5 Risk assessment for medical, surgical and trauma patients

5.1 Introduction

Risk assessment is a crucial part of deciding whether to give prophylaxis. When making a judgement on using an intervention to reduce the risk of VTE, it is important to consider:

- the reason for admission to hospital (for example, a surgical procedure or a medical problem) and factors individual to the patient concerned (for example age, gender, pre-existing medical conditions and medication use) that influence the likelihood of VTE
- the likely treatment benefit from the specific prophylactic intervention
- the possible harmful effect of the intervention.

Pharmacological methods are widely used for VTE prophylaxis. These come with the potential harm of increasing the risk of bleeding. Major bleeding is clearly a threat to life, but under some circumstances a low volume bleed can be a very major complication. A few millilitres of bleeding into the brain, or compressing the spinal cord within the vertebral canal can cause death or permanent neurological damage.

The risk assessment recommendations from the last version of the guideline (CG92) aligned with a tool produced by the Department of Health which has since become known as the National VTE Risk Assessment Tool.[73] In 2010 NICE introduced a quality standard requiring all patients to receive an assessment of VTE and bleeding risk on admission using the clinical risk assessment criteria described in the National Tool. [130] Subsequently, the Department of Health Commissioning for Quality and Innovation (CQUIN) payment framework linked the uptake of risk assessment with payments. Since 2012 over 90% of hospital admissions were risk assessed for VTE using the National Tool.

This current version of the guideline reviewed the evidence for existing risk assessment tools or checklists for VTE and bleeding. The reviews covered:

- both the predictive accuracy and clinical and cost effectiveness of tools
- tools that included VTE and bleeding risk together in a tool or as separate tools
- tools that grouped all populations together or separated them into reasons for attending hospital, for example, surgical patients, medical inpatients or patients undergoing day procedures.

After admission or a procedure at hospital a person's medical condition will usually change. As a consequence of this change their risk of VTE and bleeding may also change. The last version of the guideline (CG92) recommended patients were reassessed every 24 hours. This update reviewed the evidence for the effectiveness of reassessment of VTE and bleeding risk to establish if this time point was appropriate for some or all patients.

5.2 Accuracy of risk assessment tools for VTE in hospital admissions

5.2.1 Review question: What is the accuracy of individual risk assessment or prediction tools in predicting the likelihood of VTE in a patient who is admitted to hospital?

For full details see review protocol in appendix C.

 ISBN 978 – 1 – 4731 – 2871 - 2

Table 8: PICO characteristics of review question

Question	What is the accuracy of individual risk assessment or predication tools in predicting the likelihood of VTE in a patient who is admitted to hospital?
Population	Adults and young people (aged 16 or over) admitted to hospital
Risk tool	Derived and validated risk tools identified in literature
Target condition(s)	• VTE (symptomatic or asymptomatic) (up to 90 days) • VTE-related mortality (up to 90 days) • DVT alone (up to 90 days) • PE alone (up to 90 days)
Outcomes (in terms of predictive test accuracy, calibration)	Statistical outputs may include: • Discrimination (sensitivity, specificity, predictive values) • Area under the ROC curve (c-statistic) • Predicted risk versus observed risk (calibration) • Reclassification • Other statistical measures: for example, D statistic, R^2 statistic and Brier score
Study types	Prospective and retrospective cohort Exclusions: derivation studies

5.2.2 Clinical evidence

Twenty-two studies evaluating 13 risk assessment models were included in the review, [10, 18, 63, 64, 68, 75, 77, 79, 105, 106, 133, 140, 146, 147, 150, 164, 165, 175, 189, 190, 201, 202] these are summarised in Table 9 below. See also the study selection flow chart in appendix E, study evidence tables in appendix H, and excluded studies list in appendix N. Full details of the tools included in this review are provided in the clinical evidence tables in appendix H.

Seven studies focused on VTE risk assessment in hospitalised medical patients,[63, 133, 165, 189, 202] including one specifically on hospitalised cancer patients.[150] Ten focused on surgical patients,[10, 18, 68, 77, 106, 140, 146, 175, 190, 201] three focused on trauma patients,[75, 79, 164] and study each on VTE risk assessment in people after a stroke[105] and people with thermal (burn) injuries.[147]

The risk assessment models identified by the literature included the Caprini risk assessment model, the Kucher score, the Geneva risk score, the predictive (4 factor) IMPROVE tool, the Intermountain risk assessment model, the Khorana Score, the Padua Prediction Score and the Trauma Embolic Scoring System (TESS).

Table 9: Summary of studies included in the review

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
Risk assessment in medical patients					
Grant 2016 [63]	Caprini Score	n= 63,548 Hospitalised medical patients USA	VTE, hospital associated (90 days): Proximal upper or proximal lower extremity DVT and PE. VTE events must have occurred on the third day after admission or later (up to 90 days after admission). Diagnosis of DVT was	n= 670 (1.05%)	Retrospective cohort

 ISBN 978 – 1 – 4731 – 2871 - 2

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
			based on positive findings via compression Doppler ultrasound or venography, PE was confirmed via computed tomography (CT) scan, ventilation perfusion scan or pulmonary angiography Sensitivity and specificity		
Greene 2016 [64]	Kucher Score Padua Prediction Score International Medical Prevention Registry on Venous Thromboembolism (IMPROVE) Intermountain risk assessment model	n= 63,548 Acutely ill, hospitalised medical patients USA	VTE, hospital associated (90 days): Proximal upper or proximal lower extremity DVT and PE. VTE events must have occurred on the third day after admission or later (up to 90 days after admission). Diagnosis of DVT was based on positive findings via compression Doppler ultrasound or venography, PE was confirmed via computed tomography (CT) scan, ventilation perfusion scan or pulmonary angiography C-statistic	n= 670 (1.05%)	Prospective cohort
Nendaz 2014 [133]	Geneva Risk Score Padua Prediction Score	n=1478 Acutely medically ill patients Age: 65%(>60 years); 44% (≥ 70 years) Gender (male to female ratio): not reported Switzerland	Symptomatic VTE (90 days) including PE or DVT. PE was confirmed by contrast-enhanced computer tomography, ventilation perfusion scan or conventional pulmonary angiography, and DVT by compression ultrasound or venography. Sensitivity Specificity	n= 30 (2.3%)	Prospective cohort

