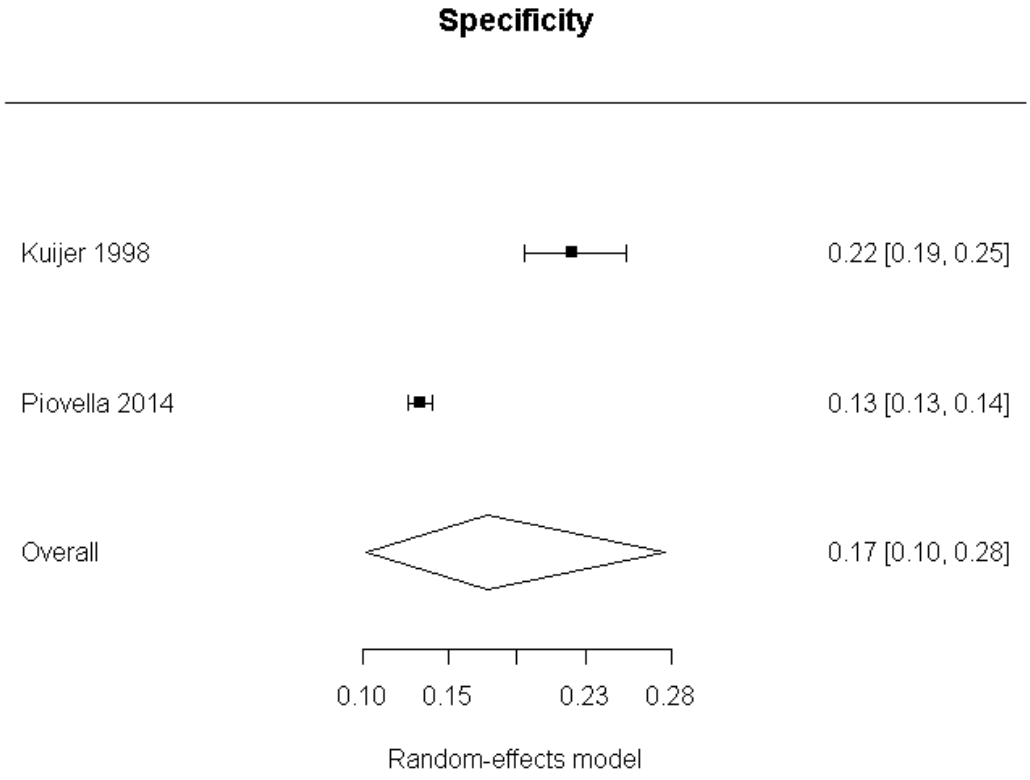


Figure 7: LR+, Kuijer score: High/moderate vs low

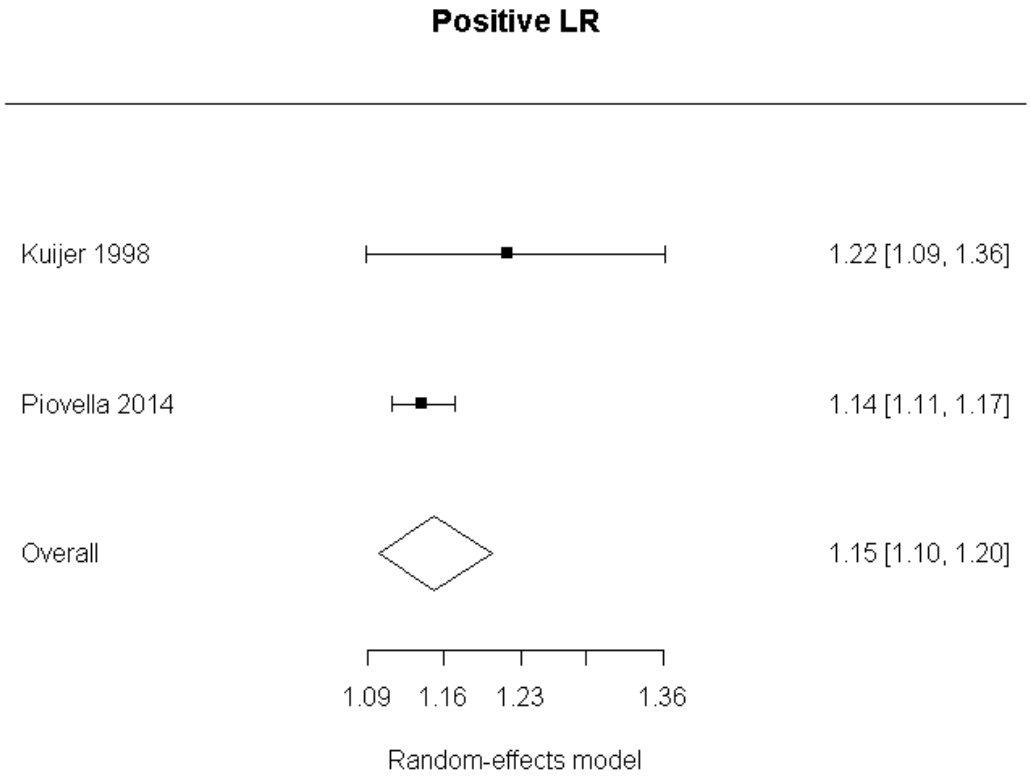
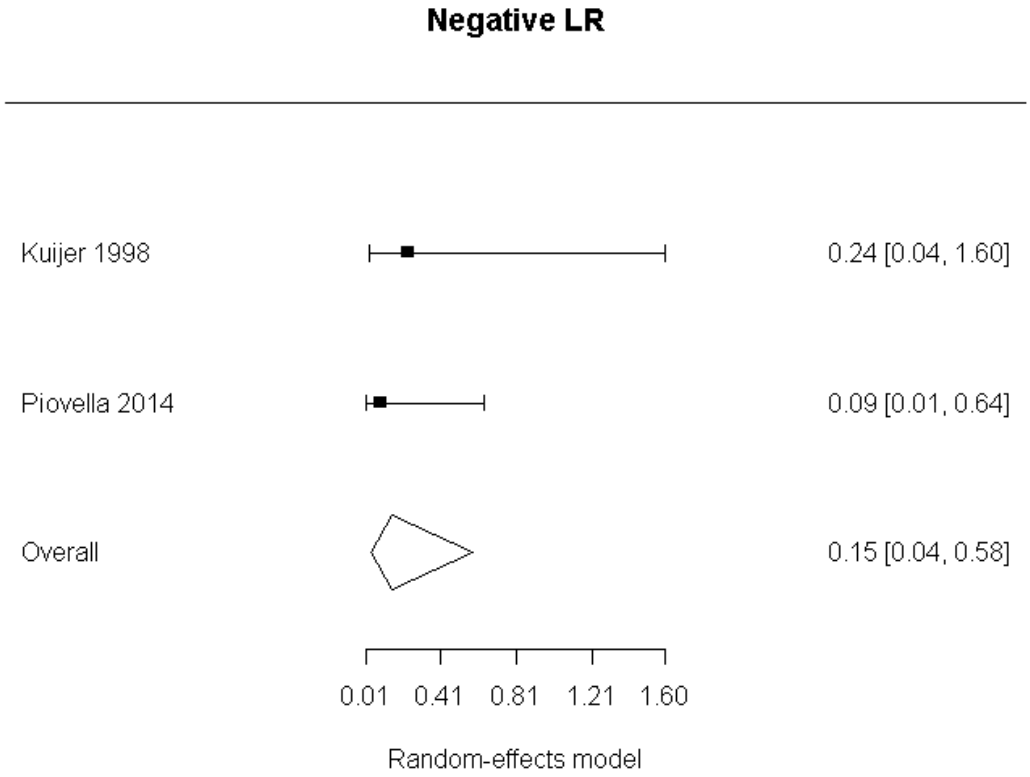


Figure 8: LR-, Kuijer score: High/moderate vs low



RIETE 1 High vs moderate/low

Figure 9: Sensitivity, RIETE 1: High vs moderate/low

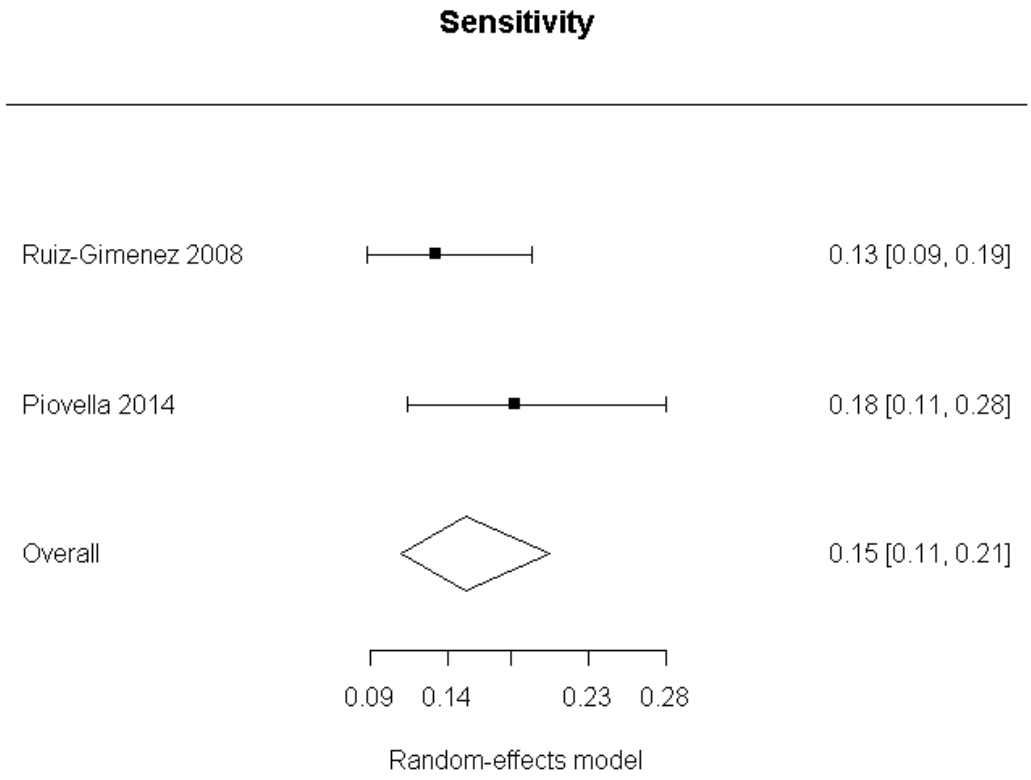


Figure 10: Specificity, RIETE 1: High vs moderate/low

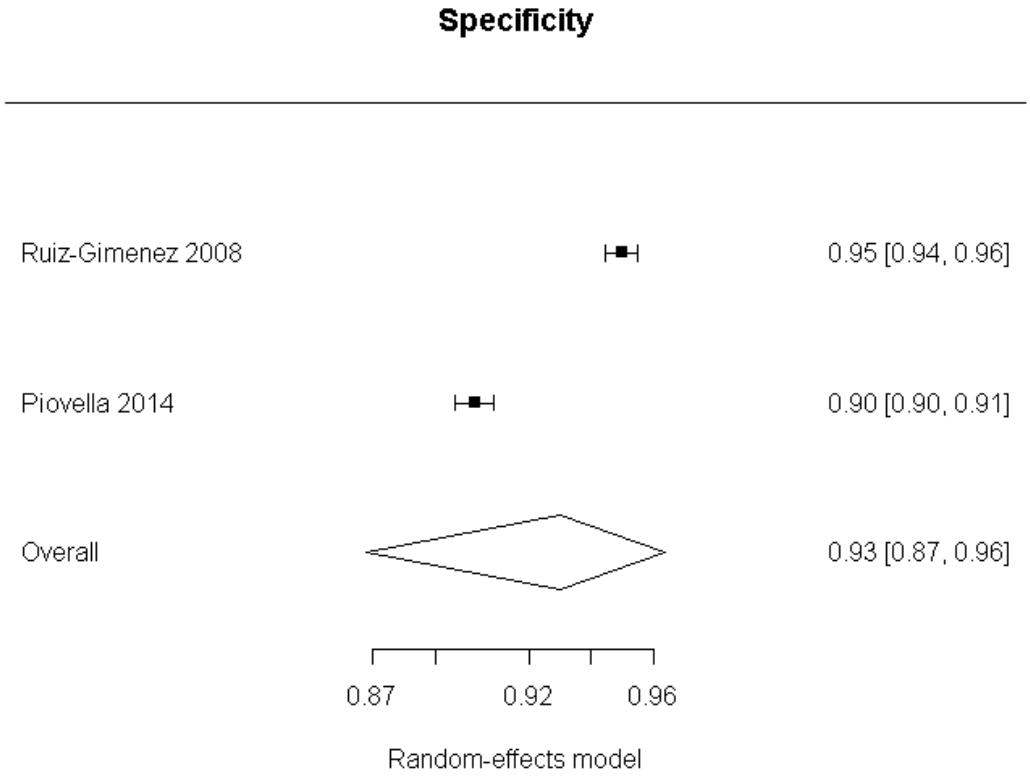


Figure 11: LR+, RIETE 1: High vs moderate/low

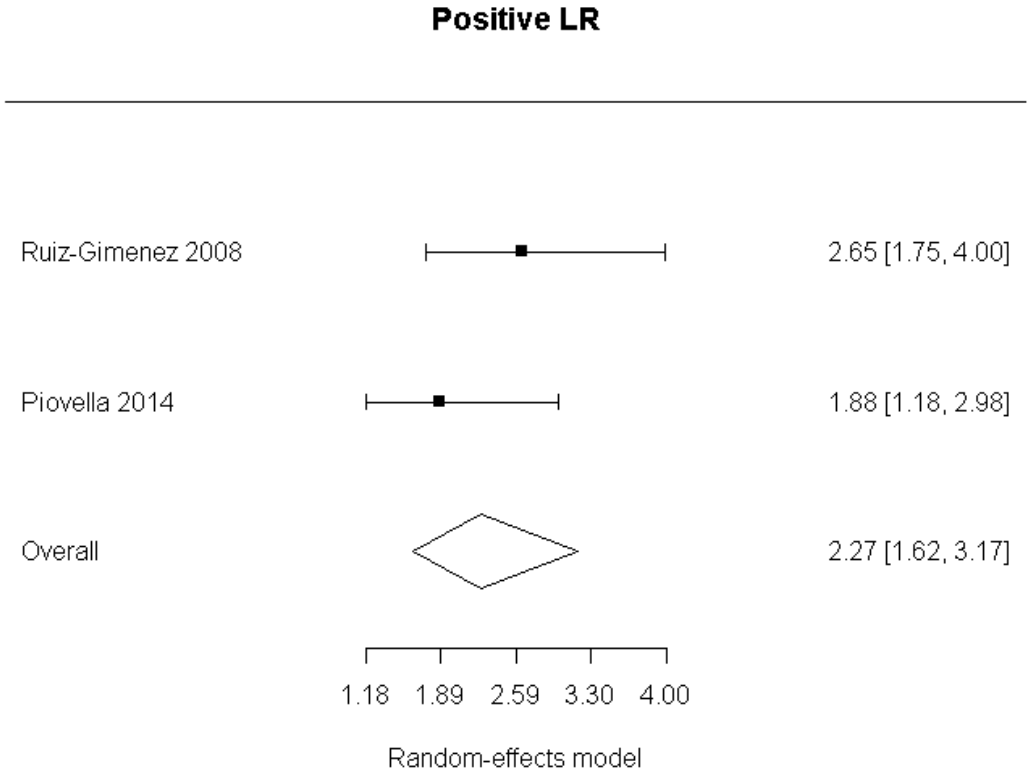
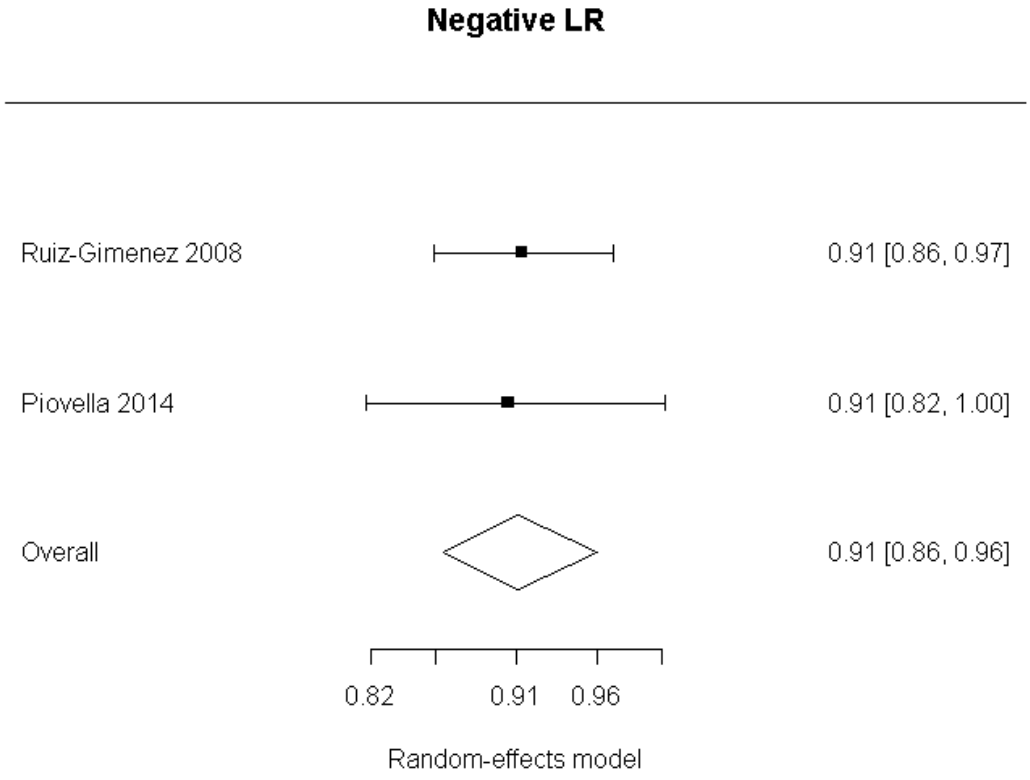