 ISBN 978 – 1 – 4731 – 2871 - 2

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
			PPV NPV NLR		
Patell 2017 [150]	Khorana Score	n=2780 Hospitalised cancer patients Age, median (range): 62 (19-98) years. Gender (male to female ratio): 1545: 1235 USA	VTE: based on ICD-9 codes Sensitivity and specificity calculated using prevalence and risk tool data reported.	n=106 (3.8%)	Retrospective cohort
Rothberg 2011 [165]	Unnamed (Rothberg 2011)	n= 48, 540 Medical patients Age: 18-49 years; 12.9%, 50-64 years, 21.1%, 65+ years 66.0% Gender (male to female ratio): 41.6 :58.4 Primary Diagnosis: Community-Acquired Pneumonia 33.5%; Septicaemia 3.2%; Chronic Obstructive Pulmonary Disease 14.5%; Respiratory Failure 2.8%; Congestive Heart Failure 19.2%; Cardiovascular Disease 13.6%; Urinary Tract Infection 13.1% USA	VTE, hospital acquired (3 days after hospitalisation - end point not reported): diagnosis by lower extremity ultrasound, venography, CT angiogram, ventilation-perfusion scan or pulmonary angiogram on hospital day 3 or later; received treatment for VTE at least 50% of the remaining hospital stay; until initiation of warfarin; appearance of a complication (e.g. transfusion or treatment for heparin-induced thrombocytopenia) and were given secondary diagnosis of VTE	n= 223 (0.46%)	Retrospective cohort
Vardi 2013 [189]	Padua Prediction Score	n= 1080 People with sepsis admitted to internal medicine departments Age (mean± SD): 74.68± 16.15	VTE (time point: For in hospital VTE our assumption is that it is an event between 48 hours after admission and discharge) Includes DVT or PE. Diagnosis of DVT by Duplex ultrasound or	n=14 (1.29%)	Prospective cohort

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
		Gender (male to female ratio): 1.09:1 Israel	computer tomography (CT) and diagnosis of PE was based on a positive CT angiography (CTA) or a high-probability ventilation perfusion scan. C-statistic		
Woller 2011 [202]	Intermountain risk assessment model Kucher Score	n=46856 (for both risk tools) Medically ill patients Age (mean): 61.14 years Gender (male to female ratio): 1.17:1 USA	VTE (90 days): not defined. C-statistic	n=2109 (4.5%)	Retrospective cohort
Risk assessment in surgical patients					
Bahl 2010 [10]	Caprini risk assessment model	n=8216 Undergoing major surgery (>45 minutes) 88.16%; general 67%, vascular 16%, 17% urologic) Age: <40 years 19.28%, 40-60 years 39.59%, 61-74 years 28.4%, 75+ years 12.73% Gender (male to female ratio): not reported USA	VTE (30 days): not defined. C-statistic Hosmer-Lemeshow test	n=118 (1.44%)	Retrospective cohort
Bilimoria 2013 [18]	American College of Surgeons National Surgical Quality Improvement Programme (ACS NSQIP) Universal Surgical Risk Calculator	Colon surgery n= 88,334 Undergoing colorectal surgery Age and gender: no details of validation cohort USA	VTE (30 days): not defined. C-statistic Brier score	Colon surgery n=3508 (4%)	Retrospective cohort

 ISBN 978 – 1 – 4731 – 2871 - 2

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
Hachey 2016 [68]	Caprini risk assessment model	n=232 People undergoing lung cancer resections: lobectomy (84.5%), segmenectomy (8.2%), pneumonectomy (7.3%) Age: Adults (with VTE mean 63.83±10.2 years, without VTE mean 64.36±11 years) Gender (male to female ratio): 100:132 USA	VTE (60 days): any PE or DVT identified via clinical imaging studies (computed tomography pulmonary angiogram or duplex ultrasound) and treated with therapeutic anticoagulation or inferior vena cava filter. Sensitivity Specificity C-statistic PPV NPV Hosmer-Lemeshow test	n=12 (5.2%)	Retrospective cohort
Hewes 2015 [77]	Modified Caprini score	n=70 Undergoing oesophagectomy for oesophageal cancer Age: with VTE mean 64.9±6.4, without VTE mean 61.6±11.7 Gender (male to female ratio): 58:12 USA	VTE (1-60 days): defined as any thromboembolic event diagnosed by appropriate imaging findings and treated with therapeutic anticoagulation or inferior vena cava filter. Sensitivity Specificity PPV NPV C-statistic Hosmer-Lemeshow test	n=10 (14.3%)	Retrospective cohort
Lobastov 2016 [106]	Caprini risk assessment model	n=140 High-risk patients who underwent emergency abdominal (48%) or cranial/spinal (52%) surgery already receiving pharmacological prophylaxis Age, mean (SD): 69.2 (12.2)	"Fresh" DVT or PE at the hospital treatment stage – occlusion of previously unaffected vein segments: duplex ultrasonography of the lower limbs, and static lung perfusion scintigraphy or combined single proton emission CT and x-ray CT of the lungs, or autopsy.	n=39 (27.9%)	Retrospective cohort

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
		Gender (male to female ratio): 68:72 Russia	Sensitivity Specificity C-statistic		
Obi 2015 [140]	Caprini risk assessment model	n=4844 Critically ill surgical patients (surgical ICU). Including general surgery, transplant, urology, and orthopaedic patients and patients with respiratory failure requiring extracorporeal membrane oxygenation 82% major operative procedures Age: <41 years 15.9%; 41-60 years 40%; 61-74 years 29.4%; ≥75 years 14.8% Gender (male to female ratio): not reported USA	VTE (time point unclear): defined as patients with DVT or PE which occurred during the patient's initial hospital admission. DVT included acute thrombosis of lower-extremity veins (iliac, femoral, popliteal, or calf veins) or upper-extremity veins (axillary, subclavian, brachial, or internal jugular veins). PE defined as acute thrombosis within the pulmonary vasculature. VTE considered present if identified with an objective imaging study, including duplex ultrasonography or PE protocol computed tomography. Patients who experienced sudden death were included if post-mortem examination documented definitive evidence of VTE C-statistic Hosmer-Lemeshow test	DVT n= 308 (6.4%) PE n=79 (1.6%)	Retrospective cohort
Pannucci 2014 [146]	Unnamed (Pannucci 2014)	n=3576 Postsurgical patients (details of surgical procedures not provided for validation sample) Overall age: ≥ 60 years:	VTE (90 days): Patients with either PE or PE. Upper extremity DVT included clots in the jugular, subclavian, axillary, or brachial veins. Lower extremity DVT included clots in the vena cava, femoral,	n= 50 (1.40%)	Prospective cohort

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
		62% Overall gender (male to female ratio): 1:1.36 USA	tibial, or popliteal veins. PE included clots in the pulmonary vasculature. All VTE events were diagnosed using an objective imaging study. C-statistic		
Shaikh 2016 [175]	Caprini risk assessment model	n=1598 People undergoing plastic surgery Age, mean (range): 49.9 (14-86) years Gender (male to female ratio): 308:1290 BMI, mean (range): 28.2 (15.9-77.5) kg/m2 USA	VTE: DVT/PE composite: not defined Sensitivity Specificity	n=24 (1.5%)	Retrospective cohort
Vaziri 2017 [190]	American College of Surgeons National Surgical Quality Improvement Programme (ACS NSQIP) Universal Surgical Risk Calculator	n=1006 People undergoing neurosurgery Age not reported. Gender (male/female): 460/546 USA	VTE: no further details provided c-statistic	n=13 (1.29%)	Retrospective cohort
Winoker 2017 [201]	American College of Surgeons National Surgical Quality Improvement Programme (ACS NSQIP) Universal Surgical Risk Calculator	n=300 People undergoing urological surgery, specifically robot-assisted partial nephrectomy Age (%): <65 (63.7); 65-73 (26.3); 75-84 (9.7); ≥85 (0.3) years Gender (male to female ratio): 185:115	VTE: no further details provided c-statistic Brier score	n=1 (0.33%)	Retrospective cohort

 ISBN 978 – 1 – 4731 – 2871 - 2

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
		BMI (%): <18.5 (0.7); 18.5-24.9 (13.3); 25-29.9 (39.7); ≥30 (46.3) kg/m2 USA			
Risk assessment in people with trauma					
Hegsted 2013 [75]	Risk Assessment Profile (RAP)	n=2281 People with trauma Age (mean): 45.2 years Gender (male to female ratio): 2.33:1 USA	DVT (time point unclear): not defined PE (time point unclear): detected by computed tomography-angiography or post-mortem examination Sensitivity Specificity PPV NPV	DVT n= 239 (10.5%) PE n=34 (1.5%)	Retrospective cohort
Ho 2014 [79]	Trauma Embolic Scoring System (TESS)	n=357 People with trauma Chest injury: 61.9% Abdominal injury: 29.1% Spinal fractures: 43.4% Pelvic fractures: 32.8% Lower limb fractures: 38.4% Age: mean (IQR): VTE event 42 (23-55) years; No VTE event 31 (21-45) years Gender (male to female ratio): VTE event 3.6:1; No VTE event 2.82:1 Australia	VTE (time point unclear): DVT and PE confirmed by colour Doppler compression ultrasound and computed tomography pulmonary angiography or post mortem examination. Sensitivity Specificity PPV NPV Hosmer-Lemeshow test	Overall VTE: n=74 (21%) Fatal PE: n= 16 (4.48%) Non-fatal PE: 22 (6.16% DVT: 47 (13.17%)	Retrospective cohort
Rogers 2012 [164]	Trauma Embolic Scoring System (TESS)	n=234,032 People with trauma Injury type: blunt 86.9%, burn 2.5%, penetrating 10.6% (missing data for	VTE (unclear time point): included deep vein thrombosis (DVT) or pulmonary embolism (PE) DVT: The formation,	n=4,881 (1.4%)	Retrospective cohort

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
		26,928) Age: <30 years 40.9%, 30-64 years 41.7%, ≥65 years 17.4% Gender (male to female ratio): 1.92:1 USA	development, or existence of a blood clot or thrombus within the vascular system, which may be coupled with inflammation. This diagnosis may be confirmed by a venogram, ultrasound, or CT. The patient must be treated with anticoagulation therapy and/or placement of a vena cava filter or clipping of the vena cava. PE: Defined as a lodging of a blood clot in a pulmonary artery with subsequent obstruction of blood supply to the lung parenchyma. The blood clots usually originate from the deep leg veins or the pelvic venous system. Consider the condition present if the patient has a V-Q scan interpreted as high probability of pulmonary embolism or a positive pulmonary arteriogram or positive CT angiogram. Sensitivity Specificity C-statistic PPV NPV Hosmer-Lemeshow test		
Risk assessment in people post-stroke					
Liu 2014 [105]	Post-stroke DVT Prediction System	n=287 Acute stroke patients	DVT (14±3 days): Diagnosis of DVT if complete compression duplex	n=30 (10.6%)	Prospective cohort

 ISBN 978 – 1 – 4731 – 2871 - 2

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
		Age: ≥65 years 58.2% Gender (male to female ratio: 1.68:1 China	ultrasonography (CCUS) showed loss of vein compressibility by ultrasonic probe pressure, a clot, or an abnormal flow pattern (loss of phasic flow signal or loss of augmentation of flow) with distal compression) C-statistic		
Risk assessment in people with thermal injuries (burns)					
Pannucci 2012 [147]	Simple Venous Thromboembolism Risk Scoring Tool	n= 5761 People with thermal injury Age (mean): 45.6 years Gender (male to female ratio): 2.33:1 USA and Canada	VTE (time point unclear: not defined) C-statistic	n=559 (9.7%)	Retrospective cohort

 ISBN 978 – 1 – 4731 – 2871 - 2

5.2.3 Discrimination

5.2.3.1 VTE

5.2.3.1.1 General medical patients

Table 10: Clinical evidence profile: risk tools for predicting VTE in general medical patients

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
Risk tool: Caprini risk assessment model										
Caprini risk assessment model Cut-off 5	1	63548	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	69.7 (66-73)	50.3 (50-51)	-	VERY LOW
Caprini risk assessment model Cut-off 7	1	63548	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	42.7 (39-47)	74.7 (74-75)	-	VERY LOW
Caprini risk assessment model Cut-off 9	1	63548	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	18.5 (16-22)	89.0 (89-89)	-	VERY LOW
Risk tool: Geneva Risk Score										
Geneva Risk Score High risk ≥3	1	1478	Serious [a]	No serious inconsistency[b]	No serious indirectness[c]	No serious imprecision[d]	90 (73.5-97.9)	35.3 (32.8-37.8)	-	MODERATE
Risk tool: IMPROVE (Predictive version - four factors available at admission)										
IMPROVE High risk ≥2	1	63548	Serious [a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	-	-	0.570 (0.565-0.576)	LOW