Figure 12: LR-, RIETE 1: High vs moderate/low



RIETE 1 High/moderate vs low

Figure 13: Sensitivity, RIETE 1: High/moderate vs low

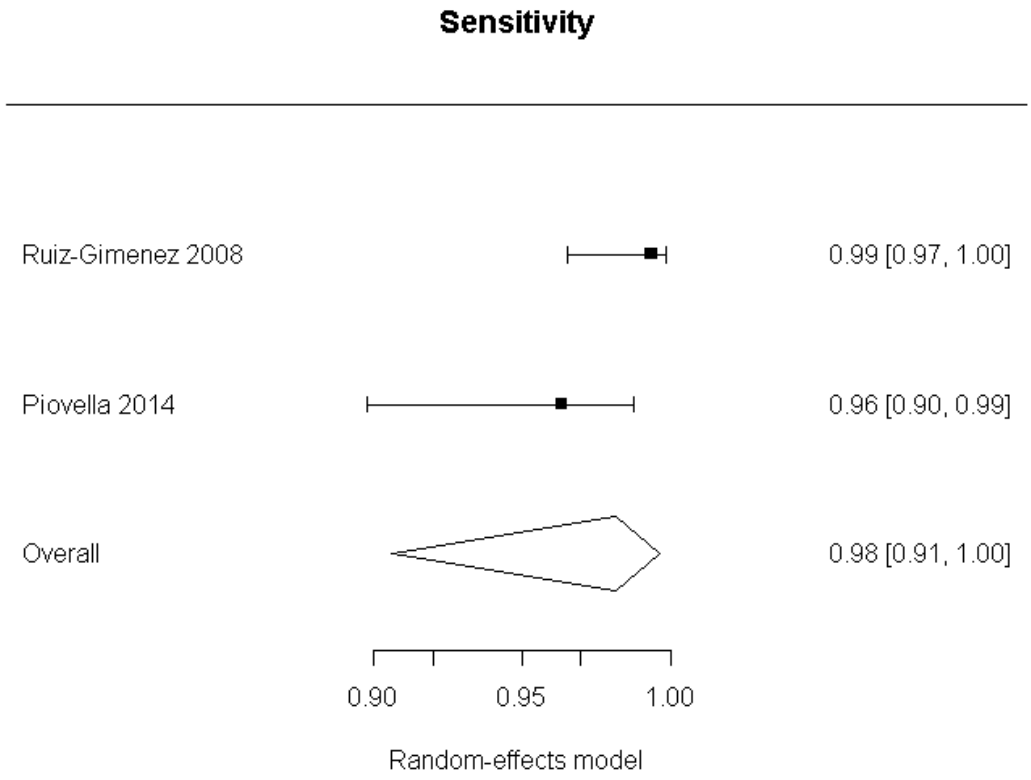


Figure 14: Specificity, RIETE 1: High/moderate vs low

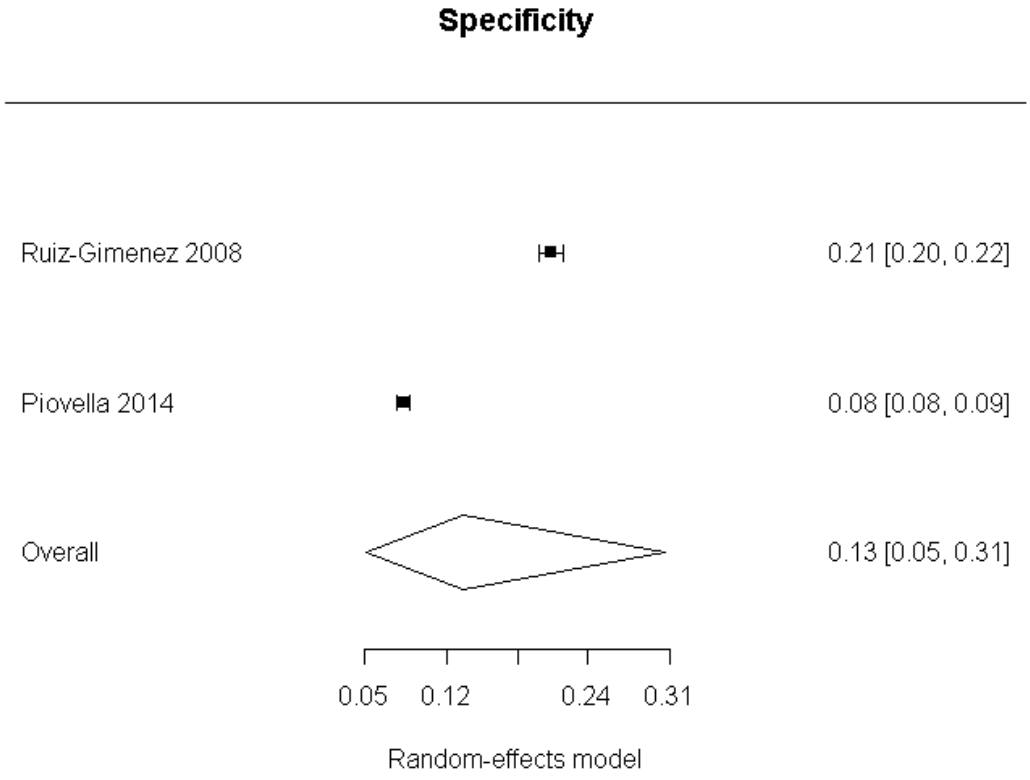


Figure 15: LR+, RIETE 1, High/moderate vs low

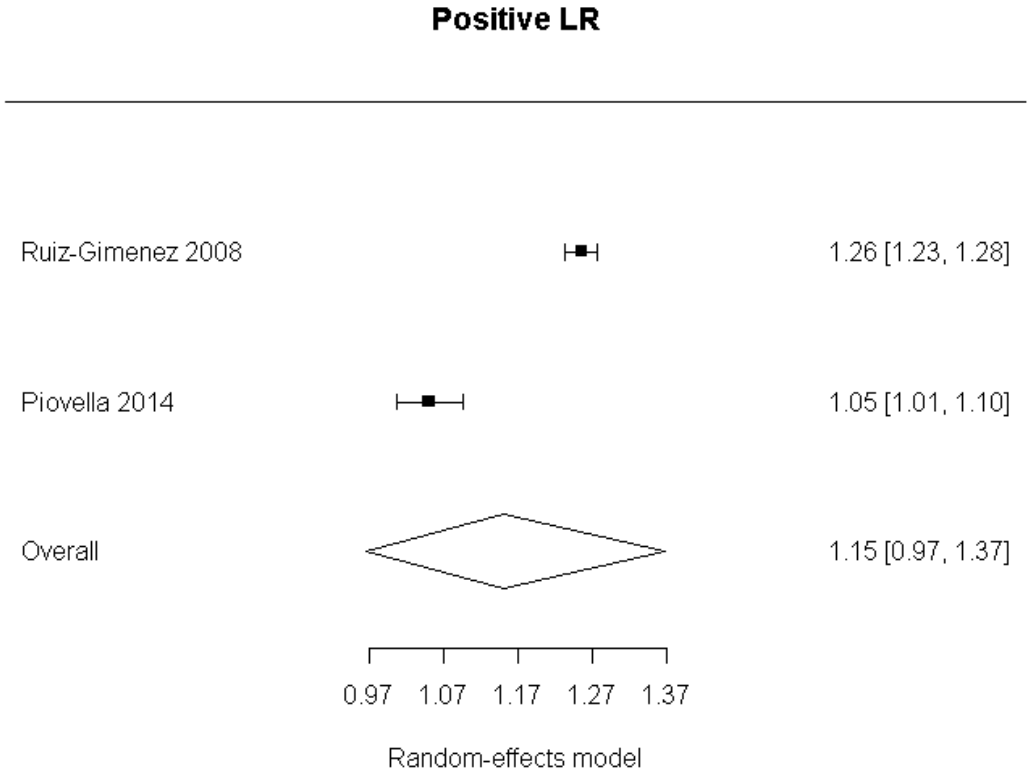
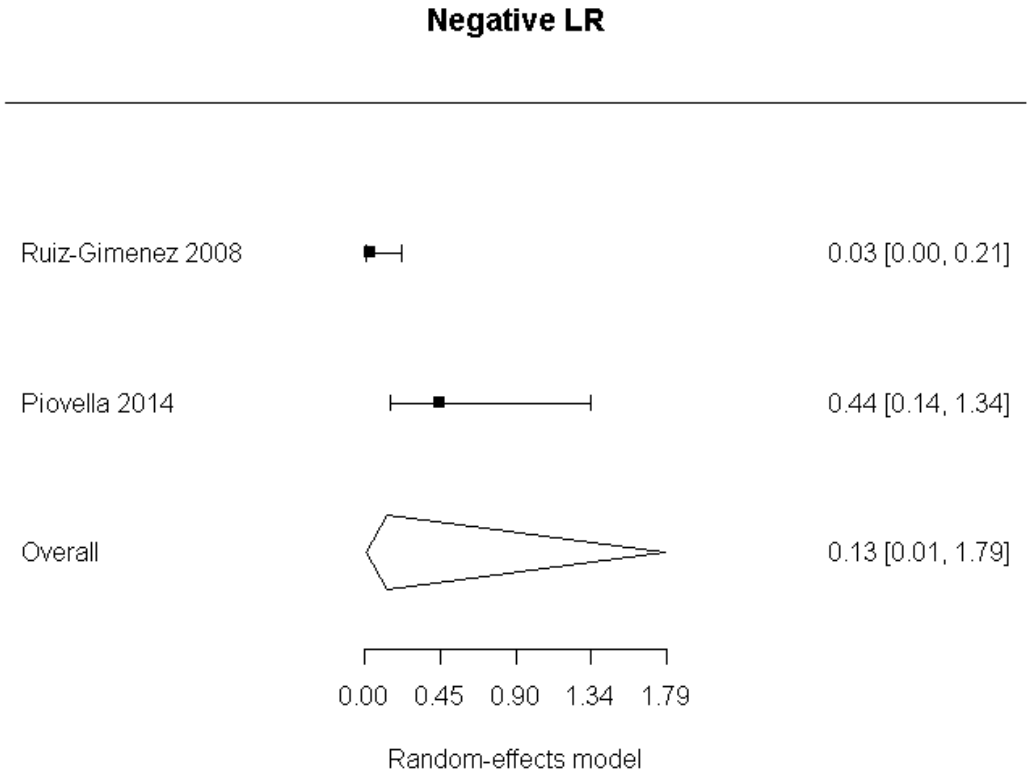


Figure 16: LR-, RIETE 1, High/moderate vs low



VTE BLEED

Figure 17: Sensitivity, VTE BLEED

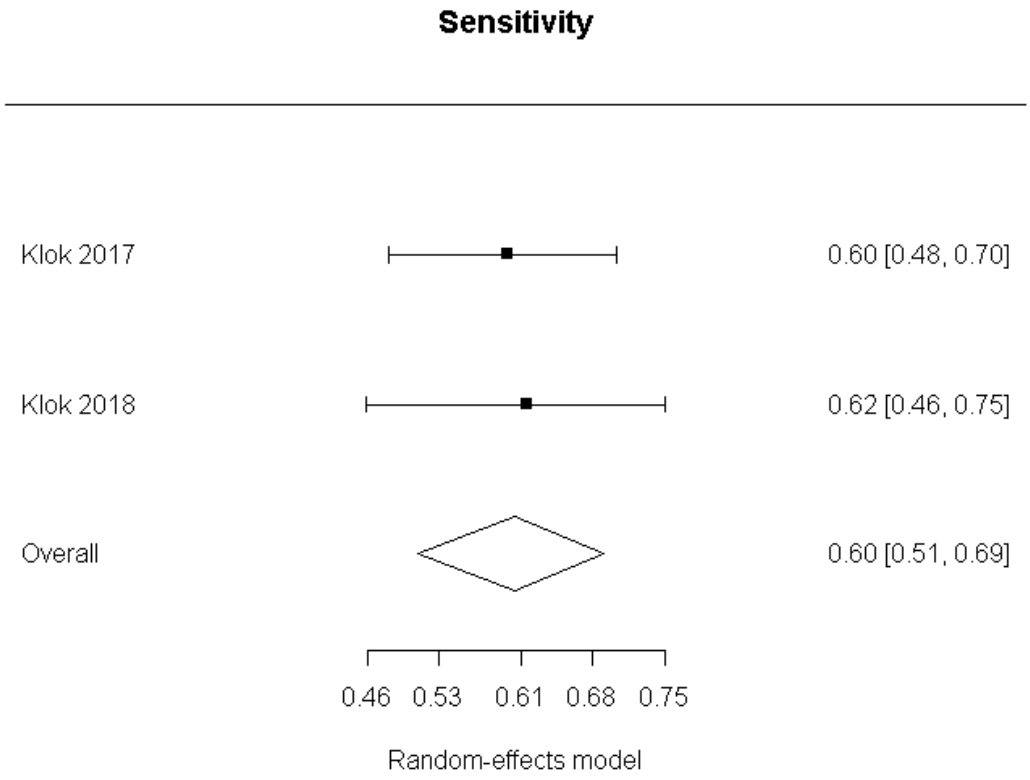


Figure 18: Specificity, VTE BLEED

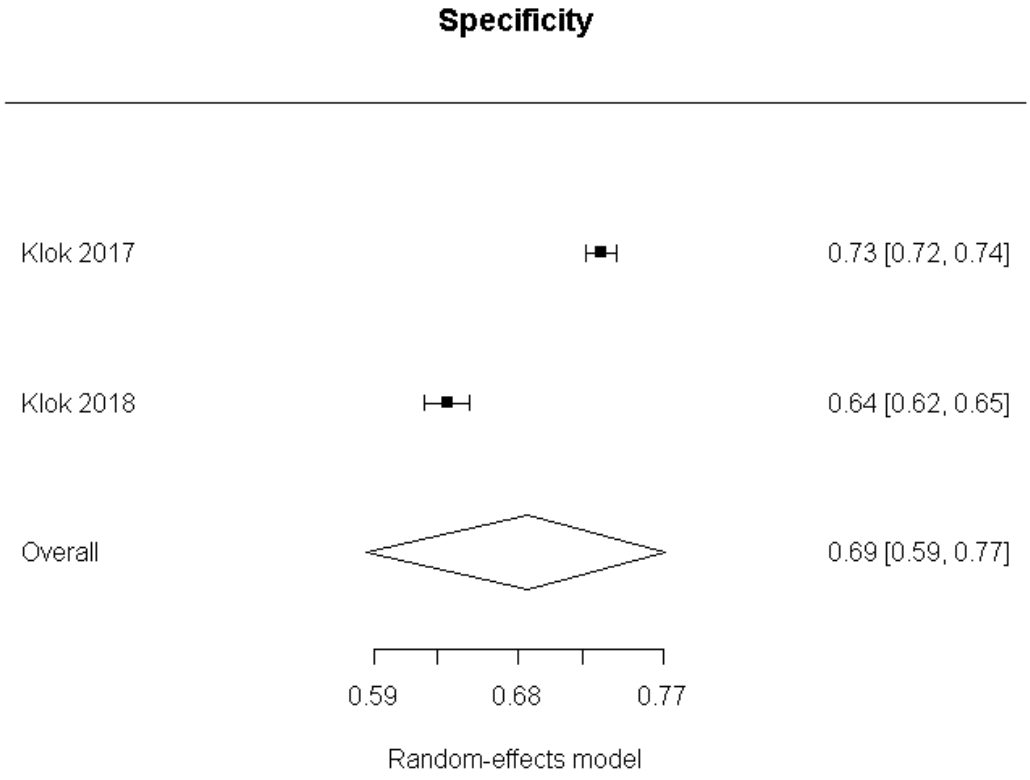


Figure 19: LR+, VTE BLEED

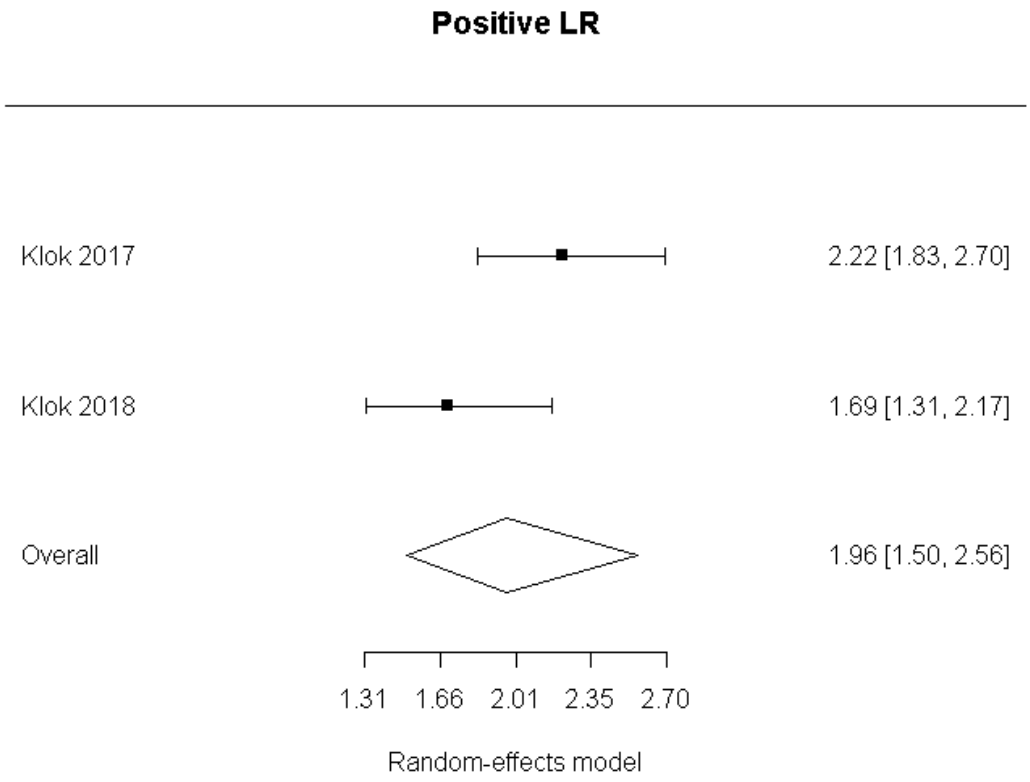
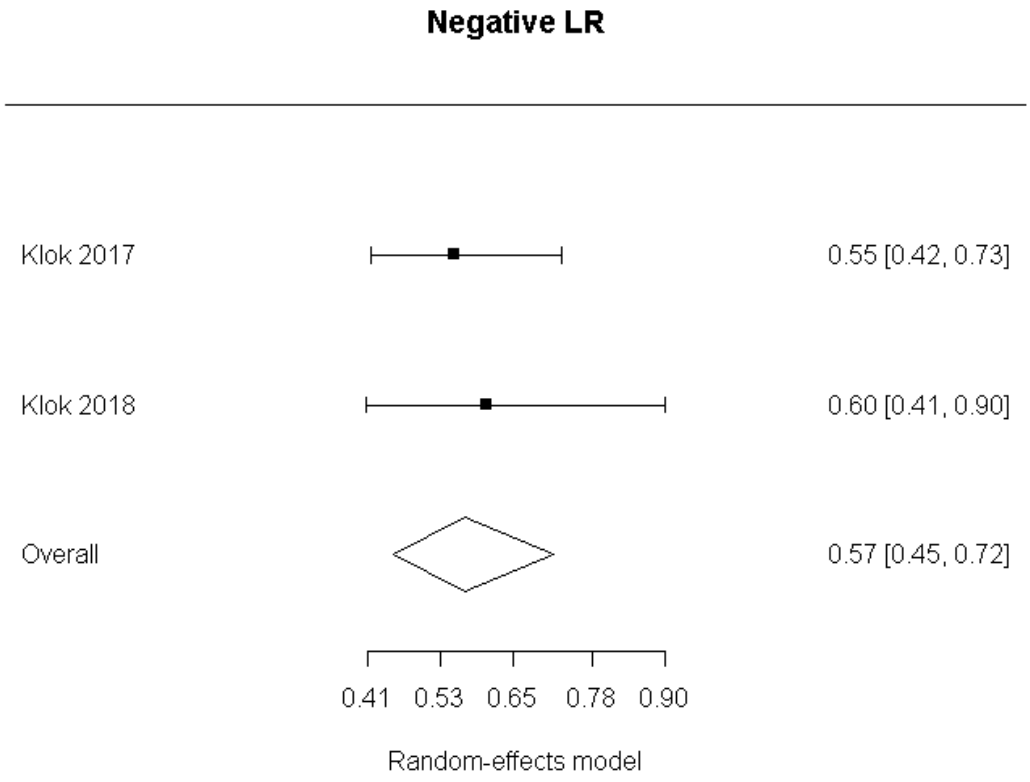


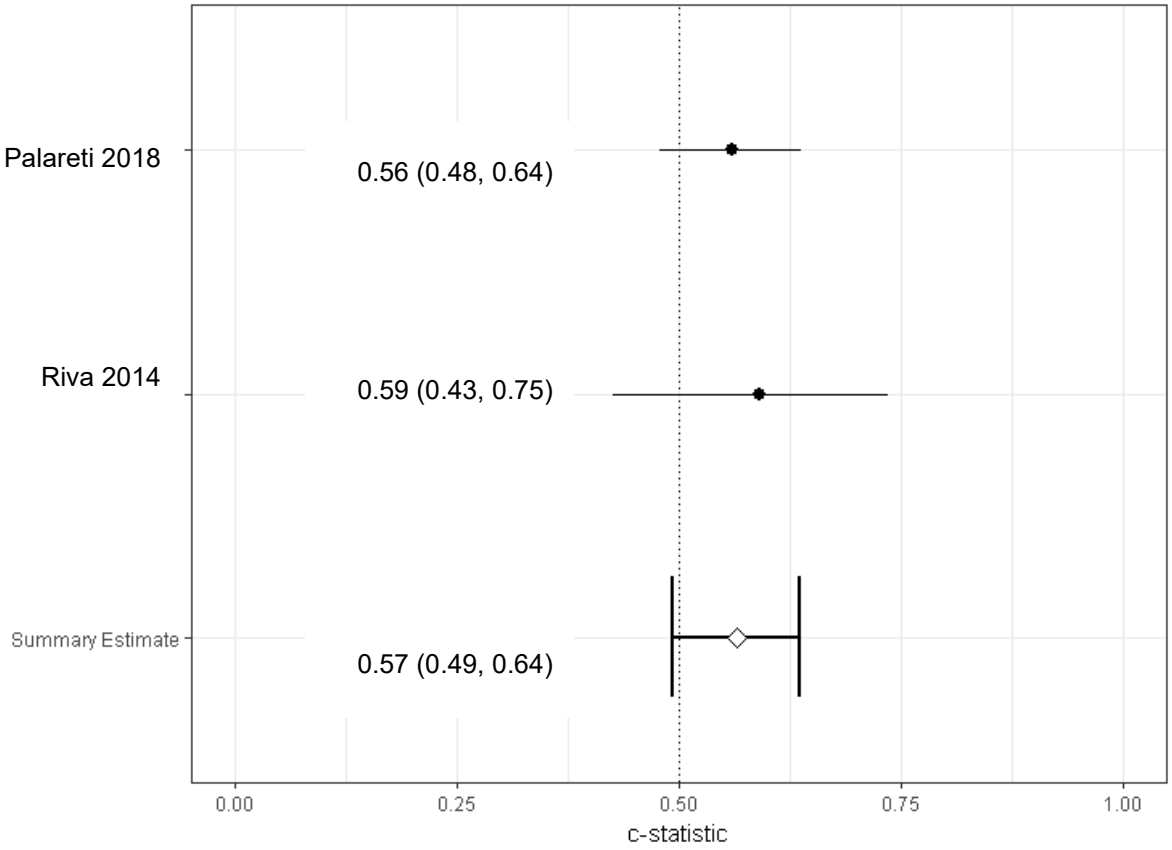
Figure 20: LR-, VTE BLEED



1 **C-statistics**

2 **ACCP**

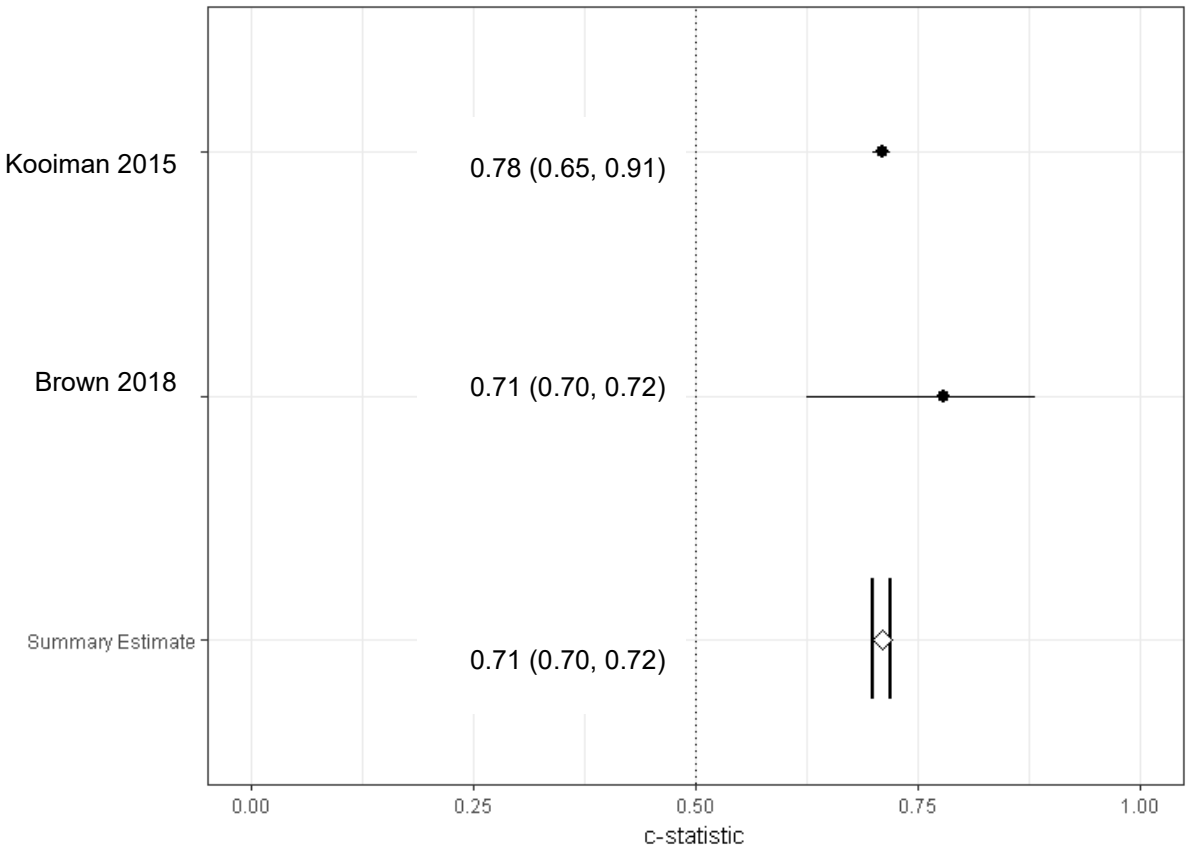
3 **Figure 21: Pooled C-statistic, ACCP**