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
Risk tool: Intermountain										
Intermountain High risk ≥1	2	110404	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	-	-	0.611 (0.605-0.618) 0.843 (0.833-0.852)	VERY LOW
Risk tool: Kucher score										
Kucher Score High risk ≥4	2	110404	Serious [a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	-	-	0.563 (0.558-0.568) 0.683 (0.673-0.691)	LOW
Risk tool: Padua Prediction score										
Padua Prediction Score High risk ≥4	3	66106	Very serious[a]	No serious inconsistency[b]	No serious indirectness[c]	Serious imprecision[d]	73.3 (54.1-87.7)	51.9 (49.3-54.5)	0.60 (0.59-0.61) 0.58 (0.43-0.73)	VERY LOW
Risk tool: Unnamed (Rothberg 2011)										
(Unnamed)	1	48540	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	-	-	0.75 (0.71-0.78)	VERY LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

(d) Imprecision was assessed according to the range of point estimates of the primary measure (sensitivity where possible, or if missing then c-statistic). The evidence was downgraded by 1 increment when there was a 20-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

 ISBN 978 – 1 – 4731 – 2871 - 2

General medical –oncology inpatients

Table 11: Clinical evidence profile: risk tools for predicting VTE in hospitalised cancer patients

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
Khorana Score High-risk ≥3	1	2780	Very serious[a]	No serious inconsistency[b]	No serious indirectness[c]	No serious imprecision[d]	18.9 (12-28)	87.2 (86-88)	-	LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

(d) Imprecision was assessed according to the range of point estimates of the primary measure (sensitivity where possible, or if missing then c-statistic). The evidence was downgraded by 1 increment when there was a 20-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

5.2.3.1.2 Surgical patients

Mixed surgical patients

Table 12: Clinical evidence profile: risk tools for predicting VTE in mixed surgical patients

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
Caprini score	2	13060	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	Not estimable	-	-	0.585 0.698	VERY LOW
Unnamed risk model (Pannucci 2014)	1	3576	Very serious[a]	No serious inconsistency[b]	No serious indirectness[c]	Not estimable	-	-	0.70	LOW

ISBN 978 – 1 – 4731 – 2871 - 2

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

(d) Imprecision was assessed according to the range of point estimates of the primary measure. The evidence was downgraded by 1 increment when there was a 20-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

Colorectal surgery patients

Table 13: Clinical evidence profile: risk tools for predicting VTE in people undergoing colorectal surgery

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
American College of Surgeons National Surgical Quality Improvement Programme: Universal Surgical Risk Calculator	1	88,334	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	Not estimable	-	-	0.7203	VERY LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

(d) Imprecision was assessed according to the range of point estimates of the primary measure. The evidence was downgraded by 1 increment when there was a 20-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

People undergoing lung cancer resections

Table 14: Clinical evidence profile: risk tools for predicting VTE in people undergoing surgery for lung cancer

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
Caprini score Moderate to high risk >5	1	232	Very serious [a]	No serious inconsistency[b]	No serious indirectness[c]	No serious imprecision[d]	100 (100 – 100)	7.2 (4.1 – 11)	-	LOW
Caprini score Cut-off >7 (chosen to ensure 100% sensitivity)	1	232	Very serious [a]	No serious inconsistency[b]	No serious indirectness[c]	No serious imprecision[d]	100 (100 – 100)	31.4 (25 – 37.3)	-	LOW
Caprini score High risk >9	1	232	Very serious [a]	No serious inconsistency[b]	No serious indirectness[c]	Very serious imprecision[d]	83.3 (58.3 – 100)	60.5 (54.4 – 67.3)	0.72	VERY LOW
Caprini score Cut-off >10 (chosen for highest c-statistic)	1	232	Very serious [a]	No serious inconsistency[b]	No serious indirectness[c]	Very serious imprecision[d]	75 (50 -100)	69.6 (64.4 – 76.4)	0.73	VERY LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability
(d) Imprecision was assessed according to the range of point estimates of the primary measure. The evidence was downgraded by 1 increment when there was a 20-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

Oesophageal cancer surgery patients

Table 15: Clinical evidence profile: risk tools for predicting VTE in people undergoing oesophagectomy for oesophageal cancer

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
Modified Caprini score (>15) [Hewes 2015]	1	70	Very serious[a]	No serious inconsistency[b]	No serious indirectness[c]	No serious imprecision[d]	100 (100 – 100)	66.7 (55 – 78.3)	0.818 (0.7111 – 0.908)	LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.
(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias
(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).
(c) Indirectness was assessed using the PROBAST checklist items relating to applicability
(d) Imprecision was assessed according to the range of point estimates of the primary measure. The evidence was downgraded by 1 increment when there was a 20-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

People undergoing plastic surgery

Table 16: Clinical evidence profile: risk tools for predicting VTE in people undergoing plastic surgery

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality

 ISBN 978 – 1 – 4731 – 2871 - 2

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
Caprini score Cut-off ≥5	1	1598	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	Serious imprecision[d]	71 (49-87)	39 (37-42)	-	VERY LOW
Caprini score Cut-off ≥6	1	1598	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	Very serious imprecision[d]	58 (37-78)	60 (58-63)	-	VERY LOW
Caprini score Cut-off ≥9	1	1598	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	Serious imprecision[d]	17 (5-37)	93 (92-94)	-	VERY LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

(d) Imprecision was assessed according to the range of point estimates of the primary measure. The evidence was downgraded by 1 increment when there was a 20-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

People undergoing neurosurgery

Table 17: Clinical evidence profile: risk tools for predicting VTE in already known high-risk people undergoing neurosurgery

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
American College of Surgeons National Surgical Quality Improvement Programme: Universal Surgical Risk Calculator	1	1006	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	Not estimable	-	-	0.767	VERY LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

People undergoing urological surgery – robot-assisted partial nephrectomy

Table 18: Clinical evidence profile: risk tools for predicting VTE in already known high-risk people undergoing urological surgery – robot-assisted partial nephrectomy

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality

 ISBN 978 – 1 – 4731 – 2871 - 2

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
American College of Surgeons National Surgical Quality Improvement Programme: Universal Surgical Risk Calculator	1	300	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	Not estimable	-	-	0.670	VERY LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

High-risk patients undergoing emergency abdominal surgery or neurosurgery

Table 19: Clinical evidence profile: risk tools for predicting VTE in already known high-risk people undergoing emergency abdominal surgery or neurosurgery

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
Caprini score Cut-off ≥10.5	1	140	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	95 (83-99)	73 (64-82)	0.87 (0.811 – 0.93)	VERY LOW

ISBN 978 – 1 – 4731 – 2871 - 2

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

(d) Imprecision was assessed according to the range of point estimates of the primary measure. The evidence was downgraded by 1 increment when there was a 20-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

5.2.3.1.3 People with trauma

Table 20: Clinical evidence profile: risk tools for predicting VTE in people with trauma

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
TESS High risk <9	1	357	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	97 (91-99)	27 (22-32)	0.71 (0.65-0.77)	VERY LOW
TESS Risk cut off >5	1	234,032	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	77.4 (76-79)	75.6 (75-76)	0.84 (0.83-0.84)	VERY LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

c) Indirectness was assessed using the PROBAST checklist items relating to applicability

d) Imprecision was assessed according to the range of point estimates of the primary measure. The evidence was downgraded by 1 increment when there was a 20-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

5.2.3.1.4 *People with thermal injuries (burns)*

Table 21: Clinical evidence profile: risk tools for predicting VTE in thermally injured (burned) people

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
Risk Scoring Tool for Thermally Injured Patients	1	5761	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	Not estimable	-	-	0.750	VERY LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

(d) Imprecision was assessed according to the range of point estimates of the primary measure. The evidence was downgraded by 1 increment when there was a 20-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

5.2.3.2 DVT

5.2.3.2.1 *People with trauma*

Table 22: Clinical evidence profile: risk tools for predicting DVT in people with trauma

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality

ISBN 978 – 1 – 4731 – 2871 - 2

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
RAP Moderate risk cut-off 5 to ≤14	1	2281	Serious[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	82 (77-87)	57 (55-59)	-	LOW
RAP High risk cut-off >14	1	2281	Serious[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	15 (11-20)	97 (97-98)	-	LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity was this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

(d) Imprecision was assessed according to the range of point estimates of the primary measure. The evidence was downgraded by 1 increment when there was a 20-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

5.2.3.2.2 People who have had a stroke

Table 23: Clinical evidence profile: risk tools for predicting DVT in stroke patients

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
Post stroke DVT Prediction System	1	287	Very serious[a]	No serious inconsistency[b]	No serious indirectness[c]	No serious imprecision[d]	-	-	0.65 (0.59-0.70)	LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a

ISBN 978 – 1 – 4731 – 2871 - 2

test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

(d) Imprecision was assessed according to the range of point estimates of the primary measure. The evidence was downgraded by 1 increment when there was a 20- 40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

5.2.3.3 PE (fatal and non-fatal PE)

5.2.3.3.1 People with trauma

Table 24: Clinical evidence profile: risk tools for predicting fatal and non-fatal PE in trauma patients

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
TESS High risk <9	1	357	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	Serious imprecision[d]	97 (87-99)	24 (20-29)	0.67 (0.59-0.75)	VERY LOW
RAP Cut-off 5 to ≤14	1	2281	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	71 (55-86)	53 (51-56)	-	VERY LOW
RAP Cut-off >14	1	2281	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	12 (10-23)	96 (95-97)	-	VERY LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

(d) Imprecision was assessed according to the range of point estimates of the primary measure. The evidence was downgraded by 1 increment when there was a 20-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

ISBN 978 – 1 – 4731 – 2871 - 2

5.2.3.4 Fatal PE

5.2.3.4.1 People with trauma

Table 25: Clinical evidence profile: risk tools for predicting fatal PE in trauma patients

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
TESS High risk <9	1	357	Very serious[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	100 (81-100)	20 (13-28)	-	VERY LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

(d) Imprecision was assessed according to the range of point estimates of the primary measure. The evidence was downgraded by 1 increment when there was a 20-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

5.2.4 Calibration

5.2.4.1 VTE

5.2.4.1.1 Surgical patients

Mixed surgical patients

Table 26: Clinical evidence profile: risk tools for predicting VTE in mixed surgical patients

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	R^2 (95%CI)	Hosmer-Lemeshow test (p-value)	Brier score (95%CI)	D statistic	Quality

ISBN 978 – 1 – 5731 – 2871 - 2

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	R^2 (95%CI)	Hosmer-Lemeshow test (p-value)	Brier score (95%CI)	D statistic	Quality
Caprini score	2	13060	Very serious[a]	Serious indirectness[b]	Not estimable	-	0.607 0.609	-	-	VERY LOW

(a) Risk of bias was assessed using the PROBAST checklist.
(b) Downgraded by 1 increment as the definition of target condition does not match protocol

Colorectal surgery patients

Table 27: Clinical evidence profile: risk tools for predicting VTE in people undergoing colorectal surgery

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	R^2 (95%CI)	Hosmer-Lemeshow test (p-value)	Brier score (95%CI)	D statistic	Quality
ACS NSQIP Universal Surgical Risk Calculator	1	88,334	Very serious[a]	Serious indirectness[b]	Not estimable	-	-	0.0218	-	VERY LOW

(a) Risk of bias was assessed using the PROBAST checklist.
(b) Downgraded by 1 increment as the definition of target condition does not match protocol

People undergoing lung cancer resections

Table 28: Clinical evidence profile: risk tools for predicting VTE in people undergoing surgery for lung cancer

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	R^2 (95%CI)	Hosmer-Lemeshow test (p-value)	Brier score (95%CI)	D statistic	Quality
Caprini score High risk >5	1	232	Very serious[a]	No serious indirectness	Not estimable	-	0.61	-	-	LOW

(a) Risk of bias was assessed using the PROBAST checklist.

Oesophageal cancer surgery patients

Table 29: Clinical evidence profile: risk tools for predicting VTE in people undergoing oesophagectomy for oesophageal cancer

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	R^2 (95%CI)	Hosmer-Lemeshow test (p-value)	Brier score (95%CI)	D statistic	Quality
Caprini score High risk >5	1	70	Very serious[a]	No serious indirectness	Not estimable	-	10.282 (0.113)	-	-	LOW

(a) Risk of bias was assessed using the PROBAST checklist.