4 FE model, I^2=0%

1 **HAS-BLED**

2 **Figure 22: Pooled C-statistic, HAS-BLED**



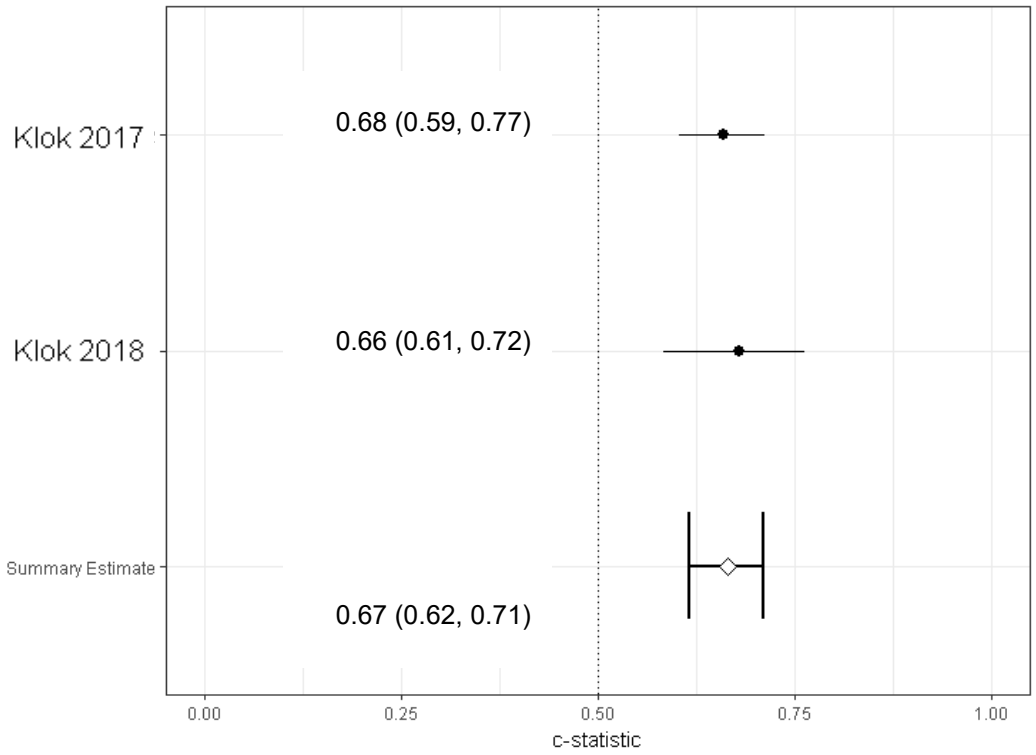
3

4 FE model, I^2=9.8%%

1 **VTE-BLEED**

2 **Figure 23: Pooled C-statistic, VTE BLEED**

3



4

5 FE model, I^2=0%

Appendix G – GRADE tables

Tools to predict VTE-recurrence

Test administered at point of discontinuation of anticoagulation treatment, ≥3 months (depending on treatment).

Sensitivity, specificity and likelihood ratios

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
HERDOO2 (follow-up 1 year post-treatment)										
1 (Rodger 2017)	Prospective management study	914	0.60 (0.43, 0.74)	0.66 (0.63, 0.69)	LR+ 1.74 (1.34, 2.27)	Serious[1]	N/A	Not serious	Serious[2]	Low
					LR- 0.61 (0.42, 0.89)	Serious[1]	N/A	Not serious	Serious[2]	Low
DAMOVES (internal validation study; follow-up median 21.3 months)										
1 (Franco 2016)	Prospective cohort study	398	0.91 (0.81, 0.96)	0.82 (0.77, 0.86)	LR+ 5.04 (3.96, 6.43)	Serious[1]	N/A	Not serious	Not serious	Moderate
					LR- 0.11 (0.05, 0.25)	Serious[1]	N/A	Not serious	Not serious	Moderate

1. Study was at moderate risk of bias.
2. 95% confidence interval for likelihood ratio crosses one end of a defined MID interval – (0.5, 1 for negative likelihood ratios and 1, 2 for positive likelihood ratios)

1 **C-statistics**

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
DASH (follow-up median 25.2 months)								
1 (Tosetto 2017)	Retrospective cohort study	827	0.65*	Not serious	N/A	Not serious	Serious[1]	Moderate
DASH (>65 years old subgroup; follow-up median 25.2 months)								
1 (Tosetto 2017)	Retrospective cohort study	315	0.54*	Not serious	N/A	Not serious	Very serious[2]	Low
DASH (≤65 years old subgroup; follow-up median 25.2 months)								
1 (Tosetto 2017)	Retrospective cohort study	512	0.72*	Not serious	N/A	Not serious	Serious[1]	Moderate
VIENNA (follow-up median 25.2 months)								
1 (Marcucci 2015)	Retrospective review of prospective cohort study	904	0.63*	Not serious	N/A	Not serious	Serious[1]	Moderate
VIENNA (≤65 years old subgroup; follow-up median 25.2 months)								
1 (Marcucci 2015)	Retrospective review of prospective cohort study	370	0.59*	Not serious	N/A	Not serious	Very serious[2]	Low
Dynamic VIENNA (measured 3 weeks after stopping anticoagulation; 68 months follow-up)								
1 (Eichinger 2014)	Prospective cohort study	553	0.63*	Not serious	N/A	Not serious	Serious[1]	Moderate
Dynamic VIENNA (measured 3 months after stopping anticoagulation; 68 months follow-up)								
1 (Eichinger 2014)	Prospective cohort study	553	0.61*	Not serious	N/A	Not serious	Serious[1]	Moderate

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Dynamic VIENNA (measured 9 months after stopping anticoagulation; 68 months follow-up)								
1 (Eichinger 2014)	Prospective cohort study	553	0.61*	Not serious	N/A	Not serious	Serious[1]	Moderate
Dynamic VIENNA (measured 15 months after stopping anticoagulation; 68 months follow-up)								
1 (Eichinger 2014)	Prospective cohort study	553	0.58*	Not serious	N/A	Not serious	Serious[1]	Moderate

* 95% confidence interval not provided
1. 95% CI not provided however sample size is over 500
2. 95% CI not provided, and sample size is under 500

Major bleeding

The test is administered either at the point of index VTE diagnosis or beginning of anticoagulation therapy. All tools are used to predict major bleeding risk apart from RIETE 2. This was designed to predict fatal bleeding, and results are presented for both major bleeding and fatal bleeding where available.

Sensitivity, specificity and likelihood ratios

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
ATRIA (High/moderate versus low risk) – 30 days										
1 (Klok 2016)	Retrospective analysis of a prospective cohort study	448	0.50 (0.27, 0.73)	0.74 (0.69, 0.77)	LR+ 1.89 (1.19, 3.02)	Serious[1]	N/A	Not serious	Serious[2]	Low
					LR- 0.68 (0.44, 1.06)	Serious[1]	N/A	Not serious	Very serious[3]	Very low
ATRIA (High versus moderate/low risk) – 30 days										
1 (Klok 2016)	Retrospective analysis of a prospective cohort study	448	0.20 (0.06, 0.44)	0.85 (0.81, 0.88)	LR+ 1.34 (0.54, 3.31)	Serious[1]	N/A	Not serious	Very serious[3]	Very low
					LR- 0.94 (0.75, 1.18)	Serious[1]	N/A	Not serious	Serious[2]	Low
HAS-BLED (High/moderate versus low risk) – 30 days										
1 (Klok 2016)	Retrospective analysis of a prospective	448	0.98 (0.74, 1.00)	0.15 (0.12, 0.18)	LR+ 1.15 (1.07, 1.23)	Serious[1]	N/A	Not serious	Not serious[4]	Moderate
					LR- 0.14 (0.01, 2.26)	Serious[1]	N/A	Not serious	Very serious[3,4]	Low

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
	cohort study									
HAS-BLED (High versus moderate/low risk) – 30 days										
1 (Klok 2016)	Retrospective analysis of a prospective cohort study	448	0.57 (0.35, 0.77)	0.37 (0.33, 0.42)	LR+ 0.90 (0.62, 1.30)	Serious[1]	N/A	Not serious	Serious[2]	Low
					LR- 1.17 (0.72, 1.89)	Serious[1]	N/A	Not serious	Serious[2]	Low
HAS-BLED (High risk [3+] versus moderate/low risk)* - up to 6 months (180 days)										
1 (Kooiman 2015)	Retrospective cohort study	537	0.55 (0.24, 0.83)	0.87 (0.84, 0.90)	LR+ 4.22 (2.35, 7.56)	Very serious[6]	N/A	Not serious	Not serious	Low
					LR- 0.52 (0.27, 1.00)	Very serious[6]	N/A	Not serious	Serious[2]	Very low
HAS-BLED (High risk [4+] versus moderate/low risk) - up to 6 months (180 days)**										
1 (Kooiman 2015)	Retrospective cohort study	537	0.18 (0.02, 0.52)	0.98 (0.96, 0.99)	LR+ 7.97 (2.02, 31.45)	Very serious[6]	N/A	Not serious	Not serious	Low
					LR- 0.84 (0.63, 1.11)	Very serious[6]	N/A	Not serious	Serious[2]	Very low
HEMORR$_2$HAGES (High/moderate versus low risk) – 30 days										
1 (Klok 2016)	Retrospective analysis of a prospective cohort study	448	0.70 (0.46, 0.88)	0.40 (0.36, 0.45)	LR+ 1.17 (0.87, 1.58)	Very serious[7]	N/A	Not serious	Serious[2]	Very low
					LR- 0.75 (0.38, 1.47)	Very serious[7]	N/A	Not serious	Very serious[3]	Very low
HEMORR$_2$HAGES (High versus moderate/low risk) – 30 days										

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Klok 2016)	Retrospective analysis of a prospective cohort study	448	0.15 (0.03, 0.38)	0.86 (0.82, 0.89)	LR+ 1.07 (0.37, 3.12)	Very serious[7]	N/A	Not serious	Very serious[3]	Very low
					LR- 0.99 (0.82, 1.19)	Very serious[7]	N/A	Not serious	Serious[2]	Very low
HEMORR$_2$HAGES (two level-optimised model) – 30 days										
1 (Klok 2016)	Retrospective analysis of a prospective cohort study	448	0.60 (0.36, 0.81)	0.67 (0.62, 0.71)	LR+ 1.81 (1.23, 2.65)	Very serious[8]	N/A	Not serious	Serious[2]	Very low
					LR- 0.60 (0.35, 1.03)	Very serious[8]	N/A	Not serious	Very serious[3]	Very low
Kearon score (High risk versus non-high risk; ≥65 year olds only)- 3 months										
1 (Scherz 2013)	Retrospective cohort study	663	0.21 (0.10, 0.40)	0.83 (0.80, 0.86)	LR+ 1.26 (0.61, 2.61)	Serious[1]	N/A	Not serious	Very serious[3]	Very low
					LR- 0.95 (0.78, 1.15)	Serious[1]	N/A	Not serious	Serious[2]	Low
Kuijer score (High risk versus moderate-low risk) – 30 days										
1 (Klok 2016)	Retrospective cohort study	448	0.20 (0.06, 0.44)	0.85 (0.81, 0.81)	LR+ 1.35 (0.54, 3.31)	Serious[1]	N/A	Not serious	Very serious[3]	Very low
					LR- 0.94 (0.75, 1.18)	Serious[1]	N/A	Not serious	Serious[2]	Low
Kuijer score (high/moderate risk versus low risk)- 30 days										
1 (Klok 2016)	Retrospective cohort study	448	0.90 (0.68, 0.99)	0.14 (0.11, 0.17)	LR+ 1.04 (0.90, 1.21)	Serious[1]	N/A	Not serious	Serious[2]	Low
					LR- 0.73 (0.19, 2.76)	Serious[1]	N/A	Not serious	Very serious[3]	Very low

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Kuijer score (High risk versus moderate/low risk) – 3 months (Figure 1, Figure 2, Figure 3 and Figure 4)										
2 (Kuijer 1998, Piovella 2014)	See above	9,497	0.21 (0.02, 0.81)	0.88 (0.73, 0.95)	LR+ 1.59 (0.46, 5.55)	Very serious[12]	Very serious[10]	Not serious	Very serious[3]	Very low
					LR- 0.81 (0.47, 1.39)	Very serious[12]	Serious[11]	Not serious	Very serious[3]	Very low
Kuijer score (high/moderate risk versus low risk)- 3 months (Figure 5, Figure 6, Figure 7 and Figure 8)										
2 (Kuijer 1998, Piovella 2014)	See above	9,497	0.97 (0.90, 0.99)	0.17 (0.10, 0.28)	LR+ 1.15 (1.10, 1.20)	Very serious[12]	Not serious	Not serious	Not serious	Low
					LR- 0.15 (0.04, 0.58)	Very serious[12]	Not serious	Not serious	Serious[2]	Very low
Kuijer score (High risk versus moderate-low risk; ≥65 year olds only) – 3 months										
1 (Scherz 2013)	Retrospective cohort study	663	0.11 (0.04, 0.27)	0.86 (0.82, 0.88)	LR+ 0.76 (0.26, 2.27)	Serious[1]	N/A	Not serious	Very serious[3]	Very low
					LR- 1.04 (0.91, 1.19)	Serious[1]	N/A	Not serious	Serious[2]	Low
Kuijer score (two level-optimised model) - 30 days										
1 (Klok 2016)	Retrospective analysis of a prospective cohort study	448	0.65 (0.41, 0.85)	0.53 (0.48, 0.58)	LR+ 1.38 (0.98, 1.93)	Very serious[8]	N/A	Not serious	Serious[2]	Very low
					LR- 0.66 (0.36, 1.21)	Very serious[8]	N/A	Not serious	Very serious[3]	Very low
MOS (Cancer patients only) – 3 months (90 days)										

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Altari 2017)	Retrospective cohort study	383	0.29 (0.08, 0.58)	0.71 (0.66, 0.76)	LR+ 0.99 (0.43, 2.31)	Serious[1]	N/A	Not serious	Very serious[3]	Very low
					LR- 1.00 (0.72, 1.40)	Serious[1]	N/A	Not serious	Serious[2]	Low
OBRI (High risk versus moderate/low risk) – 3 months										
1 (Piovella 2014)	Retrospective cohort study	8,717	0.08 (0.05, 0.14)	0.96 (0.95, 0.96)	LR+ 1.77 (1.00, 3.43)	Not serious	N/A	Not serious	Serious[2]	Moderate
					LR- 0.96 (0.91, 1.02)	Not serious	N/A	Not serious	Serious[2]	Moderate
OBRI (High/moderate risk versus low risk) – 3 months										
1 (Piovella 2014)	Retrospective cohort study	8,717	0.90 (0.82, 0.96)	0.27 (0.26, 0.28)	LR+ 1.24 (1.15, 1.33)	Not serious	N/A	Not serious	Not serious	High
					LR- 0.36 (0.19, 0.69)	Not serious	N/A	Not serious	Serious[2]	Moderate
OBRI (High risk versus moderate/low risk; ≥65 year olds only) – 3 months										
1 (Scherz 2013)	Retrospective cohort study	663	0.07 (0.02, 0.23)	0.93 (0.91, 0.95)	LR+ 1.03 (0.26, 4.04)	Serious[1]	N/A	Not serious	Very serious[3]	Very low
					LR- 1.00 (0.90, 1.11)	Serious[1]	N/A	Not serious	Serious[2]	Low
RIETE 1 (High versus moderate/low risk) – 30 days										
1 (Klok 2016)	Retrospective analysis of a prospective cohort study	448	0.10 (0.01, 0.32)	0.90 (0.87, 0.93)	LR+ 1.04 (0.27, 3.98)	Serious[1]	N/A	Not serious	Very serious[3]	Very low
					LR- 1.00 (0.86, 1.16)	Serious[1]	N/A	Not serious	Serious[2]	Low
RIETE 1 (High versus moderate/low risk) – 3 months (Figure 9, Figure 10, Figure 11 and Figure 12)										

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
2 (Ruiz-Gimenez 2008, Piovella 2014)	See above	15,289	0.15 (0.11, 0.21)	0.93 (0.87, 0.96)	LR+ 2.27 (1.62, 3.17)	Not serious	Serious[11]	Not serious	Serious[2]	Low
					LR- 0.91 (0.86, 0.96)	Not serious	Not serious	Not serious	Not serious	High
RIETE 1 (High/moderate versus low risk) – 3 months (Figure 13, Figure 14, Figure 15 and Figure 16)										
2 (Ruiz-Gimenez 2008, Piovella 2014)	See above	15,289	0.98 (0.91, 1.00)	0.13 (0.05, 0.31)	LR+ 1.15 (0.97, 1.37)	Not serious	Serious[11]	Not serious	Serious[2]	Low
					LR- 0.13 (0.01, 1.79)	Not serious	Not serious	Not serious	Very serious[3]	Low
RIETE 1 (High risk versus low-moderate; ≥65 year olds only) – 3 months										
1 (Scherz 2013)	Retrospective cohort study	663	0.14 (0.06, 0.58)	0.91 (0.89, 0.93)	LR+ 1.59 (0.62, 4.08)	Serious[1]	N/A	Not serious	Very serious[3]	Very low
					LR- 0.94 (0.81, 1.10)	Serious[1]	N/A	Not serious	Serious[2]	Low
RIETE 1 (Two level-optimised model) – 30 days										
1 (Klok 2016)	Retrospective analysis of a prospective cohort study	448	0.70 (0.46, 0.88)	0.51 (0.46, 0.56)	LR+ 1.42 (1.05, 1.92)	Very serious[8]	N/A	Not serious	Not serious	Low
					LR- 0.59 (0.30, 1.16)	Very serious[8]	N/A	Not serious	Very serious[3]	Very low
RIETE 2 (High/moderate versus low risk; predicting fatal bleeds)										
1 (Nieto 2013)	Retrospective analysis	15,206	0.81 (0.67, 0.90)	0.64 (0.63, 0.65)	LR+ 2.26 (1.98, 2.58)	Not serious	N/A	Not serious	Serious[2]	Moderate

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
	of a prospective cohort				LR- 0.30 (0.17, 0.52)	Not serious	N/A	Not serious	Serious[2]	Moderate
RIETE 2 (High versus moderate/low risk; predicting fatal bleeds)										
1 (Nieto 2013)	Retrospective analysis of a prospective cohort	15,206	0.10 (0.03, 0.21)	0.98 (0.98, 0.98)	LR+ 4.27 (1.85, 9.90)	Not serious	N/A	Not serious	Serious[2]	Moderate
					LR- 0.92 (0.85, 1.01)	Not serious	N/A	Not serious	Serious[2]	Moderate
VTE-BLEED – During in-hospital stay										
1 (Kresoja 2019)	Retrospective analysis of a prospective cohort	522	0.89 (0.65, 0.97)	0.43 (0.38, 0.47)	LR+ 1.55 (1.29, 1.86)	Serious[1]	N/A	Serious[5]	Not serious	Low
					LR- 0.26 (0.07, 0.97)	Serious[1]	N/A	Serious[5]	Serious[2]	Very low
VTE-BLEED (after 30 days of treatment)- up to 12 months (Figure 17, Figure 18, Figure 19 and Figure 20)										
2 (Klok 2017, Klok 2018)	Retrospective analysis of a prospective cohort	12,249	0.60 (0.51, 0.69)	0.69 (0.59, 0.77)	LR+ 1.96 (1.50, 2.56)	Very serious[12]	Serious[11]	Not serious	Serious[2]	Very low
					LR- 0.57 (0.45, 0.72)	Very serious[12]	Not serious	Not serious	Serious[2]	Very low
EPIPHANY index (internal validation study in cancer patients only) – 15 days										
1 (Carmona-Bayonas 2017)	Retrospective analysis of a prospective cohort study	1,075	0.94 (0.85, 0.99)	0.30 (0.27, 0.32)	LR+ 1.34 (1.24, 1.45)	Serious[1]	N/A	Not serious	Not serious	Moderate
					LR- 0.19 (0.06, 0.57)	Serious[1]	N/A	Not serious	Serious[2]	Low