People undergoing urological surgery – robot-assisted partial nephrectomy

Table 30: Clinical evidence profile: risk tools for predicting VTE in people undergoing urological surgery – robot-assisted partial nephrectomy

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	R^2 (95%CI)	Hosmer-Lemeshow test (p-value)	Brier score (95%CI)	D statistic	Quality
ACS NSQIP Universal Surgical Risk Calculator	1	300	Very serious[a]	Serious indirectness[b]	Not estimable	-	-	0.003327	-	VERY LOW

(a) Risk of bias was assessed using the PROBAST checklist.
(b) Downgraded by 1 increment as the definition of target condition does not match protocol

5.2.4.1.2 People with trauma

Table 31: Clinical evidence profile: risk tools for predicting VTE in trauma patients

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	R^2 (95%CI)	Hosmer-Lemeshow test (p-value)	Brier score (95%CI)	D statistic	Quality

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	R^2 (95%CI)	Hosmer-Lemeshow test (p-value)	Brier score (95%CI)	D statistic	Quality
TESS	2	234,389	Very serious[a]	Serious indirectness[b]	Not estimable	-	0.101 13.70	-	-	VERY LOW

(a) Risk of bias was assessed using the PROBAST checklist.
(b) Downgraded by 1 increment as the definition of target condition does not match protocol

5.2.4.2 PE (non-fatal and fatal PE)

5.2.4.2.1 People with trauma

Table 32: Clinical evidence profile: risk tools for predicting non-fatal and fatal PE in trauma patients

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	R^2 (95%CI)	Hosmer-Lemeshow test (p-value)	Brier score (95%CI)	D statistic	Quality
TESS cut off <9	1	357	Serious[a]	No serious indirectness	Not estimable	-	13.7	-	-	MODERATE

(a) Risk of bias was assessed using the PROBAST checklist.

5.2.4.3 Fatal PE

5.2.4.3.1 People with trauma

Table 33: Clinical evidence profile: risk tools for predicting fatal PE in trauma patients

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	R2 (95%CI)	Hosmer-Lemeshow test (p-value)	Brier score (95%CI)	D statistic	Quality
TESS Cut-off <9	1	357	Serious[a]	Serious indirectness[b]	Not estimable	-	13.7	-	-	LOW

(a) Risk of bias was assessed using the PROBAST checklist.

(b) Downgraded by 1 increment as the definition of target condition does not match protocol

 ISBN 978 – 1 – 4731 – 2871 - 2

5.2.5 Economic evidence

Published literature

No relevant health economic studies were identified.

See also the economic article selection flow chart in appendix F.

5.2.6 Evidence statements

Clinical

General medical patients

Evidence was available for seven tools that assessed VTE risk in general medical patients. Very low quality evidence from one study (n=63,548) that explored the predictive ability of the Caprini risk assessment model at three separate cut off points (5, 7 and 9) showed sensitivities at all thresholds did not reach the committee's pre-specified threshold for decision-making (80%). No c-statistic data was available for the Caprini RAM. Moderate quality evidence from one study (n=1478) showed that the Geneva Risk Score might be sensitive enough for consideration (90%) however the variance around this estimate dipped below the committee's decision-making threshold (95% CI 73.5-97.9) and the accompanying specificity (0.353 [0.328-0.378]) was much lower than the committee's decision-making threshold (60%). Low quality evidence from one study (n=63,548) showed that the predictive version of IMPROVE offered poor discrimination (c-statistic 0.570 [0.565-0.576]) with no corresponding sensitivity and specificity data reported. Very low quality evidence from two studies (n=110,404) using the Intermountain risk tool suggested that discrimination ranged from poor to moderate with reported c-statistics of 0.611 (0.605-0.618) and 0.843 (0.833-0.852), but no associated sensitivity and specificity data was reported in either study. Low quality evidence from two studies (110,404) suggested that the Kucher tool also offered poor discrimination with c-statistics of 0.563 (0.558-0.568) and 0.683 (0.673-0.691). Very low quality evidence from three studies (n=66,106) suggested the using the Padua Prediction Score with a cut-off of ≥4 produced sensitivity (0.733 [0.541-0.877]) and specificity (0.519 [0.493-0.545]) that did not reach the committee's pre-specified decision-making threshold; and showed poor discrimination with c-statistics of 0.60 (0.59-0.61) and 0.58 (0.43-0.73). Finally very low quality evidence from one study (n=48,540) showed that an unnamed risk tool (Rothberg 2011) showed moderate discrimination (0.75 [0.71-0.78]). A further eighth study was identified in the specific subgroup of hospitalised cancer patients. Low quality evidence from this study (n=2780) showed a sensitivity of 19% (12-28) and specificity of 87% (86-88) when using a high-risk cut-off of ≥3 to predict VTE.

One study (n=287) conducted with people who had had a stroke, provided low quality evidence that a Post-Stroke DVT Prediction System had moderate discrimination (c-stat 0.65 [0.59-0.70]) ability for predicting DVT in this particular population.

Surgical and trauma patients (including people with burn injuries)

Very low quality evidence from two studies (n=13,060) showed poor discrimination (c-statistics 0.585 and 0.698) for the Caprini RAM for predicting VTE in mixed surgical patients (Hosmer-Lemeshow test p values 0.607 and 0.609); and low quality evidence from one study (n=3,576) showed moderate discrimination for an unnamed risk model (Pannucci 2014) in a similar mixed surgical population. Very low quality evidence from one study (n=88,334) showed that the American College of Surgeons (ACS) National Surgical Quality Improvement Programme (NSQIP): Universal Surgical Risk Calculator showed moderate discrimination (0.7203) for predicting VTE in colorectal surgery patients (Brier score 0.0218). Low quality evidence from one study (n=232) looking at the Caprini RAM for predicting

VTE in people undergoing lung cancer resections showed moderate discrimination (0.72 and 0.73). At the lower cut-off points of 5 (H-L test p-value 0.61)and 7 the reported sensitivities were 100% however the associated specificities were well below the committee's pre-specified threshold for decision making (0.072 [0.041-0.11]; 0.314 [0.25-0.373]). At a cut-off of 9 the sensitivity and specificity estimates met the committee's thresholds (0.833 and 0.605) but the imprecision around these estimates fell below each of the decision-making thresholds. At a cut-off of 10 the primary measure for decision-making (sensitivity) did not meet the committee's threshold (0.75 [0.50-1.00]). Low quality evidence from one small study (n=70) showed moderate discrimination when using the modified Caprini RAM to predict VTE in oesophageal cancer surgery patients (c-statistic 0.818 [0.711-0.908]; H-L test [p-value]: 10.282 [0.113]). At a cut-off of >15 low quality evidence for this risk tool suggested 100% sensitivity and 66.7% specificity but the imprecision around the specificity measure dipped below the committee's pre-specified threshold for decision making (0.55-0.78). When using the Caprini RAM to predict VTE in people undergoing plastic surgery, very low quality evidence from one study (n=1598) showed no sensitivities that met the committee's pre-specified threshold when looking at multiple cut-offs (5, 6 and 9). Two studies explored the use of the ACS NQIP: universal surgical risk calculator for predicting VTE in patients undergoing neurosurgery (n=1006) and urological surgery (n=300). In both cases very low quality evidence was provided for the c-statistic only with no associated variance data. The c-statistic was showed moderate discrimination for the tool in the neurosurgical population (0.767) and poor discrimination in the urological surgery population (0.670; Brier score 0.003327). When looking at people already recognised at high-risk for VTE undergoing emergency abdominal or neurosurgery, low quality evidence from one study (n=140) showed moderate discrimination for the Caprini RAM (0.87 [0.81-0.93]) and sensitivity of 95% (83-99) and specificity of 73% (0.64-0.82) for predicting VTE at a cut-off of ≥10.5. Very low quality evidence from two studies suggested TESS showed moderate discrimination at predicting VTE in people with trauma (n=357, c-statistic 0.71 [0.65-0.77]; n=234032, c-stat 0.84 [0.83-0.84]). The smaller study reported sensitivity of 97% (91-99) and specificity of 27% (22-32) when using a cut-off of <9. The larger study reported sensitivity of 77% (76-79) and specificity of 76% (75-76) when using a cut-off of >5. One study (n=5761) provided very low quality evidence that a risk scoring tool for thermal injured patients showed moderate discrimination (0.750 [no CI reported]) for predicting VTE in people with burn injuries.

Low quality evidence from one study (n=2281) looked at RAP at two different thresholds for predicting DVT in people with trauma. The cut off of ≤14 showed sensitivity of 82% (77-87) and specificity of 57% (55-59). The cut-off of >14 showed sensitivity of 15% (11-20) and specificity of 97% (97-98). Very low quality evidence from this same study also reported the ability of RAP to predict PE and fatal PE. The cut off of ≤14 showed sensitivity of 71% (55-86) and specificity of 53% (51-56). The cut-off of >14 showed sensitivity of 12% (10-23) and specificity of 96% (95-97). Another study (n=357) provided very low quality evidence for the poor discrimination (0.67 [0.59-0.75]) of TESS at predicting the combination of PE and fatal PE in trauma patients. This study reported sensitivity of 97% (87-99) and specificity of 24% (20-29) for TESS at a cut-off of <9. When focusing specifically on fatal PE only, very low quality evidence showed sensitivity of 100% (81-100) and specificity of 20% (13-28) for TESS at a cut-off of <9.

Economic

No relevant economic evaluations were identified.

 ISBN 978 – 1 – 4731 – 2871 - 2

5.3 Accuracy of risk assessment tools for bleeding in hospital admissions

5.3.1 Review question: What is the accuracy of individual risk assessment or prediction tools in predicting the likelihood of major bleeding or the risk of bleeding in a patient who is admitted to hospital?

For full details see review protocol in appendix C.

Table 34: PICO characteristics of review question

Question	What is the accuracy of individual risk assessment or predication tools in predicting the likelihood of major bleeding or the risk of bleeding in a patient who is admitted to hospital?
Population	Adults and young people (aged 16 or over) admitted to hospital
Risk tool	Derived and (externally or temporally) validated risk tools identified in literature
Target condition(s)	Major bleeding (up to 90 days)
Outcomes (in terms of predictive test accuracy, calibration)	Statistical outputs may include: • Discrimination (sensitivity, specificity, predictive values) • Area under the ROC curve (c-statistic) • Predicted risk versus observed risk (calibration) • Reclassification • Other statistical measures: for example, D statistic, R^2 statistic and Brier score
Study types	Prospective and retrospective cohort Exclusions: derivation studies

5.3.2 Clinical evidence

One study evaluating the IMPROVE bleeding risk score was included in the review. [80] This is summarised in Table 35 below. See also the study selection flow chart in appendix E, study evidence tables in appendix H, and excluded studies list in appendix N.

Table 35: Summary of studies included in the review

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
Hostler 2016 [80]	IMPROVE bleeding risk score	n=1668 Adults admitted for a medical illness. Age: <40: 234 (14%), 40-84: 1144 (68.6%), ≥85: 289 (17.3%) Gender (male to female ratio): 969:699 Ethnicity: not reported USA	Major bleeding at 14 days[a] Clinically relevant non-major bleeding at 14 days Based on UCD-9 codes and a haematocrit drop >6 points to identify patients who may have bled during admission. All bleeding events were confirmed by manual chart audit.	31 14	Prospective data collection with retrospective analysis.

(a) Raw data for 2x2 tables and calculation of sensitivity and specificity provided through author correspondence.

 ISBN 978 – 1 – 4731 – 2871 - 2

5.3.3 Discrimination

5.3.3.1 Major bleeding

Table 36: Clinical evidence profile: risk tools for predicting major bleeding in patients admitted to hospital

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	Area under the curve	Quality
IMPROVE bleeding risk score										
Major bleeding at 14 days	1	1668	Very serious[a]	-	No serious indirectness	No serious imprecision[b]	48 (27, 69)	78 (76, 81)	0.67 (0.57-0.77)	LOW
Major bleeding during hospitalisation	1	1668	Very serious[a]	-	No serious indirectness	No serious imprecision[b]	48 (30, 67)	78 (76, 81)	-	LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making. Where sensitivity was not reported specificity values were used for the assessment, if these were not available the assessment was based on the C-statistic. The committee set the following thresholds as an acceptable level to recommend a test: sensitivity 80%; specificity 60%.

(a) Risk of bias was assessed using the PROBAST checklist.

(b) Imprecision was assessed according to the range of point estimates of the primary decision measure (specificity). The evidence was downgraded by 1 increment when there was a 40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

 ISBN 978 – 1 – 4731 – 2871 - 2

5.3.4 Calibration

No calibration data reported.

5.3.5 Economic evidence

Published literature

No relevant health economic studies were identified.