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality

1. Study was at moderate risk of bias
2. 95% confidence interval for likelihood ratio crosses one end of a defined MID interval (0.5 or 1 for negative likelihood ratios and 1 or 2 for positive likelihood ratios)
3. 95% confidence interval for likelihood ratio crosses both sides of an MID (1 and either 0.5 or 2)
4. 0.5 added to TP, FP, TN and FN as previously had zero false negatives and therefore confidence intervals for negative likelihood ratio were not calculable.
5. Study was only partially applicable to the review question
6. Study was at high risk of bias
7. Data pertaining to the presence of CYP-2C9 gene was missing in this study, this was a variable of the HEMMOR$_2$HAGES tool.
8. Study was at moderate risk of bias; however it was marked down once more for this result as the cut-off used was modified based on its relationship with the outcome variable.
9. Although study was at low risk of bias overall it was marked down for this tool as it is unclear whether this study re-sampled from the same participants used in the RIETE 1 derivation study.
10. $33.3\% < I^2 < 66.6\%$
11. $I^2 > 66.6\%$
12. >33.3% of studies were at high risk of bias.

c-statistics

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
ACCP- up to 12 months (Figure 21)								
2 (Palareti 2018 and Riva 2014)	Retrospective review of prospective cohort	2,944	0.57 (0.49, 0.64)	Serious[6]	Not serious	Not serious	Very serious[4]	Very low
ACCP ((≥ 80 years old, continuous variable data)								
1 (Poli 2013)	Retrospective review of prospective cohort	1,078	0.55 (0.45, 0.64)	Very serious[3]	N/A	Not serious	Serious[2]	Very low
ATRIA- up to 12 months								
1 (Riva 2014)	Retrospective review	681	0.47 (0.31, 0.63)	Very serious[3]	N/A	Not serious	Serious[2]	Very low
ATRIA ((≥ 80 years old, continuous variables data)								
1 (Poli 2013)	Retrospective review of prospective cohort	1,078	0.58 (0.48, 0.67)	Very serious[3]	N/A	Not serious	Serious[2]	Very low
HAS-BLED – during in-hospital stay								
1 (Kresoja 2019)	Retrospective review of prospective cohort	522	0.58 (0.48, 0.69)	Serious[1]	N/A	Serious[5]	Very serious[4]	Very low
HAS-BLED – up to 6 months (Figure 22)								

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
2 (Kooiman 2015, Brown 2018)	Retrospective cohort study	132,817	0.71 (0.70, 0.72)	Serious[1]	Not serious	Not serious	Not serious	Moderate
HAS-BLED (cancer patients only) – up to 6 months								
1 (Brown 2018)	Retrospective cohort study	24,915	0.69 (0.67, 0.71)	Serious[1]	N/A	Not serious	Serious[2]	Low
HAS-BLED – up to 12 months								
1 (Riva 2014)	Retrospective cohort study	681	0.60 (0.45, 0.75)	Very serious[3]	N/A	Not serious	Very serious[4]	Very low
HAS-BLED ((≥ 80 years old, continuous variable data)								
1 (Poli 2013)	Retrospective review of prospective cohort	1,078	0.55 (0.46, 0.64)	Very serious[3]	N/A	Not serious	Serious[2]	Very low
HEMMOR$_2$HAGES – up to 12 months								
1 (Riva 2014)	Retrospective review	681	0.51 (0.32, 0.70)	Very serious[3]	N/A	Not serious	Very serious[5]	Very low
HEMMOR$_2$HAGES ((≥ 80 years old, continuous variables data)								
1 (Poli 2013)	Retrospective review of prospective cohort	1,078	0.60 (0.50, 0.70)	Very serious[3]	N/A	Not serious	Very serious[4]	Very low
Kearon score- 3 months								
1 (Zhang 2019)	Prospective review	539	0.75 (0.60, 0.89)	Serious[1]	N/A	Not serious	Very serious[4]	Very low
Kearon score (≥ 65 years old)- 3 months								

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Scherz 2013)	Retrospective review of prospective cohort	663	0.59 (0.55, 0.63)	Serious[1]	N/A	Not serious	Serious[2]	Low
Kearon score (≥ 65 years old and already received 3 months of extended duration treatment)- 3 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.54 (0.38, 0.69)	Not serious	N/A	Not serious	Serious[2]	Moderate
Kearon score (≥ 65 years old and already received 3 months of extended duration treatment)- 6 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.55 (0.43, 0.67)	Not serious	N/A	Not serious	Serious[2]	Moderate
Kearon score (≥ 65 years old and already received 3 months of extended duration treatment)- 12 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.58 (0.49, 0.67)	Not serious	N/A	Not serious	Serious[2]	Moderate
Kearon score (≥ 65 years old and already received 3 months of extended duration treatment)- 24 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.57 (0.50, 0.64)	Not serious	N/A	Not serious	Serious[2]	Moderate
Kearon score (≥ 65 years old and already received 3 months of extended duration treatment)- 36 months								
1 (Seiler 2017)	Retrospective review of	743	0.59 (0.52, 0.66)	Not serious	N/A	Not serious	Serious[2]	Moderate

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
	prospective cohort							
Kuijer score- 3 months								
1 (Zhang 2019)	Prospective review	539	0.57 (0.44, 0.68)	Serious[1]	N/A	Not serious	Very serious[4]	Very low
Kuijer score- up to 12 months								
1 (Riva 2014)	Retrospective review	681	0.51 (0.32, 0.70)	Very serious[3]	N/A	Not serious	Very serious[4]	Very low
Kuijer score (≥ 65 years old)– 3 months								
1 (Scherz 2013)	Retrospective review of prospective cohort	663	0.49 (0.45, 0.52)	Serious[1]	N/A	Not serious	Not serious	Moderate
Kuijer score (≥ 65 years old and already received 3 months of extended duration treatment) – 3 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.67 (0.54, 0.81)	Not serious	N/A	Not serious	Very serious[4]	Low
Kuijer score (≥ 65 years old and already received 3 months of extended duration treatment) – 6 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.61 (0.50, 0.72)	Not serious	N/A	Not serious	Very serious[4]	Low
Kuijer score (≥ 65 years old and already received 3 months of extended duration treatment) – 12 months								
1 (Seiler 2017)	Retrospective review of	743	0.57 (0.48, 0.66)	Not serious	N/A	Not serious	Serious[2]	Moderate

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
	prospective cohort							
Kuijer score (≥ 65 years old and already received 3 months of extended duration treatment) – 24 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.57 (0.50, 0.64)	Not serious	N/A	Not serious	Serious[2]	Moderate
Kuijer score (≥ 65 years old and already received 3 months of extended duration treatment) – 36 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.57 (0.50, 0.64)	Not serious	N/A	Not serious	Serious[2]	Moderate
Nieuwenhuis score- 3 months								
1 (Zhang 2019)	Prospective review	539	0.59 (0.41, 0.74)	Serious[1]	N/A	Not serious	Very serious[4]	Very low
OBRI – up to 12 months								
1 (Riva 2014)	Retrospective review	681	0.59 (0.42, 0.76)	Very serious[3]	N/A	Not serious	Very serious[4]	Very low
OBRI (≥ 65 years old)- 3 months								
1 (Scherz 2013)	Retrospective review of prospective cohort	663	0.54 (0.50, 0.58)	Serious[1]	N/A	Not serious	Not serious	Moderate
OBRI (≥ 65 years old and already received 3 months of extended duration treatment)- 3 months								
1 (Seiler 2017)	Retrospective review of	743	0.54 (0.43, 0.65)	Not serious	N/A	Not serious	Serious[2]	Moderate

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
	prospective cohort							
OBRI (≥ 65 years old and already received 3 months of extended duration treatment)- 6 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.51 (0.42, 0.60)	Not serious	N/A	Not serious	Serious[2]	Moderate
OBRI (≥ 65 years old and already received 3 months of extended duration treatment)- 12 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.52 (0.44, 0.60)	Not serious	N/A	Not serious	Serious[2]	Moderate
OBRI (≥ 65 years old and already received 3 months of extended duration treatment)- 24 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.52 (0.46, 0.59)	Not serious	N/A	Not serious	Not serious	High
OBRI (≥ 65 years old and already received 3 months of extended duration treatment)- 36 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.53 (0.47, 0.59)	Not serious	N/A	Not serious	Not serious	High
OBRI (≥ 80 years old, continuous variables data)								
1 (Poli 2013)	Retrospective review of prospective cohort	1,078	0.58 (0.49, 0.67)	Very serious[3]	N/A	Not serious	Serious[2]	Very low

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
RIETE 1 - 3 months								
1 (Zhang 2019)	Prospective review	539	0.56 (0.45, 0.71)	Serious[1]	N/A	Not serious	Very serious[4]	Very low
RIETE 1 (≥ 65 years old)- 3 months								
1 (Scherz 2013)	Retrospective review of prospective cohort	663	0.60 (0.56, 0.64)	Serious[1]	N/A	Not serious	Serious[2]	Low
RIETE 1 (≥ 65 years old and already received 3 months of extended duration treatment)- 3 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.59 (0.45, 0.72)	Not serious	N/A	Not serious	Very serious[4]	Low
RIETE 1 (≥ 65 years old and already received 3 months of extended duration treatment)- 6 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.59 (0.48, 0.70)	Not serious	N/A	Not serious	Very serious[4]	Low
RIETE 1 (≥ 65 years old and already received 3 months of extended duration treatment)- 12 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.63 (0.54, 0.72)	Not serious	N/A	Not serious	Very serious[4]	Low
RIETE 1 (≥ 65 years old and already received 3 months of extended duration treatment)- 24 months								
1 (Seiler 2017)	Retrospective review of	743	0.62 (0.55, 0.69)	Not serious	N/A	Not serious	Serious[2]	Moderate

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
	prospective cohort							
RIETE 1 (≥ 65 years old and already received 3 months of extended duration treatment)- 36 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.63 (0.56, 0.70)	Not serious	N/A	Not serious	Very serious[4]	Low
RIETE 1 (continuous variables) (≥ 80 years old)								
1 (Poli 2013)	Retrospective review of prospective cohort	1,078	0.61 (0.51, 0.71)	Very serious[3]	N/A	Not serious	Very serious[4]	Very low
RIETE 1- up to 12 months								
1 (Riva 2014)	Retrospective review of prospective cohort	681	0.54 (0.38, 0.71)	Very serious[3]	N/A	Not serious	Very serious[4]	Very low
RIETE 2 – 3 months								
1 (Nieto 2013)	Retrospective review of prospective cohort	15,206	0.72 (0.69, 0.75)	Not serious	N/A	Not serious	Serious[2]	Moderate
RIETE 2 (fatal bleeds only)- 3 months								
1 (Nieto 2013)	Retrospective review of prospective cohort	15,206	0.78 (0.72, 0.83)	Not serious	N/A	Not serious	Serious[2]	Moderate

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Shireman score- up to 12 months								
1 (Riva 2014)	Retrospective review	681	0.63 (0.45, 0.81)	Very serious[3]	N/A	Not serious	Very serious[5]	Very low
VTE-BLEED – during in-hospital stay								
1 (Kresoja 2019)	Retrospective review of prospective cohort	522	0.69 (0.58, 0.80)	Serious[1]	N/A	Serious[5]	Very serious[4]	Very low
VTE-BLEED (after 30 days of treatment)- up to 12 months (Figure 23)								
2 (Klok 2017 and 2018)	Retrospective review of prospective cohort	12,687	0.67 (0.62, 0.71)	Serious[6]	Not serious	Not serious	Serious[2]	Low
Seiler score (internal validation and in participants ≥ 65 years old that have already received 3 months of extended duration treatment) – 3 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.75 (0.61, 0.88)	Serious[1]	N/A	Not serious	Very serious[4]	Very low
Seiler score (internal validation and in participants ≥ 65 years old that have already received 3 months of extended duration treatment) – 6 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.69 (0.58, 0.79)	Serious[1]	N/A	Not serious	Very serious[4]	Very low
Seiler score (internal validation and in participants ≥ 65 years old and already received 3 months of extended duration treatment) – 12 months								
1 (Seiler 2017)	Retrospective review of	743	0.68 (0.60, 0.76)	Serious[1]	N/A	Not serious	Serious[2]	Low

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
	prospective cohort							
Seiler score (internal validation and in participants ≥ 65 years old and already received 3 months of extended duration treatment) – 24 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.67 (0.60, 0.73)	Serious[1]	N/A	Not serious	Serious[2]	Low
Seiler score (internal validation and in participants ≥ 65 years old and already received 3 months of extended duration treatment) – 36 months								
1 (Seiler 2017)	Retrospective review of prospective cohort	743	0.68 (0.61, 0.74)	Serious[1]	N/A	Not serious	Serious[2]	Low

* 95% confidence interval not provided or calculable
1. Study was at moderate risk of bias
2. 95% confidence interval spans two categories of test effectiveness
3. Study was at high risk of bias
4. 95% confidence interval spans three or more categories of test effectiveness
5. Study was only partially applicable to the review question.
6. >33.3% of analysis weight comes from studies moderate risk of bias.

Hazard ratios and odds ratios

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
HAS-BLED – 180 days								
1 (Brown 2018)	Retrospective cohort study	132,280	1 vs. 0: HR 1.34 (1.15, 1.56) 2 vs. 1: HR 1.31 (1.16, 1.49) 3 vs. 2: HR 1.22 (1.06, 1.39) 4+ vs. 3: HR 1.71 (1.44, 2.04)	Serious[1]	N/A	Not serious	Not serious	Moderate
1 (Kooiman 2015)	Retrospective cohort study	537	High versus moderate/low risk*: HR 8.7 (2.7, 28.4) High versus moderate/low risk**: HR 10.8 (2.3, 50.0)	Very serious[2]	N/A	Not serious	Not serious	Low
HAS-BLED (during in-hospital stay only)								
1 (Kresoja 2019)	Retrospective review of prospective cohort	522	3+ vs. 0-2: OR 1.10 (0.40, 2.90)	Serious[1]	N/A	Serious[4]	Serious[3]	Very low
HAS-BLED (cancer patients only)								
1 (Brown 2018)	Retrospective cohort study	24,915	1 vs. 0: HR 0.88 (0.67, 1.14) 2 vs. 1: HR 1.22 (0.99, 1.49) 3 vs. 2: HR 0.95 (0.77, 1.19) 4+ vs. 3: HR 1.66 (1.26, 2.20)	Serious[1]	N/A	Not serious	Not serious	Moderate
VTE-BLEED (during in-hospital stay only)								
1 (Kresoja 2019)	Retrospective review of prospective cohort	522	2+ vs. 0-1: OR 3.70 (1.10, 13.00)	Serious[1]	N/A	Serious[4]	Not serious	Low
VTE-BLEED per 1 point increase (after 30 days of treatment, adjusted for type of anticoagulant used)								

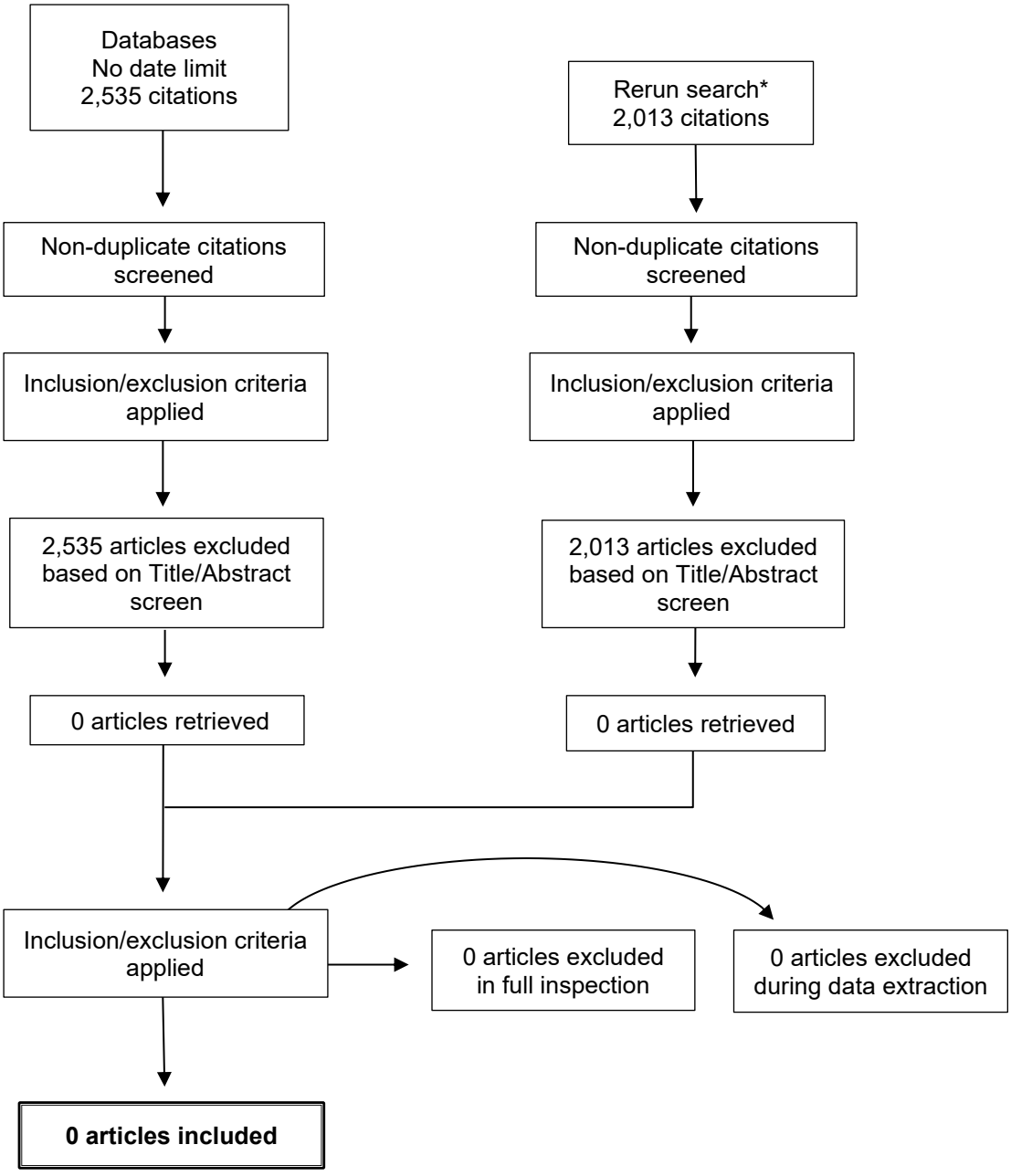
No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Klok 2018)	Retrospective review of prospective cohort	4,447	OR 1.40 (1.10,1.60)	Very serious[2]	N/A	Not serious	Not serious	Low
VTE-BLEED per 1 point increase (after 30 days of treatment, adjusted for type of anticoagulant used; unprovoked VTE patients only)								
1 (Klok 2018)	Retrospective review of prospective cohort	2,856	OR 1.30 (1.00,1.70)	Very serious[2]	N/A	Not serious	Not serious	Low
VTE-BLEED (after 30 days of treatment, adjusted for type of anticoagulant used)								
1 (Klok 2018)	Retrospective review of prospective cohort	4,447	VTE-BLEED ≥2 vs.<2: HR 2.30 (1.10,4.50)	Very serious[2]	N/A	Not serious	Serious[5]	Very low
VTE-BLEED (after 30 days of treatment, adjusted for type of anticoagulant used; unprovoked VTE patients only)								
1 (Klok 2018)	Retrospective review of prospective cohort	2,856	VTE-BLEED ≥2 vs.<2: HR 1.70 (0.69, 4.40)	Very serious[2]	N/A	Not serious	Very serious[6]	Very low

*High risk classified as 3+ points. This is in-line with the original model, designed for use in atrial fibrillation.

**High risk classified as 4+ points. This adjusted the original HAS-BLED model such that the cut off is in-line with the definition of high risk in VTE (7.3% chance of major bleed)

1. Study was at moderate risk of bias
2. Study was at high risk of bias
3. 95% confidence interval crosses the line of no effect
4. Study was only partially applicable to the review question.
5. 95% CI crosses one side of a MID boundaries (0.8, 1.25)
6. 95% CI crosses both MID boundaries (0.8, 1.25)

Appendix H – Economic evidence study selection



**Combined for all questions in the guideline*

Appendix I – Excluded studies

Clinical studies (main search)

Short Title	Title	Reason for exclusion
Ahn (2013)	Validation of the clinical prediction rule for recurrent venous thromboembolism in cancer patients: the Ottawa score	• Participants were on anticoagulants during follow up VTE-recurrence prediction
Ahn (2017)	Prognostic Value of Treatment Setting in Patients with Cancer Having Pulmonary Embolism: Comparison with the Pulmonary Embolism Severity Index	• Study does not contain any relevant prognostic variables
Ahn (2018)	Validation of the EPIPHANY index for predicting risk of serious complications in cancer patients with incidental pulmonary embolism	• Study does not contain any relevant prognostic variables
Airaksinen (2010)	Usefulness of outpatient bleeding risk index to predict bleeding complications in patients with long-term oral anticoagulation undergoing coronary stenting	• Does not contain a population of people meeting the PICO
Alatri (2017)	Low discriminating power of the modified Ottawa risk score in a cohort of patients with active cancer and acute venous thromboembolism	• Conference abstract
Albertsen (2018)	Risk of Recurrent Venous Thromboembolism: A Danish Nationwide Cohort Study	• Study did not include prognostic risk tool
Al-Ogaili (2018)	Risk assessment as a guide for the prevention of cancer-associated thromboembolism	• Review article but not a systematic review
Alper (2018)	Risk Stratification Model: Lower-Extremity Ultrasonography for Hospitalized Patients with Suspected Deep Vein Thrombosis	• Study does not contain any relevant prognostic variables • Does not contain a population of people meeting the PICO
Angelini (2016)	A novel risk assessment model to predict venous thromboembolism (VTE) in cancer inpatients: the canclot score	• predicted outcome(s) not of relevance to this review
Astruc (2016)	External validation of the modified Ottawa score for risk stratification of recurrent cancer-associated thrombosis	• Conference abstract

Short Title	Title	Reason for exclusion
Aujesky (2005)	Derivation and validation of a prognostic model for pulmonary embolism	• predicted outcome(s) not of relevance to this review
Aujesky (2006)	Validation of a model to predict adverse outcomes in patients with pulmonary embolism	• predicted outcome(s) not of relevance to this review
Baglin (2004)	High risk of recurrent venous thromboembolism in men	• Participants were on anticoagulants during follow up
Banerjee (2014)	Composite risk scores and composite endpoints in the risk prediction of outcomes in anticoagulated patients with atrial fibrillation: The Loire Valley Atrial Fibrillation Project	• Does not contain a population of people meeting the PICO
Barco (2017)	Clinical course of patients with symptomatic isolated superficial vein thrombosis: the ICARO follow-up study	• Does not contain a population of people meeting the PICO
Barillari (2016)	Recurrence of venous thromboembolism in patients with recent gestational deep vein thrombosis or pulmonary embolism: Findings from the RIETE Registry	• individual risk factors only
Bateman (2017)	Correlation of the Caprini Score and Venous Thromboembolism Incidence Following Primary Total Joint Arthroplasty-Results of a Single-Institution Protocol	• Does not contain a population of people meeting the PICO
Bernaitis (2017)	A High HASBLED Score Identifies Poor Warfarin Control in Patients Treated for Non-Valvular Atrial Fibrillation in Australia and Singapore	• Does not contain a population of people meeting the PICO
Bertoletti (2011)	Prognostic value of the Geneva prediction rule in patients in whom pulmonary embolism is ruled out	• Does not contain a population of people meeting the PICO
Bertoletti (2013)	Prognostic value of the Geneva prediction rule in patients with pulmonary embolism	• Does not contain a population of people meeting the PICO
Bledsoe (2016)	Intermountain risk score strati pulmonary embolism severity index fies to better predict mortality across all classes	• Conference abstract
Bledsoe (2016)	Intermountain risk score is highly predictive of pulmonary embolism mortality	• Conference abstract

Short Title	Title	Reason for exclusion
Blondon (2018)	Comparative Performance of Clinical Risk Assessment Models for Hospital-Acquired Venous Thromboembolism in Medical Patients	• Does not contain a population of people meeting the PICO
Carpenter (2009)	Differentiating low-risk and no-risk PE patients: the PERC score	• Participants sample size less than 250
Chagnon (2002)	Comparison of two clinical prediction rules and implicit assessment among patients with suspected pulmonary embolism	• Does not contain a population of people meeting the PICO
Chai-Adisaksopha (2018)	Vitamin K Antagonists After 6 Months of Low-Molecular-Weight Heparin in Cancer Patients with Venous Thromboembolism	• Study did not include prognostic risk tool
Chatterjee (2016)	HAS-BLED Versus ATRIA Risk Scores for Intracranial Hemorrhage in Patients Receiving Thrombolytics for Pulmonary Embolism	• Conference abstract
Chee (2014)	Predictors of venous thromboembolism recurrence and bleeding among active cancer patients: a population-based cohort study	• predicted outcome(s) not of relevance to this review
Coleman (2015)	Validation of the multivariable In-hospital Mortality for PulmonAry embolism using Claims daTa (IMPACT) prediction rule within an all-payer inpatient administrative claims database	• predicted outcome(s) not of relevance to this review
Coleman (2016)	External validation of a multivariable claims-based rule for predicting in-hospital mortality and 30-day post-pulmonary embolism complications	• predicted outcome(s) not of relevance to this review
Dang (2018)	Predicting venous thromboembolism following laparoscopic bariatric surgery: development of the BariClot tool using the MBSAQIP database	• Does not contain a population of people meeting the PICO
de Bastos (2016)	Derivation of a risk assessment model for hospital-acquired venous thrombosis: the NAVAL score	• Does not contain a population of people meeting the PICO

Short Title	Title	Reason for exclusion
Decousus (2011)	Factors at admission associated with bleeding risk in medical patients: findings from the IMPROVE investigators	• Does not contain a population of people meeting the PICO
Dentali (2015)	Rate and duration of hospitalization for deep vein thrombosis and pulmonary embolism in real-world clinical practice	• predicted outcome(s) not of relevance to this review
Desai (2016)	Utility of Inferior Vena Cava Filters in Severe Pulmonary Embolism, Catheter-directed Therapy in Massive and Submassive Pulmonary Embolism, and HAS-BLED Score to Determine Risk of Major Hemorrhage in Pulmonary Embolism	• Review article but not a systematic review
Development and validation... (Janssen)	Development and validation of clinical prediction models: Marginal differences between logistic regression, penalized maximum likelihood estimation, and genetic programming	• No specific tool investigated the study explores the methods of validating
Donze (2008)	Prospective validation of the Pulmonary Embolism Severity Index. A clinical prognostic model for pulmonary embolism	• predicted outcome(s) not of relevance to this review
Eichinger (2010)	Risk assessment of recurrence in patients with unprovoked deep vein thrombosis or pulmonary embolism: the Vienna prediction model	• Secondary paper of an included study
Elias (2016)	Prognostic models in acute pulmonary embolism: a systematic review and meta-analysis	• Systematic review
Elias (2017)	Automating Venous Thromboembolism Risk Calculation Using Electronic Health Record Data upon Hospital Admission: The Automated Padua Prediction Score	• Does not contain a population of people meeting the PICO
Engelberger (2011)	Comparison of the diagnostic performance of the original and modified Wells score in inpatients and outpatients with suspected deep vein thrombosis	• Does not contain a population of people meeting the PICO
Ensor (2013)	Protocol for a systematic review of prognostic models for the recurrence of venous	• Systematic review

Short Title	Title	Reason for exclusion
	thromboembolism (VTE) following treatment for a first unprovoked VTE	
Ensor (2016)	Systematic review of prognostic models for recurrent venous thromboembolism (VTE) post-treatment of first unprovoked VTE	• Systematic review
Erkens (2012)	Does the Pulmonary Embolism Severity Index accurately identify low risk patients eligible for outpatient treatment?	• Participants sample size less than 250
Ferrer (2013)	Validation of two clinical prognostic models in patients with acute symptomatic pulmonary embolism	• predicted outcome(s) not of relevance to this review
Franco (2017)	Predicting recurrence after a first unprovoked venous thromboembolism: Retrospective validation of the DAMOVES score	• Duplicate reference
Galanaud (2012)	Superficial vein thrombosis and recurrent venous thromboembolism: a pooled analysis of two observational studies	• Individual risk factor and not risk tool assessed
Galanaud (2017)	Long-term risk of venous thromboembolism recurrence after isolated superficial vein thrombosis	• Does not contain a population of people meeting the PICO
Grant (2016)	Assessing the Caprini Score for Risk Assessment of Venous Thromboembolism in Hospitalized Medical Patients	• Does not contain a population of people meeting the PICO
Greene (2016)	Validation of Risk Assessment Models of Venous Thromboembolism in Hospitalized Medical Patients	• Does not contain a population of people meeting the PICO
Gruettner (2015)	Importance of Wells score and Geneva score for the evaluation of patients suspected of pulmonary embolism	• Diagnostic study and not predicting recurrence or major bleeding
Heit (2000)	Predictors of recurrence after deep vein thrombosis and pulmonary embolism: a population-based cohort study	• Study did not include prognostic risk tool
Hendriksen (2015)	Diagnostic prediction models for suspected pulmonary embolism: systematic review and independent external validation in primary care	• Systematic review

Short Title	Title	Reason for exclusion
Hippisley-Cox (2014)	QBleed predicted risk for gastrointestinal and intracranial bleeding	• Does not contain a population of people meeting the PICO
Hippisley-Cox (2014)	The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study	• Does not contain a population of people meeting the PICO
Hostler (2016)	Validation of the International Medical Prevention Registry on Venous Thromboembolism Bleeding Risk Score	• Does not contain a population of people meeting the PICO
Hron (2006)	Prediction of recurrent venous thromboembolism by the activated partial thromboplastin time	• individual risk factors only
Janjua (2008)	Treatment of acute pulmonary embolism as outpatients or following early discharge - A systematic review	• Systematic review
Janssen (2012)	Development and validation of clinical prediction models: marginal differences between logistic regression, penalized maximum likelihood estimation, and genetic programming	• Review article but not a systematic review
Jaquet (2018)	Prediction of short-term prognosis in elderly patients with acute pulmonary embolism: validation of the RIETE score	• predicted outcome(s) not of relevance to this review
Jimenez (2007)	Prognostic models for selecting patients with acute pulmonary embolism for initial outpatient therapy	• predicted outcome(s) not of relevance to this review
Jimenez (2010)	Simplification of the pulmonary embolism severity index for prognostication in patients with acute symptomatic pulmonary embolism	• predicted outcome(s) not of relevance to this review
Kilic (2014)	Prognostic role of simplified Pulmonary Embolism Severity Index and the European Society of Cardiology Prognostic Model in short- and long-term risk stratification in pulmonary embolism	• Participants sample size less than 250
Kim (2017)	Utility of the simplified Wells and revised Geneva scores to exclude pulmonary embolism in femur fracture patients	• predicted outcome(s) not of relevance to this review

Short Title	Title	Reason for exclusion
Kline (2018)	Utility of a Clinical Prediction Rule to Exclude Pulmonary Embolism Among Low-Risk Emergency Department Patients: Reason to PERC Up	• Review article but not a systematic review
Klok (2008)	Comparison of the revised Geneva score with the Wells rule for assessing clinical probability of pulmonary embolism	• Does not contain a population of people meeting the PICO
Kohn (2016)	External validation of the In-hospital Mortality for PulmonAry embolism using Claims daTa (IMPACT) multivariable prediction rule	• predicted outcome(s) not of relevance to this review
Kruger (2017)	HERDOO2 identified women at low risk for recurrence after 5 to 12 mo of anticoagulation for a first unprovoked VTE	• Conference abstract
Kyrle (2012)	Clinical scores to predict recurrence risk of venous thromboembolism	• Review article but not a systematic review
Kyrle (2013)	Predicting the risk of recurrent venous thromboembolism. The Austrian study on recurrent venous thromboembolism (AUREC)	• Review article but not a systematic review
Lankeit (2012)	A strategy combining imaging and laboratory biomarkers in comparison with a simplified clinical score for risk stratification of patients with acute pulmonary embolism	• predicted outcome(s) not of relevance to this review
Lauber (2018)	Predictors and Outcomes of Recurrent Venous Thromboembolism in Elderly Patients	• Individual risk factor and not risk tool assessed
Lauque (2014)	Predictive value of the heart-type fatty acid-binding protein and the Pulmonary Embolism Severity Index in patients with acute pulmonary embolism in the emergency department	• Participants sample size less than 250
Le Moigne (2014)	Validation of the LEFt score, a newly proposed diagnostic tool for deep vein thrombosis in pregnant women	• Does not contain a population of people meeting the PICO
Li (2018)	Stratification of venous thromboembolism risk in burn patients by Caprini score	• Does not contain a population of people meeting the PICO
Lin (2014)	Incremental health care resource utilization and economic burden of venous	• Study did not include prognostic risk tool

Short Title	Title	Reason for exclusion
	thromboembolism recurrence from a U.S. payer perspective	
Long-term risk of venous... (2016)	Long-term risk of venous thrombosis after stopping anticoagulants for a first unprovoked event: a multi-national cohort	• Individual risk factor and not risk tool assessed
Louzada (2012)	Development of a clinical prediction rule for risk stratification of recurrent venous thromboembolism in patients with cancer-associated venous thromboembolism	• Individual risk factor and not risk tool assessed
Madsen (2010)	PERC bleeding risk calculation and resultant test threshold may be inappropriate	• predicted outcome(s) not of relevance to this review
Maestre (2015)	Identification of Low-Risk Patients with Acute Symptomatic Pulmonary Embolism for Outpatient Therapy	• Individual risk factor and not risk tool assessed
Mansfield (2016)	Predictors of active cancer thromboembolic outcomes: validation of the Khorana score among patients with lung cancer	• Does not contain a population of people meeting the PICO
Mizuno (2015)	Pulmonary embolism severity index and simplified pulmonary embolism severity index risk scores are useful to predict mortality in Japanese patients with pulmonary embolism	• predicted outcome(s) not of relevance to this review
Molnar (2018)	Risk and complications of venous thromboembolism in dialysis patients	• Study was not investigating risk tools but a hopsital protocol
Moorehead (2017)	A Retrospective Cohort Analysis of Pharmacologic VTE Prophylaxis and Padua Prediction Score in Hospitalized Patients With Chronic Liver Disease	• Does not contain a population of people meeting the PICO
Moores (2013)	Changes in PESI scores predict mortality in intermediate-risk patients with acute pulmonary embolism	• predicted outcome(s) not of relevance to this review
Mueller (2017)	HAS-BLED Predicts Warfarin Control in Australian Patients treated for Deep Vein Thrombosis	• predicted outcome(s) not of relevance to this review
Munoz (2018)	Multivariable clinical-genetic risk model for predicting venous thromboembolic events in patients with cancer	• Participants sample size less than 250 only 71 participants had VTE

Short Title	Title	Reason for exclusion
Murthy (2018)	Ability of the Khorana score to predict recurrent thromboembolism in cancer patients with ischemic stroke	• Does not contain a population of people meeting the PICO
Nendaz (2004)	Validation of a risk score identifying patients with acute pulmonary embolism, who are at low risk of clinical adverse outcome	• Participants sample size less than 250
Nendaz (2014)	Multicentre validation of the Geneva Risk Score for hospitalised medical patients at risk of venous thromboembolism. Explicit ASsessment of Thromboembolic RIsk and Prophylaxis for Medical PATients in SwitzErland (ESTIMATE)	• Does not contain a population of people meeting the PICO
Nieto (2006)	Acute venous thromboembolism in patients with recent major bleeding. The influence of the site of bleeding and the time elapsed on outcome	• No specific tool investigated the study explores the methods of validating
Nieto (2010)	Fatal bleeding in patients receiving anticoagulant therapy for venous thromboembolism: findings from the RIETE registry	• Secondary paper of an included study
Ozsu (2014)	Combination and comparison of two models in prognosis of pulmonary embolism: results from TUrkey Pulmonary Embolism Group (TUPEG) study	• predicted outcome(s) not of relevance to this review
Paiva (2013)	Cardiovascular risk assessment of pulmonary embolism with the GRACE risk score	• Participants sample size less than 250
Pannucci (2012)	Assessment of postoperative venous thromboembolism risk in plastic surgery patients using the 2005 and 2010 Caprini Risk score	• Comparison of two tools
Parker (2018)	Risk stratification for the development of venous thromboembolism in hospitalized patients with cancer	• Does not contain a population of people meeting the PICO
Penaloza (2011)	Comparison of the Wells score with the simplified revised Geneva score for assessing	• Comparison of two tools

Short Title	Title	Reason for exclusion
	pretest probability of pulmonary embolism	
Penaloza (2017)	Pulmonary embolism rule-out criteria (PERC) rule in European patients with low implicit clinical probability (PERCEPIC): a multicentre, prospective, observational study	• Not a peer-reviewed publication
Piovella (2014)	Comparison of four scores to predict major bleeding in patients receiving anticoagulation for venous thromboembolism: findings from the RIETE registry	• Duplicate reference
Poli (2013)	Assessing recurrence risk following acute venous thromboembolism: use of algorithms	• Letter to the editor
Poli (2013)	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	• Duplicate reference
Prandoni (2010)	Major bleeding as a predictor of mortality in patients with venous thromboembolism: Findings from the RIETE Registry	• Letter to the editor
Rief (2018)	The HAS-BLED score is useful for the identification of venous thromboembolism patients with high risk for major bleeding complications: A prospective outpatient cohort study	• Study not reported in English
Rief (2018)	Calculation of HAS-BLED Score Is Useful for Early Identification of Venous Thromboembolism Patients at High Risk for Major Bleeding Events: A Prospective Outpatients Cohort Study	• Study not reported in English
Rieken (2018)	Risk of Bleeding Versus Venous Thromboembolism in Urological Surgery-Finding the Right Balance Is Not Always Easy!	• Letter to the editor
Robert-Ebadi (2010)	Differences in clinical presentation of pulmonary embolism in women and men	• Individual risk factor and not risk tool assessed

Short Title	Title	Reason for exclusion
Rodger (2008)	Identifying unprovoked thromboembolism patients at low risk for recurrence who can discontinue anticoagulant therapy.	• Does not contain a population of people meeting the PICO
Rodger (2016)	Long-term risk of venous thrombosis after stopping anticoagulants for a first unprovoked event: A multi-national cohort	• Study did not include prognostic risk tool
Rosenberg (2014)	External validation of the risk assessment model of the International Medical Prevention Registry on Venous Thromboembolism (IMPROVE) for medical patients in a tertiary health system	• Does not contain a population of people meeting the PICO
Rosenberg (2016)	External validation of the IMPROVE Bleeding Risk Assessment Model in medical patients	• Does not contain a population of people meeting the PICO
Sam (2011)	The shock index and the simplified PESI for identification of low-risk patients with acute pulmonary embolism	• Comparison of two tools
Sanden (2017)	Venous thromboembolism and cancer risk	• Study did not include prognostic risk tool
Shen (2016)	Comparison of the Wells score with the revised Geneva score for assessing suspected pulmonary embolism: a systematic review and meta-analysis	• Systematic review
Smith (2012)	Triple antithrombotic therapy following an acute coronary syndrome: Prevalence, outcomes and prognostic utility of the HAS-BLED score	• Does not contain a population of people meeting the PICO
Spirk (2011)	Cardiac troponin testing and the simplified Pulmonary Embolism Severity Index. The SWIss Venous ThromboEmbolism Registry (SWIVTER)	• Individual risk factor and not risk tool assessed
Squizzato (2012)	Prognostic clinical prediction rules to identify a low-risk pulmonary embolism: a systematic review and meta-analysis	• Systematic review
Stoker (2016)	Managing risk after intracerebral hemorrhage in concomitant atrial fibrillation	• Review article but not a systematic review

Short Title	Title	Reason for exclusion
	and cerebral amyloid angiopathy	
Suarez (2008)	Predictive variables for major bleeding events in patients presenting with documented acute venous thromboembolism. Findings from the RIETE Registry	• Duplicate reference
Tan (2013)	Validation of the San Francisco Syncope Rule in two hospital emergency departments in an Asian population	• Does not contain a population of people meeting the PICO
Thomas (2014)	Bleeding risk prediction models in atrial fibrillation topical collection on invasive electrophysiology and pacing	• Review article but not a systematic review
Tosetto (2012)	Predicting disease recurrence in patients with previous unprovoked venous thromboembolism: a proposed prediction score (DASH)	• Primary study
Touhami (2018)	Are the Wells Score and the Revised Geneva Score valuable for the diagnosis of pulmonary embolism in pregnancy?	• Participants sample size less than 250
Tritschler (2015)	Predicting recurrence after unprovoked venous thromboembolism: prospective validation of the updated Vienna Prediction Model	• Participants sample size less than 250
Trujillo-Santos (2015)	A prognostic score to identify low-risk outpatients with acute deep vein thrombosis in the lower limbs	• Study did not include prognostic risk tool
Tsu (2015)	Modified HAS-BLED score and risk of major bleeding in patients receiving dabigatran and rivaroxaban: A retrospective, case-control study	• Participants sample size less than 250
van Es (2017)	Bleeding risk in patients with unprovoked venous thromboembolism: A critical appraisal of clinical prediction scores	• Systematic review
van Es (2017)	Comparison of risk prediction scores for venous thromboembolism in cancer patients: a prospective cohort study	• Comparison of two tools

Short Title	Title	Reason for exclusion
Vink (2003)	Individualized duration of oral anticoagulant therapy for deep vein thrombosis based on a decision model	• Study did not include prognostic risk tool
Vuilleumier (2015)	Cardiac biomarkers and clinical scores for risk stratification in elderly patients with non-high-risk pulmonary embolism	• Study did not include prognostic risk tool
Wells (2003)	The outpatient bleeding risk index: validation of a tool for predicting bleeding rates in patients treated for deep venous thrombosis and pulmonary embolism	• Participants sample size less than 250
Wells (2016)	Prediction of bleeding risk in patients on extended oral anticoagulation for venous thromboembolism	• Conference abstract
Wicki (2000)	Predicting adverse outcome in patients with acute pulmonary embolism: a risk score	• Comparison of two tools
Zhou (2012)	The prognostic value of pulmonary embolism severity index in acute pulmonary embolism: A meta-analysis	• Systematic review
Zhou (2012)	Validation of the Caprini risk assessment model in Chinese hospitalized patients with venous thromboembolism	• Participants sample size less than 250

Clinical studies (search update)

Short Title	Title	Reason for exclusion
Alatri (2017	Low discriminating power of the modified Ottawa risk score in a cohort of patients with active cancer and acute venous thromboembolism	• Abstract only
Dhanda (2018)	Low discriminating power of the modified Ottawa risk score in a cohort of patients with active cancer and acute venous thromboembolism	• Outcome not of interest
Jaquet (2018)	Low discriminating power of the modified Ottawa risk score in a cohort of patients with active cancer and acute venous thromboembolism	• Outcome not of interest
Palareti (2018)	The American College of Chest Physician score to assess the risk of bleeding during anticoagulation in patients with venous thromboembolism	• Reference already contained in main review
Wells (2019)	The value of sPESI for risk stratification in patients with pulmonary embolism	• Outcome not of interest

Appendix J – References

Included studies

Alatri A, Mazzolai L, Kucher N, Aujesky D, Beer J H, Baldi T, Banyai M, Hayoz D, Kaeslin T, Korte W, Escher R, Husmann M, Frauchiger B, Engelberger R P, Baumgartner I, and Spirk D (2017) The Modified Ottawa Score and Clinical Events in Hospitalized Patients with Cancer-Associated Thrombosis from the Swiss VTE Registry. Seminars in Thrombosis and Hemostasis 43(8), 871-876

Brown J D, Goodin A J, Lip G Y. H, and Adams V R (2018) Risk Stratification for Bleeding Complications in Patients With Venous Thromboembolism: Application of the HAS-BLED Bleeding Score During the First 6 Months of Anticoagulant Treatment. Journal of the American Heart Association 7(6), 07

Carmona-Bayonas A, Jimenez-Fonseca P, Font C, Fenoy F, Otero R, Beato C, Plasencia J M, Biosca M, Sanchez M, Benegas M, Calvo-Temprano D, Varona D, Faez L, de la Haba , I , Antonio M, Madridano O, Solis M P, Ramchandani A, Castanon E, Marchena P J, Martin M, Ayala de la Pena, F , and Vicente V (2017) Predicting serious complications in patients with cancer and pulmonary embolism using decision tree modelling: the EPIPHANY Index. British Journal of Cancer 116(8), 994-1001

Eichinger S, Heinze G, and Kyrle P A (2014) D-dimer levels over time and the risk of recurrent venous thromboembolism: an update of the Vienna prediction model. Journal of the American Heart Association 3(1), e000467

Franco Moreno, A I, Garcia Navarro, M J, Ortiz Sanchez, J , Martin Diaz, R M, Madronal Cerezo, E , de Ancos Aracil, C L, Cabello Clotet, N , Perales Fraile, I , Gimeno Garcia, S , Montero Hernandez, C , Zapatero Gaviria, A , Ruiz Giardin, and J M (2016) A risk score for prediction of recurrence in patients with unprovoked venous thromboembolism (DAMOVES). European Journal of Internal Medicine 29, 59-64

Kresoja, K. P., Ebner, M., Rogge, N. I. J. et al. (2019) Prediction and prognostic importance of in-hospital major bleeding in a real-world cohort of patients with pulmonary embolism. International Journal of Cardiology 16: 16

Klok F A, Niemann C, Dellas C, Hasenfus G, Konstantinides S, and Lankeit M (2016) Performance of five different bleeding-prediction scores in patients with acute pulmonary embolism. Journal of Thrombosis & Thrombolysis 41(2), 312-20

Klok F A, Barco S, and Konstantinides S V (2017) External validation of the VTE-BLEED score for predicting major bleeding in stable anticoagulated patients with venous thromboembolism. Thrombosis & Haemostasis 117(6), 1164-1170

Klok F A, Barco S, Turpie A G. G, Haas S, Kreutz R, Mantovani L G, Gebel M, Herpers M, Bugge J P, Kostantinides S V, and Ageno W (2018) Predictive value of venous thromboembolism (VTE)-BLEED to predict major bleeding and other adverse events in a practice-based cohort of patients with VTE: results of the XALIA study. British Journal of Haematology 19, 19

Kooiman J, van Hagen , N , Iglesias Del Sol, A , Planken E V, Lip G Y, van der Meer , F J, Cannegieter S C, Klok F A, and Huisman M V (2015) The HAS-BLED Score Identifies Patients with Acute Venous Thromboembolism at High Risk of Major Bleeding Complications

during the First Six Months of Anticoagulant Treatment. PLoS ONE [Electronic Resource] 10(4), e0122520

Kuijer P M, Hutten B A, Prins M H, and Buller H R (1999) Prediction of the risk of bleeding during anticoagulant treatment for venous thromboembolism. Archives of Internal Medicine 159(5), 457-60

Marcucci M, Iorio A, Douketis J D, Eichinger S, Tosetto A, Baglin T, Cushman M, Palareti G, Poli D, Tait R C, and Kyrle P A (2015) Risk of recurrence after a first unprovoked venous thromboembolism: external validation of the Vienna Prediction Model with pooled individual patient data. Journal of Thrombosis & Haemostasis 13(5), 775-81

Nieto J A, Solano R, Trapero Iglesias, N , Ruiz-Gimenez N, Fernandez-Capitan C, Valero B, Tiberio G, Bura-Riviere A, Monreal M, and Investigators Riete (2013) Validation of a score for predicting fatal bleeding in patients receiving anticoagulation for venous thromboembolism. Thrombosis Research 132(2), 175-9

Palareti G, Antonucci E, Mastroiacovo D, Ageno W, Pengo V, Poli D, Testa S, Tosetto A, and Prandoni P (2018) The American College of Chest Physician score to assess the risk of bleeding during anticoagulation in patients with venous thromboembolism. Journal of Thrombosis & Haemostasis 16(10), 1994-2002

Piovella, C., Dalla Valle, F., Trujillo-Santos, J., Pesavento, R., López, L., Font, L., ... & RIETE investigators. (2014). Comparison of four scores to predict major bleeding in patients receiving anticoagulation for venous thromboembolism: findings from the RIETE registry. *Internal and emergency medicine*, *9*(8), 847-852.

Poli D, Antonucci E, Cosmi B, Palareti G, Testa S, Paoletti O, Nante G, Pengo V, Carini U, Guazzaloca G, Scortechini A R, Canafoglia L, Tomassetti S, Restifo D, Ciampa A, Pignatelli P, Basili S, Saliola M, Di Gennaro , L , De Cristofaro , R , Ageno W, Caprioli M, Pedrini S, Orlandini F, Benedetti R, Ruocco L, Tiraferri E, Cappelli R, Piana A, Armani U, Porcu A, Falco P, Da Col , P , Marongiu F, Barcellona D, Falanga A, Lerede T, Galbo L, Bucherini E, Insana A, Masciocco L, and Pini F (2013) 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 11(6), 1053-1058

Riva N, Bellesini M, Di Minno , M N, Mumoli N, Pomero F, Franchini M, Fantoni C, Lupoli R, Brondi B, Borretta V, Bonfanti C, Ageno W, and Dentali F (2014) Poor predictive value of contemporary bleeding risk scores during long-term treatment of venous thromboembolism. A multicentre retrospective cohort study. Thrombosis & Haemostasis 112(3), 511-21

Rodger M A, Le Gal , G , Anderson D R, Schmidt J, Pernod G, Kahn S R, Righini M, Mismetti P, Kearon C, Meyer G, Elias A, Ramsay T, Ortel T L, Huisman M V, Kovacs M J, and Investigators Reverse Ii Study (2017) Validating the HERDOO2 rule to guide treatment duration for women with unprovoked venous thrombosis: multinational prospective cohort management study. BMJ 356, j1065

Ruiz-Gimenez N, Suarez C, Gonzalez R, Nieto J A, Todoli J A, Samperiz A L, Monreal M, and Investigators Riete (2008) Predictive variables for major bleeding events in patients presenting with documented acute venous thromboembolism. Findings from the RIETE Registry. Thrombosis & Haemostasis 100(1), 26-31

Scherz N, Mean M, Limacher A, Righini M, Jaeger K, Beer H J, Frauchiger B, Osterwalder J, Kucher N, Matter C M, Banyai M, Angelillo-Scherrer A, Lammle B, Husmann M, Egloff M, Aschwanden M, Bounameaux H, Cornuz J, Rodondi N, and Aujesky D (2013) Prospective,

multicenter validation of prediction scores for major bleeding in elderly patients with venous thromboembolism. Journal of Thrombosis & Haemostasis 11(3), 435-43

Seiler E, Limacher A, Mean M, Beer H J, Osterwalder J, Frauchiger B, Righini M, Aschwanden M, Matter C M, Banyai M, Kucher N, Staub D, Lammle B, Rodondi N, Squizzato A, and Aujesky D (2017) Derivation and validation of a novel bleeding risk score for elderly patients with venous thromboembolism on extended anticoagulation. Thrombosis & Haemostasis 117(10), 1930-1936

Tosetto A, Testa S, Martinelli I, Poli D, Cosmi B, Lodigiani C, Ageno W, De Stefano , V , Falanga A, Nichele I, Paoletti O, Bucciarelli P, Antonucci E, Legnani C, Banfi E, Dentali F, Bartolomei F, Barcella L, and Palareti G (2017) External validation of the DASH prediction rule: a retrospective cohort study. Journal of Thrombosis & Haemostasis 15(10), 1963-1970

Zhang, Z., Lei, J., Zhai, Z. et al. (2019) Comparison of prediction value of four bleeding risk scores for pulmonary embolism with anticoagulation: A real-world study in Chinese patients. The clinical respiratory journal 13(3): 139-147

Excluded studies (main search)

Ahn S, Lim K S, Lee Y S, and Lee J L. (2013). Validation of the clinical prediction rule for recurrent venous thromboembolism in cancer patients: the Ottawa score. Supportive Care in Cancer, 21(8), pp.2309-13.

Ahn S, Lee Y S, Kim W Y, Lim K S, and Lee J L. (2017). Prognostic Value of Treatment Setting in Patients with Cancer Having Pulmonary Embolism: Comparison with the Pulmonary Embolism Severity Index. Clinical and Applied Thrombosis/Hemostasis, 23(6), pp.615-621.

Ahn S, Cooksley T, Banala S, Buffardi L, and Rice T W. (2018). Validation of the EPIPHANY index for predicting risk of serious complications in cancer patients with incidental pulmonary embolism. Supportive Care in Cancer, 26(10), pp.3601-3607.

Airaksinen K E, Suurmunne H, Porela P, Niemela M, Vikman S, Puurunen M, Annala A P, Biancari F, and Karjalainen P P. (2010). Usefulness of outpatient bleeding risk index to predict bleeding complications in patients with long-term oral anticoagulation undergoing coronary stenting. American Journal of Cardiology, 106(2), pp.175-9.

Alatri A, Mazzolai L, Font C, Tafur A, Valle R, Marchena P J, Ballaz A, Tiraferri E, Font L, and Monreal M. (2017). Low discriminating power of the modified Ottawa risk score in a cohort of patients with active cancer and acute venous thromboembolism. Vasa - European Journal of Vascular Medicine, 46 (Supplement 98), pp.20-21.

Albertsen I E, Nielsen P B, Sogaard M, Goldhaber S Z, Overvad T F, Rasmussen L H, and Larsen T B. (2018). Risk of Recurrent Venous Thromboembolism: A Danish Nationwide Cohort Study. American Journal of Medicine, 131(9), pp.1067-1074.e4.

Al-Ogaili A, Fuentes H E, Tafur A J, and Caprini J. (2018). Risk assessment as a guide for the prevention of cancer-associated thromboembolism. International Angiology, 37(4), pp.269-276.

Alper E C, Ip I K, Balthazar P, Piazza G, Goldhaber S Z, Benson C B, Lacson R, and Khorasani R. (2018). Risk Stratification Model: Lower-Extremity Ultrasonography for Hospitalized Patients with Suspected Deep Vein Thrombosis. Journal of General Internal Medicine, 33(1), pp.21-25.

Angelini De, Greene Mt, Wietzke Jn, Flanders Sa, and Sood Sl. (2016). A novel risk assessment model to predict venous thromboembolism (VTE) in cancer inpatients: the canclot score. Blood. Conference: 58th annual meeting of the american society of hematology, and ASH 2016. United states. Conference start: 20161203. Conference end: 20161206, 128(22) (no pagination), pp..

Astruc N, Ianotto J C, Metges J P, Lacut K, and Delluc A. (2016). External validation of the modified Ottawa score for risk stratification of recurrent cancer-associated thrombosis. European Journal of Internal Medicine, 36, pp.e11-e12.

Aujesky D, Obrosky D S, Stone R A, Auble T E, Perrier A, Cornuz J, Roy P M, and Fine M J. (2005). Derivation and validation of a prognostic model for pulmonary embolism. American Journal of Respiratory & Critical Care Medicine, 172(8), pp.1041-6.

Aujesky D, Roy P M, Le Manach , C P, Verschuren F, Meyer G, Obrosky D S, Stone R A, Cornuz J, and Fine M J. (2006). Validation of a model to predict adverse outcomes in patients with pulmonary embolism. European Heart Journal, 27(4), pp.476-81.

Baglin T, Luddington R, Brown K, and Baglin C. (2004). High risk of recurrent venous thromboembolism in men. Journal of Thrombosis & Haemostasis, 2(12), pp.2152-5.

Banerjee A, Fauchier L, Bernard-Brunet A, Clementy N, and Lip G Y. H. (2014). 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, 111(3), pp.549-556.

Barco S, Pomero F, Di Minno , M N D, Tamborini Permunian, E , Malato A, Pasca S, Barillari G, Fenoglio L, Siragusa S, Di Minno , G , Ageno W, and Dentali F. (2017). Clinical course of patients with symptomatic isolated superficial vein thrombosis: the ICARO follow-up study. Journal of Thrombosis & Haemostasis, 15(11), pp.2176-2183.

Barillari G, Londero A P, Brenner B, Nauffal D, Munoz-Torrero J F. S, Del Molino , F , Moustafa F, Madridano O, Martin-Martos F, Monreal M, Decousus H, Prandoni P, Barba R, Di Micco , P , Bertoletti L, Schellong S, Tzoran I, Reis A, Bosevski M, Bounameaux H, Maly R, Wells P, and Papadakis M. (2016). Recurrence of venous thromboembolism in patients with recent gestational deep vein thrombosis or pulmonary embolism: Findings from the RIETE Registry. European Journal of Internal Medicine, 32, pp.53-59.

Bateman D K, Dow R W, Brzezinski A, Bar-Eli H Y, and Kayiaros S T. (2017). Correlation of the Caprini Score and Venous Thromboembolism Incidence Following Primary Total Joint Arthroplasty-Results of a Single-Institution Protocol. Journal of Arthroplasty, 32(12), pp.3735-3741.

Bernaitis N, Ching C K, Chen L, Hon J S, Teo S C, Badrick T, Davey A K, and Anoopkumar-Dukie S. (2017). A High HASBLED Score Identifies Poor Warfarin Control in Patients Treated for Non-Valvular Atrial Fibrillation in Australia and Singapore. Basic & Clinical Pharmacology & Toxicology, 121(6), pp.499-504.

Bertoletti L, Le Gal , G , Aujesky D, Roy P M, Sanchez O, Verschuren F, Bounameaux H, Perrier A, and Righini M. (2011). Prognostic value of the Geneva prediction rule in patients in whom pulmonary embolism is ruled out. Journal of Internal Medicine, 269(4), pp.433-40.

Bertoletti L, Le Gal , G , Aujesky D, Sanchez O, Roy P M, Verschuren F, Bounameaux H, Perrier A, and Righini M. (2013). Prognostic value of the Geneva prediction rule in patients with pulmonary embolism. Thrombosis Research, 132(1), pp.32-6.

Bledsoe J, Aston V, and Horne B. (2016). Intermountain risk score strati pulmonary embolism severity index fies to better predict mortality across all classes. Annals of emergency medicine. Conference: american college of emergency physicians, and ACEP 2016 research forum. United states, 68(4 Supplement 1), pp.S54-s55.

Bledsoe J, Aston V, and Horne B. (2016). Intermountain risk score is highly predictive of pulmonary embolism mortality. Annals of emergency medicine. Conference: american college of emergency physicians, and ACEP 2016 research forum. United states, 68(4 Supplement 1), pp.S55-s56.

Blondon M, Spirk D, Kucher N, Aujesky D, Hayoz D, Beer J H, Husmann M, Frauchiger B, Korte W, Wuillemin W A, Bounameaux H, Righini M, and Nendaz M. (2018). Comparative Performance of Clinical Risk Assessment Models for Hospital-Acquired Venous Thromboembolism in Medical Patients. Thrombosis & Haemostasis, 118(1), pp.82-89.

Carpenter C R, Keim S M, Seupaul R A, Pines J M, Best Evidence in Emergency Medicine Investigator, and Group . (2009). Differentiating low-risk and no-risk PE patients: the PERC score. Journal of Emergency Medicine, 36(3), pp.317-22.

Chagnon I, Bounameaux H, Aujesky D, Roy P M, Gourdier A L, Cornuz J, Perneger T, and Perrier A. (2002). Comparison of two clinical prediction rules and implicit assessment among patients with suspected pulmonary embolism. American Journal of Medicine, 113(4), pp.269-75.

Chai-Adisaksopha C, Iorio A, Crowther M A, Aibar M A, Arcelus J I, Ballaz A, Barba R, Barron M, Barron-Andres B, Bascunana J, Blanco-Molina A, Bueso T, Calvo B, Canada G, Canas I, Casado I, Culla A, de Miguel , J , del Toro , J , Diaz-Peromingo J A, Falga C, Fernandez-Capitan C, Font C, Font L, Gallego P, Garcia-Bragado F, Gomez V, Gonzalez J, Grau E, Guil M, Guirado L, Gutierrez J, Hernandez G, Hernandez-Blasco L, Isern V, Jara-Palomares L, Jaras M J, Jimenez D, Lacruz B, Lecumberri R, Lobo J L, Lopez-Jimenez L, Lopez-Reyes R, Lopez-Saez J B, Lorente M A, Lorenzo A, Madridano O, Maestre A, Manrique-Abos I, Marchena P J, Martin-Antoran J M, Martin-Martos F, Monreal M, Morales M V, Morillo R, Nauffal D, Nieto J A, Nieto S, Nunez M J, Odriozola M, Otalora S, Otero R, Pagan B, Pedrajas J M, Perez C, Peris M L, Pons I, Porras J A, Ramirez L, Riera A, Rivas A, Rodriguez C, Rodriguez-Davila M A, Rosa V, Ruiz-Gimenez N, Samperiz A, Sanchez R, Sala M C, Sahuquillo J C, Sanz O, Soler S, Suarez-Gonzalez I, Surinach J M, Tiberio G, Tolosa C, Trujillo-Santos J, Uresandi F, Valero B, Valle R, Vela J, Vicente M P, Vidal G, Vilella-Tomas V, Villalta J, Malfante P C, Vanassche T, Verhamme P, Wells P, Hirmerova J, Maly R, Tomko T, Celis G, Salgado E, Sanchez G T, Bertoletti L, Bura-Riviere A, Farge-Bancel D, Hij A, Mahe I, Merah A, Quere I, Papadakis M, Braester A, Brenner B, Tzoran I, Apollonio A, Barillari G, Bertone A, Bilora F, Bucherini E, Candelero G, Ciammaichella M, Di Micco , P , Ferrazzi P, Grandone E, Lessiani G, Lodigiani C, Mastroiacovo D, Pace F, Pesavento R, Pinelli M, Prandoni P, Rosa M, Rota L, Tiraferri E, Tonello D, Tufano A, Venturelli U, Visona A, Zalunardo B, Drucka E, Kigitovica D, Skride A, Sousa M S, Bosevski M, Zdraveska M, Bounameaux H, Mazzolai L, and Serrano J C. (2018). Vitamin K Antagonists After 6 Months of Low-Molecular-Weight Heparin in Cancer Patients with Venous Thromboembolism. American Journal of Medicine, 131(4), pp.430-437.

Chatterjee S, Lip G Y, and Giri J. (2016). HAS-BLED Versus ATRIA Risk Scores for Intracranial Hemorrhage in Patients Receiving Thrombolytics for Pulmonary Embolism. Journal of the American College of Cardiology, 67(24), pp.2904-5.

Chee C E, Ashrani A A, Marks R S, Petterson T M, Bailey K R, Melton L J, 3rd , and Heit J A. (2014). Predictors of venous thromboembolism recurrence and bleeding among active cancer patients: a population-based cohort study. Blood, 123(25), pp.3972-8.

Coleman C I, Kohn C G, Crivera C, Schein J R, and Peacock W F. (2015). Validation of the multivariable In-hospital Mortality for PulmonAry embolism using Claims daTa (IMPACT) prediction rule within an all-payer inpatient administrative claims database. BMJ Open, 5(10), pp.e009251.

Coleman C I, Peacock W F, Fermann G J, Crivera C, Weeda E R, Hull M, DuCharme M, Becker L, and Schein J R. (2016). External validation of a multivariable claims-based rule for predicting in-hospital mortality and 30-day post-pulmonary embolism complications. BMC Health Services Research, 16(1), pp.610.

Dang J T, Switzer N, Delisle M, Laffin M, Gill R, Birch D W, and Karmali S. (2018). Predicting venous thromboembolism following laparoscopic bariatric surgery: development of the BariClot tool using the MBSAQIP database. Surgical Endoscopy, 12, pp.12.

de Bastos , M , Barreto S M, Caiafa J S, Boguchi T, Silva J L, and Rezende S M. (2016). Derivation of a risk assessment model for hospital-acquired venous thrombosis: the NAVAL score. Journal of Thrombosis & Thrombolysis, 41(4), pp.628-35.

Decousus H, Tapson V F, Bergmann J F, Chong B H, Froehlich J B, Kakkar A K, Merli G J, Monreal M, Nakamura M, Pavanello R, Pini M, Piovella F, Spencer F A, Spyropoulos A C, Turpie A G, Zotz R B, Fitzgerald G, Anderson F A, and Investigators Improve. (2011). Factors at admission associated with bleeding risk in medical patients: findings from the IMPROVE investigators. Chest, 139(1), pp.69-79.

Dentali F, Di Micco , G , Giorgi Pierfranceschi, M , Gussoni G, Barillari G, Amitrano M, Fontanella A, Lodigiani C, Guida A, Visona A, Monreal M, Di Micco , and P . (2015). Rate and duration of hospitalization for deep vein thrombosis and pulmonary embolism in real-world clinical practice. Annals of Medicine, 47(7), pp.546-54.

Desai A K, Pang J, Aparnath M, and Ilowite J. (2016). Utility of Inferior Vena Cava Filters in Severe Pulmonary Embolism, Catheter-directed Therapy in Massive and Submassive Pulmonary Embolism, and HAS-BLED Score to Determine Risk of Major Hemorrhage in Pulmonary Embolism. American Journal of Respiratory & Critical Care Medicine, 193(11), pp.1301-3.

Janssen K J. M, Siccama I, Vergouwe Y, Koffijberg E, Debray T P. A, Keijzer M, Grobbee D E, and Moons K G. M. ( ). Development and validation of clinical prediction models: Marginal differences between logistic regression, penalized maximum likelihood estimation, and genetic programming. Journal of Clinical Epidemiology., , pp..

Donze J, Le Gal , G , Fine M J, Roy P M, Sanchez O, Verschuren F, Cornuz J, Meyer G, Perrier A, Righini M, and Aujesky D. (2008). Prospective validation of the Pulmonary Embolism Severity Index. A clinical prognostic model for pulmonary embolism. Thrombosis & Haemostasis, 100(5), pp.943-8.

Eichinger S, Heinze G, Jandeck L M, and Kyrle P A. (2010). Risk assessment of recurrence in patients with unprovoked deep vein thrombosis or pulmonary embolism: the Vienna prediction model. Circulation, 121(14), pp.1630-6.

Elias A, Mallett S, Daoud-Elias M, Poggi J N, and Clarke M. (2016). Prognostic models in acute pulmonary embolism: a systematic review and meta-analysis. BMJ Open, 6(4), pp.e010324.

Elias P, Khanna R, Dudley A, Davies J, Jacolbia R, McArthur K, and Auerbach A D. (2017). Automating Venous Thromboembolism Risk Calculation Using Electronic Health Record

Data upon Hospital Admission: The Automated Padua Prediction Score. Journal of Hospital Medicine (Online), 12(4), pp.231-237.

Engelberger R P, Aujesky D, Calanca L, Staeger P, Hugli O, and Mazzolai L. (2011). Comparison of the diagnostic performance of the original and modified Wells score in inpatients and outpatients with suspected deep vein thrombosis. Thrombosis Research, 127(6), pp.535-9.

Ensor J, Riley R D, Moore D, Bayliss S, Jowett S, and Fitzmaurice D A. (2013). Protocol for a systematic review of prognostic models for the recurrence of venous thromboembolism (VTE) following treatment for a first unprovoked VTE. Systematic Reviews, 2, pp.91.

Ensor J, Riley R D, Moore D, Snell K I, Bayliss S, and Fitzmaurice D. (2016). Systematic review of prognostic models for recurrent venous thromboembolism (VTE) post-treatment of first unprovoked VTE. BMJ Open, 6(5), pp.e011190.

Erkens P M, Gandara E, Wells P S, Shen A Y, Bose G, Le Gal , G , Rodger M, Prins M H, and Carrier M. (2012). Does the Pulmonary Embolism Severity Index accurately identify low risk patients eligible for outpatient treatment?. Thrombosis Research, 129(6), pp.710-4.

Ferrer M, Morillo R, Elias T, Jara L, Garcia L, Nieto R, Sandoval E, Uresandi F, Otero R, and Jimenez D. (2013). Validation of two clinical prognostic models in patients with acute symptomatic pulmonary embolism. Archivos de Bronconeumologia, 49(10), pp.427-31.

Franco Moreno, A I, Garcia Navarro, M J, Ortiz Sanchez, J , Ruiz Giardin, and J M. (2017). Predicting recurrence after a first unprovoked venous thromboembolism: Retrospective validation of the DAMOVES score. European Journal of Internal Medicine, 41, pp.e15-e16.

Galanaud J P, Bosson J L, Genty C, Presles E, Cucherat M, Sevestre M A, Quere I, Decousus H, and Leizorovicz A. (2012). Superficial vein thrombosis and recurrent venous thromboembolism: a pooled analysis of two observational studies. Journal of Thrombosis & Haemostasis, 10(6), pp.1004-11.

Galanaud J P, Sevestre M A, Pernod G, Kahn S R, Genty C, Terrisse H, Brisot D, Gillet J L, Quere I, and Bosson J L. (2017). Long-term risk of venous thromboembolism recurrence after isolated superficial vein thrombosis. Journal of Thrombosis & Haemostasis, 15(6), pp.1123-1131.

Grant P J, Greene M T, Chopra V, Bernstein S J, Hofer T P, and Flanders S A. (2016). Assessing the Caprini Score for Risk Assessment of Venous Thromboembolism in Hospitalized Medical Patients. American Journal of Medicine, 129(5), pp.528-535.

Greene M T, Spyropoulos A C, Chopra V, Grant P J, Kaatz S, Bernstein S J, and Flanders S A. (2016). Validation of Risk Assessment Models of Venous Thromboembolism in Hospitalized Medical Patients. American Journal of Medicine, 129(9), pp.1001.e9-1001.e18.

Gruettner J, Walter T, Lang S, Meyer M, Apfaltrer P, Henzler T, and Viergutz T. (2015). Importance of Wells score and Geneva score for the evaluation of patients suspected of pulmonary embolism. In Vivo, 29(2), pp.269-72.

Heit J A, Mohr D N, Silverstein M D, Petterson T M, O'Fallon W M, Melton L J, and 3rd . (2000). Predictors of recurrence after deep vein thrombosis and pulmonary embolism: a population-based cohort study. Archives of Internal Medicine, 160(6), pp.761-8.

Hendriksen J M, Geersing G J, Lucassen W A, Erkens P M, Stoffers H E, van Weert , H C, Buller H R, Hoes A W, and Moons K G. (2015). Diagnostic prediction models for suspected

pulmonary embolism: systematic review and independent external validation in primary care. BMJ, 351, pp.h4438.

Hippisley-Cox, J., & Coupland, C. (2014). Predicting risk of upper gastrointestinal bleed and intracranial bleed with anticoagulants: cohort study to derive and validate the QBleed scores. *Bmj*, *349*, g4606.

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

Hostler D C, Marx E S, Moores L K, Petteys S, Hostler J M, Mitchell J D, Holley P R, Collen J F, Foster B, and Holley A B. (2015). Validation of the IMPROVE Bleeding Risk Score. Chest, 23, pp.23.

Hostler D C, Marx E S, Moores L K, Petteys S K, Hostler J M, Mitchell J D, Holley P R, Collen J F, Foster B E, and Holley A B. (2016). Validation of the International Medical Prevention Registry on Venous Thromboembolism Bleeding Risk Score. Chest, 149(2), pp.372-9.

Hron G, Eichinger S, Weltermann A, Quehenberger P, Halbmayer W M, and Kyrle P A. (2006). Prediction of recurrent venous thromboembolism by the activated partial thromboplastin time. Journal of Thrombosis and Haemostasis, 4(4), pp.752-756.

Janjua M, Badshah A, Matta F, Danescu L G, Yaekoub A Y, and Stein P D. (2008). Treatment of acute pulmonary embolism as outpatients or following early discharge - A systematic review. Thrombosis and Haemostasis, 100(5), pp.756-761.

Janssen K J, Siccama I, Vergouwe Y, Koffijberg H, Debray T P, Keijzer M, Grobbee D E, and Moons K G. (2012). Development and validation of clinical prediction models: marginal differences between logistic regression, penalized maximum likelihood estimation, and genetic programming. Journal of Clinical Epidemiology, 65(4), pp.404-12.

Jaquet E, Tritschler T, Stalder O, Limacher A, Mean M, Rodondi N, and Aujesky D. (2018). Prediction of short-term prognosis in elderly patients with acute pulmonary embolism: validation of the RIETE score. Journal of Thrombosis & Haemostasis, 16(7), pp.1313-1320.

Jimenez D, Yusen R D, Otero R, Uresandi F, Nauffal D, Laserna E, Conget F, Oribe M, Cabezudo M A, and Diaz G. (2007). Prognostic models for selecting patients with acute pulmonary embolism for initial outpatient therapy. Chest, 132(1), pp.24-30.

Jimenez D, Aujesky D, Moores L, Gomez V, Lobo J L, Uresandi F, Otero R, Monreal M, Muriel A, Yusen R D, and Investigators Riete. (2010). Simplification of the pulmonary embolism severity index for prognostication in patients with acute symptomatic pulmonary embolism. Archives of Internal Medicine, 170(15), pp.1383-9.

Kilic T, Gunen H, Gulbas G, Hacievliyagil S S, and Ozer A. (2014). Prognostic role of simplified Pulmonary Embolism Severity Index and the European Society of Cardiology Prognostic Model in short- and long-term risk stratification in pulmonary embolism. Pakistan Journal of Medical Sciences, 30(6), pp.1259-64.

Kim Y J, Choi D H, Lee E S, Ryoo S M, Ahn S, Sohn C H, Seo D W, Lim K S, and Kim W Y. (2017). Utility of the simplified Wells and revised Geneva scores to exclude pulmonary embolism in femur fracture patients. American Journal of Emergency Medicine, 35(8), pp.1131-1135.

Kline J A. (2018). Utility of a Clinical Prediction Rule to Exclude Pulmonary Embolism Among Low-Risk Emergency Department Patients: Reason to PERC Up. JAMA, 319(6), pp.551-553.

Klok F A, Kruisman E, Spaan J, Nijkeuter M, Righini M, Aujesky D, Roy P M, Perrier A, Le Gal , G , and Huisman M V. (2008). Comparison of the revised Geneva score with the Wells rule for assessing clinical probability of pulmonary embolism. Journal of Thrombosis & Haemostasis, 6(1), pp.40-4.

Kohn C G, Peacock W F, Fermann G J, Bunz T J, Crivera C, Schein J R, and Coleman C I. (2016). External validation of the In-hospital Mortality for PulmonAry embolism using Claims daTa (IMPACT) multivariable prediction rule. International Journal of Clinical Practice, 70(1), pp.82-8.

Kruger P, and Eikelboom J. (2017). HERDOO2 identified women at low risk for recurrence after 5 to 12 mo of anticoagulation for a first unprovoked VTE. Annals of Internal Medicine, 167(6), pp.JC33.

Kyrle P A, and Eichinger S. (2012). Clinical scores to predict recurrence risk of venous thromboembolism. Thrombosis & Haemostasis, 108(6), pp.1061-4.

Kyrle P A, and Eischer L. (2013). Predicting the risk of recurrent venous thromboembolism. The Austrian study on recurrent venous thromboembolism (AUREC). Hamostaseologie, 33(3), pp.201-9.

Lankeit M, Gomez V, Wagner C, Aujesky D, Recio M, Briongos S, Moores C L. K, Yusen R D, Konstantinides S, Jimenez D, Instituto Ramon y Cajal de Investigacion Sanitaria Pulmonary Embol, and Group . (2012). A strategy combining imaging and laboratory biomarkers in comparison with a simplified clinical score for risk stratification of patients with acute pulmonary embolism. Chest, 141(4), pp.916-922.

Lauber S, Limacher A, Tritschler T, Stalder O, Mean M, Righini M, Aschwanden M, Beer J H, Frauchiger B, Osterwalder J, Kucher N, Lammle B, Cornuz J, Angelillo-Scherrer A, Matter C M, Husmann M, Banyai M, Staub D, Mazzolai L, Hugli O, Rodondi N, and Aujesky D. (2018). Predictors and Outcomes of Recurrent Venous Thromboembolism in Elderly Patients. American Journal of Medicine, 131(6), pp.703.e7-703.e16.

Lauque D, Maupas-Schwalm F, Bounes V, Juchet H, Bongard V, Roshdy A, Botella J M, and Charpentier S. (2014). Predictive value of the heart-type fatty acid-binding protein and the Pulmonary Embolism Severity Index in patients with acute pulmonary embolism in the emergency department. Academic Emergency Medicine, 21(10), pp.1143-50.

Le Moigne , E , Genty C, Meunier J, Arnoult A C, Righini M, Bressollette L, Bosson J L, Le Gal , G , and Investigators Optimev. (2014). Validation of the LEFt score, a newly proposed diagnostic tool for deep vein thrombosis in pregnant women. Thrombosis Research, 134(3), pp.664-7.

Li Q, Ba T, Wang L F, Chen Q, Li F, and Xue Y. (2018). Stratification of venous thromboembolism risk in burn patients by Caprini score. Burns, 22, pp.22.

Lin J, Lingohr-Smith M, and Kwong W J. (2014). Incremental health care resource utilization and economic burden of venous thromboembolism recurrence from a U.S. payer perspective. Journal of Managed Care Pharmacy, 20(2), pp.174-86.

. (2016). Long-term risk of venous thrombosis after stopping anticoagulants for a first unprovoked event: a multi-national cohort. Thrombosis research. 143 (pp 152-158), and 2016. Date of publication: 18 feb 2016., , pp..

Louzada M L, Carrier M, Lazo-Langner A, Dao V, Kovacs M J, Ramsay T O, Rodger M A, Zhang J, Lee A Y, Meyer G, and Wells P S. (2012). Development of a clinical prediction rule for risk stratification of recurrent venous thromboembolism in patients with cancer-associated venous thromboembolism. Circulation, 126(4), pp.448-54.

Madsen T E. (2010). PERC bleeding risk calculation and resultant test threshold may be inappropriate. Annals of Emergency Medicine, 56(5), pp.585-6; author reply 586-7.

Maestre A, Trujillo-Santos J, Riera-Mestre A, Jimenez D, Di Micco , P , Bascunana J, Vela J R, Peris L, Malfante P C, Monreal M, and Investigators Riete. (2015). Identification of Low-Risk Patients with Acute Symptomatic Pulmonary Embolism for Outpatient Therapy. Annals of the American Thoracic Society, 12(8), pp.1122-9.

Mansfield A S, Tafur A J, Wang C E, Kourelis T V, Wysokinska E M, and Yang P. (2016). Predictors of active cancer thromboembolic outcomes: validation of the Khorana score among patients with lung cancer. Journal of Thrombosis and Haemostasis, 14(9), pp.1773-1778.

Mizuno A, Yamamoto T, Tanabe Y, Obayashi T, Takayama M, Nagao K, and Tokyo C C. U. Network Scientific Committee. (2015). Pulmonary embolism severity index and simplified pulmonary embolism severity index risk scores are useful to predict mortality in Japanese patients with pulmonary embolism. Circulation Journal, 79(4), pp.889-91.

Molnar A O, Bota S E, McArthur E, Lam N N, Garg A X, Wald R, Zimmerman D, and Sood M M. (2018). Risk and complications of venous thromboembolism in dialysis patients. Nephrology Dialysis Transplantation, 33(5), pp.874-880.

Moorehead K J, Jeffres M N, and Mueller S W. (2017). A Retrospective Cohort Analysis of Pharmacologic VTE Prophylaxis and Padua Prediction Score in Hospitalized Patients With Chronic Liver Disease. Journal of Pharmacy Practice, 30(1), pp.58-63.

Moores L, Zamarro C, Gomez V, Aujesky D, Garcia L, Nieto R, Yusen R, Jimenez D, Instituto Ramon y Cajal de Investigacion Sanitaria Pulmonary Embol, and Group . (2013). Changes in PESI scores predict mortality in intermediate-risk patients with acute pulmonary embolism. European Respiratory Journal, 41(2), pp.354-9.

Mueller K, Bernaitis N, Badrick T, and Anoopkumar-Dukie S. (2017). HAS-BLED Predicts Warfarin Control in Australian Patients treated for Deep Vein Thrombosis. Basic & Clinical Pharmacology & Toxicology, 120(3), pp.299-302.

Munoz Martin, A J, Ortega I, Font C, Pachon V, Castellon V, Martinez-Marin V, Salgado M, Martinez E, Calzas J, Ruperez A, Souto J C, Martin M, Salas E, and Soria J M. (2018). Multivariable clinical-genetic risk model for predicting venous thromboembolic events in patients with cancer. British Journal of Cancer, 118(8), pp.1056-1061.

Murthy S B, Cushman M, Bobrow D, Kamel H, Merkler A E, Elkind M S. V, DeAngelis L M, and Navi B B. (2018). Ability of the Khorana score to predict recurrent thromboembolism in cancer patients with ischemic stroke. Journal of Clinical Neuroscience, 57, pp.111-115.

Nendaz M R, Bandelier P, Aujesky D, Cornuz J, Roy P M, Bounameaux H, and Perrier A. (2004). Validation of a risk score identifying patients with acute pulmonary embolism, who are at low risk of clinical adverse outcome. Thrombosis & Haemostasis, 91(6), pp.1232-6.

Nendaz M, Spirk D, Kucher N, Aujesky D, Hayoz D, Beer J H, Husmann M, Frauchiger B, Korte W, Wuillemin W A, Jager K, Righini M, and Bounameaux H. (2014). Multicentre validation of the Geneva Risk Score for hospitalised medical patients at risk of venous

thromboembolism. Explicit ASsessment of Thromboembolic RIsk and Prophylaxis for Medical PATients in SwitzErland (ESTIMATE). Thrombosis & Haemostasis, 111(3), pp.531-8.

Nieto J A, Bruscas M J, Ruiz-Ribo D, Trujillo-Santos J, Valle R, Monreal M, Arcelus J I, Barba R, Barron M, Blanco A, Buges J, Cabezudo M A, Casado I, Conget F, De las Heras, G , Falga C, Fernandez-Capitan C, Font L, Gallego P, Garcia-Bragado F, Grau E, Guijarro R, Guil M, Gutierrez J, Hernandez L, Jimenez D, Laserna E, Lobo J L, Lopez F, Lopez L, Lopez I, Maestre A, Martin J J, Naufall M D, Otero R, Pedrajas J M, Portillo J, Rabunal R, Raguer E, Raventos A, Roman R, Romero C, Rosa V, Ruiz-Gimenez N, Samperiz A L, Sanchez R, Sanchez J F, Soler S, Tiberio G, Tirado R, Tolosa C, Trujillo J, Uresandi F, Vasco B, Llobet X, Mismetti P, Rivron-Guillot K, and Bottaro F. (2006). Acute venous thromboembolism in patients with recent major bleeding. The influence of the site of bleeding and the time elapsed on outcome. Journal of Thrombosis and Haemostasis, 4(11), pp.2367-2372.

Nieto J A, Solano R, Ruiz-Ribo M D, Ruiz-Gimenez N, Prandoni P, Kearon C, Monreal M, and Riete Investigators. (2010). Fatal bleeding in patients receiving anticoagulant therapy for venous thromboembolism: findings from the RIETE registry. Journal of Thrombosis & Haemostasis, 8(6), pp.1216-22.

Ozsu S, Ozlu T, Senturk A, Ucar E Y, Kirkil G, Kadioglu E E, Altinsoy B, Saylan B, Selimoglu H S, Dabak G, Tutar N, Uysal A, and Investigators Tupeg Study. (2014). Combination and comparison of two models in prognosis of pulmonary embolism: results from TUrkey Pulmonary Embolism Group (TUPEG) study. Thrombosis Research, 133(6), pp.1006-10.

Paiva L V, Providencia R C, Barra S N, Faustino A C, Botelho A M, and Marques A L. (2013). Cardiovascular risk assessment of pulmonary embolism with the GRACE risk score. American Journal of Cardiology, 111(3), pp.425-31.

Pannucci C J, Barta R J, Portschy P R, Dreszer G, Hoxworth R E, Kalliainen L K, and Wilkins E G. (2012). Assessment of postoperative venous thromboembolism risk in plastic surgery patients using the 2005 and 2010 Caprini Risk score. Plastic & Reconstructive Surgery, 130(2), pp.343-53.

Parker A, Peterson E, Lee A Y. Y, de Wit , C , Carrier M, Polley G, Tien J, and Wu C. (2018). Risk stratification for the development of venous thromboembolism in hospitalized patients with cancer. Journal of Thrombosis & Haemostasis, 16(7), pp.1321-1326.

Penaloza A, Melot C, and Motte S. (2011). Comparison of the Wells score with the simplified revised Geneva score for assessing pretest probability of pulmonary embolism. Thrombosis Research, 127(2), pp.81-4.

Penaloza A, Soulie C, Moumneh T, Delmez Q, Ghuysen A, El Kouri , D , Brice C, Marjanovic N S, Bouget J, Moustafa F, Trinh-Duc A, Le Gall , C , Imsaad L, Chretien J M, Gable B, Girard P, Sanchez O, Schmidt J, Le Gal , G , Meyer G, Delvau N, and Roy P M. (2017). Pulmonary embolism rule-out criteria (PERC) rule in European patients with low implicit clinical probability (PERCEPIC): a multicentre, prospective, observational study. The Lancet Haematology, 4(12), pp.e615-e621.

Piovella C, Dalla Valle, F , Trujillo-Santos J, Pesavento R, Lopez L, Font L, Valle R, Nauffal D, Monreal M, Prandoni P, and Investigators Riete. (2014). Comparison of four scores to predict major bleeding in patients receiving anticoagulation for venous thromboembolism: findings from the RIETE registry. Internal & Emergency Medicine, 9(8), pp.847-52.

Poli D, Antonucci E, Testa S, Cosmi B, Palareti G, Ageno W, and Clinics Fcsa Italian Federation of Anticoagulation. (2013). 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 & Haemostasis, 11(6), pp.1053-8.

Poli D, and Palareti G. (2013). Assessing recurrence risk following acute venous thromboembolism: use of algorithms. Current Opinion in Pulmonary Medicine, 19(5), pp.407-12.

Prandoni P, Trujillo-Santos J, Sanchez-Cantalejo E, Dalla Valle, F , Piovella C, Pesavento R, Nieto RodriGuez, J A, and Monreal M. (2010). Major bleeding as a predictor of mortality in patients with venous thromboembolism: Findings from the RIETE Registry. Journal of Thrombosis and Haemostasis, 8(11), pp.2575-2577.

Rief P, and Raggam R B. (2018). The HAS-BLED score is useful for the identification of venous thromboembolism patients with high risk for major bleeding complications: A prospective outpatient cohort study. Vasomed, 30(1), pp.36-37.

Rief P, Raggam R B, Hafner F, Avian A, Hackl G, Cvirn G, Brodmann M, and Gary T. (2018). Calculation of HAS-BLED Score Is Useful for Early Identification of Venous Thromboembolism Patients at High Risk for Major Bleeding Events: A Prospective Outpatients Cohort Study. Seminars in Thrombosis & Hemostasis, 44(4), pp.348-352.

Rieken M, and Bachmann A. (2018). Risk of Bleeding Versus Venous Thromboembolism in Urological Surgery-Finding the Right Balance Is Not Always Easy!. European Urology, 73(2), pp.252-253.

Robert-Ebadi H, Le Gal , G , Carrier M, Couturaud F, Perrier A, Bounameaux H, and Righini M. (2010). Differences in clinical presentation of pulmonary embolism in women and men. Journal of Thrombosis & Haemostasis, 8(4), pp.693-8.

Rodger MA, Kahn SR, Wells PS, Anderson DA, Chagnon I, Le Gal G, Solymoss S, Crowther M, Perrier A, White R, Vickars L, Ramsay T, Betancourt MT, and Kovacs MJ. (2008). Identifying unprovoked thromboembolism patients at low risk for recurrence who can discontinue anticoagulant therapy.. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne, 179(5), pp.417-26.

Rodger M A, Scarvelis D, Kahn S R, Wells P S, Anderson D A, Chagnon I, Le Gal , G , Gandara E, Solymoss S, Sabri E, Kovacs J, and Kovacs M J. (2016). Long-term risk of venous thrombosis after stopping anticoagulants for a first unprovoked event: A multi-national cohort. Thrombosis Research, 143, pp.152-8.

Rosenberg D, Eichorn A, Alarcon M, McCullagh L, McGinn T, and Spyropoulos A C. (2014). External validation of the risk assessment model of the International Medical Prevention Registry on Venous Thromboembolism (IMPROVE) for medical patients in a tertiary health system. Journal of the American Heart Association, 3(6), pp.e001152.

Rosenberg D J, Press A, Fishbein J, Lesser M, McCullagh L, McGinn T, and Spyropoulos A C. (2016). External validation of the IMPROVE Bleeding Risk Assessment Model in medical patients. Thrombosis & Haemostasis, 116(3), pp.530-6.

Sam A, Sanchez D, Gomez V, Wagner C, Kopecna D, Zamarro C, Moores L, Aujesky D, Yusen R, Jimenez Castro, and D . (2011). The shock index and the simplified PESI for identification of low-risk patients with acute pulmonary embolism. European Respiratory Journal, 37(4), pp.762-766.

Sanden P, Svensson P J, and Sjalander A. (2017). Venous thromboembolism and cancer risk. Journal of Thrombosis & Thrombolysis, 43(1), pp.68-73.

Shen J H, Chen H L, Chen J R, Xing J L, Gu P, and Zhu B F. (2016). Comparison of the Wells score with the revised Geneva score for assessing suspected pulmonary embolism: a systematic review and meta-analysis. Journal of Thrombosis & Thrombolysis, 41(3), pp.482-92.

Smith J G, Wieloch M, Koul S, Braun O O, Lumsden J, Rydell E, Ohman J, Schersten F, Svensson P J, Van Der Pals , and J . (2012). Triple antithrombotic therapy following an acute coronary syndrome: Prevalence, outcomes and prognostic utility of the HAS-BLED score. EuroIntervention, 8(6), pp.672-678.

Spirk D, Aujesky D, Husmann M, Hayoz D, Baldi T, Frauchiger B, Banyai M, Baumgartner I, and Kucher N. (2011). Cardiac troponin testing and the simplified Pulmonary Embolism Severity Index. The SWIss Venous ThromboEmbolism Registry (SWIVTER). Thrombosis & Haemostasis, 106(5), pp.978-84.

Squizzato A, Donadini M P, Galli L, Dentali F, Aujesky D, and Ageno W. (2012). Prognostic clinical prediction rules to identify a low-risk pulmonary embolism: a systematic review and meta-analysis. Journal of Thrombosis & Haemostasis, 10(7), pp.1276-90.

Stoker T B, and Evans N R. (2016). Managing risk after intracerebral hemorrhage in concomitant atrial fibrillation and cerebral amyloid angiopathy. Stroke, 47(7), pp.e190-e192.

Suarez C, Gonzalez R, Arcelus J I, Barba R, Blanco A, Barron M, Buges J, Calvo J M, Casado I, Epelde F, Falga C, Fernandez-Capitan C, Gallego P G, Arcia-Bragado F, Grau E, Guijarro R, Guil M, Gutierrez J, Gutierrez M R, Hernandez L, Jimenez D, Lecumberri R, Lopez F, Lopez L, Lopez I, Madridano O, Maestre A, Martin-Villasclaras J J, Monreal M, Montes J, Naufall M D, Nieto J A, Nunez M J, Orue M T, Perez J L, Portillo J, Rabunal R, Raguer E, Roman P, Ruiz-Gimenez N, Samperiz A L, Sanchez R, Sanchez J F, Soler S, Tiberio G, Tirado R, Todoli J A, Tolosa C, Trujillo J, Valle R, Vela J, Mismetti P, Rivron-Guillot K, Le Gal , G , Di Micco , P , Enea I, Ghirarduzzi A, Iannuzo M T, Poggio R, Prandoni P, Quintavalla R, and Tiraferri E. (2008). Predictive variables for major bleeding events in patients presenting with documented acute venous thromboembolism. Findings from the RIETE Registry. Thrombosis and Haemostasis, 100(1), pp.26-31.

Tan C, Sim T B, and Thng S Y. (2013). Validation of the San Francisco Syncope Rule in two hospital emergency departments in an Asian population. Academic Emergency Medicine, 20(5), pp.487-97.

Thomas I C, and Sorrentino M J. (2014). Bleeding risk prediction models in atrial fibrillation topical collection on invasive electrophysiology and pacing. Current Cardiology Reports, 16 (1) (no pagination)(432), pp..

Tosetto A, Iorio A, Marcucci M, Baglin T, Cushman M, Eichinger S, Palareti G, Poli D, Tait R C, and Douketis J. (2012). Predicting disease recurrence in patients with previous unprovoked venous thromboembolism: a proposed prediction score (DASH). Journal of Thrombosis & Haemostasis, 10(6), pp.1019-25.

Touhami O, Marzouk S B, Bennasr L, Touaibia M, Souli I, Felfel M A, Kehila M, Channoufi M B, and Magherbi H E. (2018). Are the Wells Score and the Revised Geneva Score valuable for the diagnosis of pulmonary embolism in pregnancy?. European Journal of Obstetrics, Gynecology, and & Reproductive Biology, 221, pp.166-171.

Tritschler T, Mean M, Limacher A, Rodondi N, and Aujesky D. (2015). Predicting recurrence after unprovoked venous thromboembolism: prospective validation of the updated Vienna Prediction Model. Blood, 126(16), pp.1949-51.

Trujillo-Santos J, Lozano F, Lorente M A, Adarraga D, Hirmerova J, Del Toro , J , Mazzolai L, Barillari G, Barron M, Monreal M, and Investigators Riete. (2015). A prognostic score to identify low-risk outpatients with acute deep vein thrombosis in the lower limbs. American Journal of Medicine, 128(1), pp.90.e9-15.

Tsu L V, Berry A, Wald E, and Ehrlich C. (2015). Modified HAS-BLED score and risk of major bleeding in patients receiving dabigatran and rivaroxaban: A retrospective, case-control study. Consultant Pharmacist, 30(7), pp.395-402.

van Es , N , Di Nisio , M , Cesarman G, Kleinjan A, Otten H M, Mahe I, Wilts I T, Twint D C, Porreca E, Arrieta O, Stepanian A, Smit K, De Tursi , M , Bleker S M, Bossuyt P M, Nieuwland R, Kamphuisen P W, and Buller H R. (2017). Comparison of risk prediction scores for venous thromboembolism in cancer patients: a prospective cohort study. Haematologica, 102(9), pp.1494-1501.

van Es , N , Wells P S, and Carrier M. (2017). Bleeding risk in patients with unprovoked venous thromboembolism: A critical appraisal of clinical prediction scores. Thrombosis Research, 152, pp.52-60.

Vink R, Kraaijenhagen R A, Levi M, and Buller H R. (2003). Individualized duration of oral anticoagulant therapy for deep vein thrombosis based on a decision model. Journal of Thrombosis & Haemostasis, 1(12), pp.2523-30.

Vuilleumier N, Limacher A, Mean M, Choffat J, Lescuyer P, Bounameaux H, Aujesky D, and Righini M. (2015). Cardiac biomarkers and clinical scores for risk stratification in elderly patients with non-high-risk pulmonary embolism. Journal of Internal Medicine, 277(6), pp.707-16.

Wells P S, Forgie M A, Simms M, Greene A, Touchie D, Lewis G, Anderson J, and Rodger M A. (2003). The outpatient bleeding risk index: validation of a tool for predicting bleeding rates in patients treated for deep venous thrombosis and pulmonary embolism. Archives of Internal Medicine, 163(8), pp.917-20.

Wells Ps, Kovacs Mj, Anderson D, Kahn Sr, Kearon C, Schulman S, Keeling Dm, Kaatz S, Solymoss S, Corsi D, and et al. (2016). Prediction of bleeding risk in patients on extended oral anticoagulation for venous thromboembolism. Blood. Conference: 58th annual meeting of the american society of hematology, and ASH 2016. United states. Conference start: 20161203. Conference end: 20161206, 128(22) (no pagination), pp..

Wicki J, Perrier A, Perneger T V, Bounameaux H, and Junod A F. (2000). Predicting adverse outcome in patients with acute pulmonary embolism: a risk score. Thrombosis & Haemostasis, 84(4), pp.548-52.

Zhou H X, Peng L Q, Yan Y, Yi Q, Tang Y J, Shen Y C, Feng Y L, and Wen F Q. (2012). Validation of the Caprini risk assessment model in Chinese hospitalized patients with venous thromboembolism. Thrombosis Research, 130(5), pp.735-40.

Zhou X Y, Ben S Q, Chen H L, and Ni S S. (2012). The prognostic value of pulmonary embolism severity index in acute pulmonary embolism: A meta-analysis. Respiratory Research, 13 (no pagination)(111), pp..

Excluded studies (search update)

Alatri, A., Mazzolai, L., Font, C. et al. (2017) Low discriminating power of the modified Ottawa risk score in a cohort of patients with active cancer and acute venous thromboembolism. Vasa - European Journal of Vascular Medicine 46 (Supplement 98): 20-21

Dhanda, S., Davies, M., Roy, D. et al. (2018) Risk of major and clinically relevant non-major (CRNM) bleeding in patients prescribed rivaroxaban in primary care in England. Drug Safety 41 (11): 1130-1131

Jaquet, E., Tritschler, T., Stalder, O. et al. (2018) Prediction of short-term prognosis in elderly patients with acute pulmonary embolism: validation of the RIETE score. Journal of Thrombosis & Haemostasis 16(7): 1313-1320

Palareti, G., Antonucci, E., Mastroiacovo, D. et al. (2018) The American College of Chest Physician score to assess the risk of bleeding during anticoagulation in patients with venous thromboembolism. Journal of Thrombosis and Haemostasis 16(10): 1994-2002

Wells, P., Peacock, W. F., Fermann, G. J. et al. (2019) The value of sPESI for risk stratification in patients with pulmonary embolism. Journal of Thrombosis and Thrombolysis.

Other references

Khan, F., Rahman, A., Carrier, M., Kearon, C., Weitz, J. I., Schulman, S., ... & Agnelli, G. (2019). Long term risk of symptomatic recurrent venous thromboembolism after discontinuation of anticoagulant treatment for first unprovoked venous thromboembolism event: systematic review and meta-analysis. *bmj*, *366*, l4363.

Wolff R, Moons K, Riley R, Whiting P, Westwood M, Collins G, Reitsma J, Kleijnen J, Mallett S. PROBAST – A risk-of-bias tool for prediction-modelling studies. Abstracts of the Global Evidence Summit, Cape Town, South Africa. Cochrane Database of Systematic Reviews 2017, Issue 9 (Suppl 1). https://doi.org/10.1002/14651858.CD201702.

Now published as:

Wolff RF, Moons KGM, Riley RD, Whiting PF et al. (2019) PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019; 170: 51-58

Moons KGM, Wolff RF, Riley RD, Whiting PF et al. (2019) PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med. 2019; 170: W1-W33

Appendix K – Research recommendations

Research recommendation 1

Research recommendation	What is the prognostic accuracy of a tool to predict both VTE-recurrence and major bleeding after 3 months of initial anticoagulation treatment and in the long-term? **[2019]**
Population	Adults (aged 18+) with confirmed VTE who have already received ≥3 months' anticoagulation treatment.
Intervention(s)	Any risk tool using multiple predictors to predict both VTE recurrence (if treatment is stopped) and major bleeding (if treatment is continued) following an initial treatment duration of ≥3 months.
Outcomes	Recurrence of VTE and major bleeding
Outcome measures	For each outcome, prognostic accuracy measures will be reported where available, for example: • Odds ratios/hazard ratios • Model fit (e.g. r squared) • Sensitivity, specificity, positive and negative predictive values. • C-statistics • Observed and estimated risks
Study design	A prognostic test accuracy study designing and externally validating a combined tool using multiple predictors to predict the risk of VTE recurrence following cessation of anticoagulant treatment and major bleeding during continued treatment that gives a combined score regarding the decision to stop or continue anticoagulation treatment.
Subgroup analyses	• Age • Sex (male/female) • D-dimer (as a dichotomous variable) • Hormone therapy (yes vs. no)

Potential criterion	**Explanation**
Importance to patients, service users or the population	There is uncertainty surrounding the best way to estimate the risk associated with stopping or continuing anticoagulation treatment. Current prediction tools only predict recurrence **or** major bleeding individually, with no tool predicting both. Current tools therefore do no translate that well into a clinical decision to continue or stop anticoagulation treatment. The development of a tool that could assess the combined risks of major bleeding **and** VTE recurrence will be helpful in providing information about the balance the benefits and harms of continuing treatment at the individual level and could be used as part of the decision- making process for stopping treatment after the initial treatment period or continuing treatment longer term.
Relevance to NICE guidance	Medium priority: the committee were able to make recommendations based on current evidence, recommending the use of a tool to predict major bleeding and a different tool to predict VTE-recurrence in subsets of the

Potential criterion	Explanation
	population, but they could not recommend a tool that was able to do both as none were identified. This research will help to create better prediction tools and improve the recommendations.
Current evidence base	After initial treatment, a clinician must balance the benefits and harms of extended treatment, by weighing the risk of recurrence associated with stopping treatment against the risk of bleeding associated with continuing treatment. This review identified 18 studies reporting data on 15 different prognostic tools to predict major bleeding. However, none of these tools have better than "good" prognostic classification accuracy, and most tools were found to have only poor to adequate classification accuracy. Additionally, most of the tools were developed (and externally validated) in cohorts of people beginning treatment however the intended use of a tool to predict major bleeding is for it to be used after a person with VTE has completed 3 months of anticoagulation treatment to help inform the decision to continue or stop treatment. Only one study (Seiler, 2017) tested tools in people who had received 3 months treatment. This review only identified 5 tools for predicting VTE-recurrence and have only been assessed in one external validation study each. The committee were concerned with the poor classification accuracy of these tests and their limited clinical utility. Additionally, the reporting of the diagnostic accuracy of these tools has been poor, typically reporting only C-statistics without confidence intervals. The committee advised that a better tool is needed to predict VTE-recurrence and that this tool needs to be applicable to all people with VTE. No tools were identified that assessed the combined risks of VTE recurrence **and** major bleeding.
Equality	No specific equality concerns are relevant to this research recommendation.
Feasibility	There is a sufficiently large and well-defined population available that a high quality prognostic accuracy study should be feasible.

Research recommendation 2

Research recommendation	What is the prognostic accuracy of tools to predict both VTE-recurrence **and** major bleeding compared with clinician's judgment in people with unprovoked proximal DVT or PE? **[2019]**
Population	Adults (aged 18+) with confirmed unprovoked VTE who have already received ≥3 months' anticoagulation treatment.
Intervention(s)	Any risk tool using multiple predictors to predict both VTE recurrence (if treatment is stopped) **and** major bleeding (if treatment is continued) following an initial treatment duration of ≥3 months.
Comparator	Clinician judgement
Outcomes	Recurrence of VTE and major bleeding
Outcome measures	• Model fit (e.g. r squared) • Sensitivity, specificity, positive and negative predictive values. • C-statistics • Risk ratios/hazard ratios
Study design	Note that it is intended that this research is only commenced after research recommendation 1 has produced a new prognostic tool capable of predicting both VTE-recurrence **and** major bleeding. Prospective cohort study or test and treat RCT. **Prospective cohort study**: treating physicians will estimate the risk of recurrence and major bleeding (high or low risk only) for each participant (without using a prognostic tool) and treat as normal. At the same time, the clinical characteristics needed for the prognostic tool will be collected amongst other relevant information. At the end of the study, the risk of VTE recurrence and major bleeding will be assessed using the joint tool to give a decision to continue (low risk group) or stop treatment (high risk group). The tools rating of low and high risk and the clinician's judgement will then be compared to the patient outcomes (VTE recurrence and major bleeding with major bleed on treatment or VTE recurrence off treatment counting as the wrong decision.) The results for prognostic accuracy of the tool and clinician gestalt can then be compared directly using the same patient population. **Test and treat RCT**: A study randomly assigning participants to have their risk of bleeding **and** recurrence assessed by either a joint risk tool or clinician's judgement, with subsequent treatment being administered accordingly and the effects on major bleeding and VTE recurrence, amongst others, assessed.
Subgroup analyses	• Age • Sex (male/female) • D-dimer (as a dichotomous variable) • Hormone therapy (yes vs. no)

Potential criterion	Explanation
Importance to patients, service users or the population	After initial treatment, it is important that any decision to continue or stop treatment balances the risk of recurrence against the risk of bleeding. There is uncertainty as to whether predictive tools improve decision making over clinician's gestalt. This research will help to identify the best method of estimating a person's combined risk of major bleeding and recurrence and help to better balance the benefits and harms of continuing and stopping anticoagulant treatment. Any tool capable of better prognostic accuracy than clinician judgement would be useful in the decision to continue or stop anticoagulant treatment, regardless of the tool's level of accuracy.
Relevance to NICE guidance	High priority: The committee were able to make recommendations to use tools that predict major bleeding or VTE-recurrence, but these tools were not useful in all scenarios and did not give a combined output to inform the decision to continue or stop treatment. The results of testing a new tool that was able to predict combined VTE recurrence **and** major bleeding risk and that was shown to be better than clinician judgement would substantially change and improve the recommendations.
Current evidence base	After initial treatment, a clinician must balance the benefits and harms of extended treatment, by weighing the risk of recurrence associated with stopping treatment against the risk of bleeding associated with continuing treatment. The currently available prognostic tools offer less than ideal prognostic accuracy and there is uncertainty as to whether these tools are better able to classify risk than clinician's gestalt, as the evidence identified for this review did not compare any of the tools to clinical judgment. After a new prognostic tool (capable of predicting both VTE-recurrence **and** major bleeding) is established and externally verified, a study is needed to see whether this prediction tool has better classification accuracy than clinician's gestalt alone, and whether it improves outcomes.
Equality	No specific equality concerns are relevant to this research recommendation.
Feasibility	There is a sufficiently large and well-defined population available that a high quality prospective cohort study should be feasible. A high quality RCT may be less feasible than a prospective study as there are a likely to be more limited numbers of people willing to take part. Of note, there are no currently available tools that area capable of predicting both VTE-recurrence and major bleeding in a single score and therefore such a tool needs to be derived and externally validated – in accordance with research recommendation 1 before the study in this research recommendation is possible.