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

5.3.6 Evidence statements

Clinical

Low quality evidence from one study (n=1668) suggested that calculating the IMPROVE bleeding risk score at admission was a poor predictor of major bleeding in medical inpatients (AUC 0.67 [95% CI 0.57-0.77]). The sensitivity of the IMPROVE bleeding risk score (0.48 [0.27-0.69]), the primary outcome for decision making, did not reach the committee's pre-specified thresholds (80%).

Economic

No relevant economic evaluations were identified.

5.4 Effectiveness of risk assessment tools in hospital admissions

5.4.1 Review question: How clinically and cost effective are risk assessment tools at reducing the rate of VTE in patients who are admitted to hospital?

For full details see review protocol in appendix C.

Table 37: PICO characteristics of review question

Population	Adults (aged 16 or over) admitted to hospital
Intervention(s)	Intervention: Derived and validated risk tool for predicting the risk of VTE/DVT/PE/major bleeding The Department of Health risk tool (not validated)
Comparison(s)	No risk tool, other risk tools
Outcomes	**Critical:** • All-cause mortality (up to 90 days from hospital discharge) • VTE (symptomatic or asymptomatic) (up to 90 days from hospital discharge) • DVT (symptomatic or asymptomatic) (up to 90 days from hospital discharge) • Pulmonary embolism (up to 90 days from hospital discharge) • Fatal pulmonary embolism (up to 90 days from hospital discharge) • Major bleeding (up to 90 days from hospital discharge) • Quality of life (validated scores) (up to 90 days from hospital discharge) **Important:** • Fatal bleeding (up to 90 days from hospital discharge) • Clinically relevant non-major bleeding (up to 45 days from hospital discharge)

 ISBN 978 – 1 – 4731 – 2871 - 2

	• Heparin-induced thrombocytopenia (up to 90 days from hospital discharge) • Hospital length of stay (up to 90 days from hospital discharge) • Unplanned readmission (up to 90 days from hospital discharge) • Haemorrhagic stroke (up to 90 days from hospital discharge)
Study design	Systematic reviews of RCTs or RCTs. If no RCTs are identified, observational studies (including before and after studies) will be considered

5.4.2 Clinical evidence

As no randomised controlled trials were identified, observational studies were considered for inclusion in this review. Five studies were included in the review ; one retrospective cohort study[102], one prospective cohort study[59], and three before-and-after studies[24, 25, 162]; these are summarised in Table 38 below.

Three studies [24, 25,59] compared use of a risk tool with no risk tool (Department of Health risk tool, Caprini risk tool and the Padua prediction score). Two studies [102, 162] compared achieving the quality standard of 90% of admissions being assessed with the Department of Health risk tool with not achieving the quality standard.

Evidence from these studies is summarised in the clinical evidence summary below (Table 38). See also the study selection flow chart in appendix E, forest plots in appendix L, study evidence tables in appendix H, GRADE tables in appendix K and excluded studies list in appendix N.

Table 38: Summary of studies included in the review: studies comparing use of risk tool versus no risk tool

Study	Intervention and comparison	Population	Outcomes
Cassidy 2014 [24]	Before and after study Before: Before development of the standardised program, no VTE prevention guidelines were formally used (2009). Surgeons generally acknowledged the American College of Chest Physicians guidelines, but no structured system existed and no individualised risk stratification was performed. There were no electronic reminders about VTE prophylaxis, and no surgeons used the Caprini system to guide decisions After: Post-implementation (July 2011-June 2012). Electronic order system is customised to require that a Caprini score be calculated for every patient at the time of operation and/or admission within general surgery and vascular surgery standardised order sets. Standardised VTE prophylaxis regimens were	Before implementation n=1,569 After implementation n=1,323 People undergoing general or vascular surgery, including people admitted to an ICU Age: Not reported Gender (male to female ratio): Not reported USA	DVT (30 days):new diagnosis of venous thrombosis, confirmed by imaging study or autopsy, which is treated with anticoagulation or placement of vena cava filter PE (30 days): new diagnosis of a new blood clot in a pulmonary artery, which is confirmed by imaging or autopsy.

 ISBN 978 – 1 – 4731 – 2871 - 2

Study	Intervention and comparison	Population	Outcomes
	created and linked to Caprini risk categories, the surgeon may decline VTE prophylaxis when it is contrary to his or her judgement by choosing the "opt out" selection in the order sets. Mobilisation program was also implemented, encouraging mobilisation of patients.		
Catterick 2014 [25]	Before and after study Before: 1 year before the implementation of Department of Health risk tool (2009) After: Two years after the implementation of Department of Health risk tool (2010/11)	n= not reported All people admitted to NHS hospitals in England. Age: Not reported Gender (male to female ratio): Not reported UK	VTE-related mortality (90 days) VTE-related readmission (30 days) VTE-related readmission (90 days) VTE: defined using ICD-10 codes used by the UK All Party Parliamentary Thrombosis Group. PE defined as I26.0 and I26.9. DVT defined as I80.1, I80.2, I80.3, I80.9 and I82.9
Germini 2016 [59]	Prospective cohort (quasi RCT) Intervention: Those admitted to Internal Medicine section 1 allocated to Padua prediction score decision strategy. Comparison: Those admitted to Internal Medicine section 2 allocated to clinical judgment-based strategy.	n = 628 All hospitalised acutely ill medical patients admitted into one of two Internal Medicine sections at the University Hospital in Perugia. Age: Range of medians 72-75 years Gender (male to female ratio): 340/288 Italy	DVT: defined with complete compression ultrasonography. PE: defined with CT angiography or V/Q lung scanning Fatal PE All-cause mortality Major bleeding: not defined.

Table 39: Summary of studies included in the review: studies comparing achievement of >90% of admissions assessed using risk tool with <90%

Study	Intervention and comparison	Population	Outcomes
Lester 2013 [102]	Retrospective cohort study Intervention: Use of Department of Health risk tool from July 2010 in achieving <90% VTE risk assessment Comparison:	n=17,712,681 All people admitted to 163 NHS hospitals in England (including general medical and surgical patients). Age: Not reported	VTE-related mortality post-discharge (90 days): death anywhere within the first three positions where VTE is considered either the direct cause or a contributing cause of death. Primary VTE-related mortality post-discharge (90

 ISBN 978 – 1 – 4731 – 2871 - 2

Study	Intervention and comparison	Population	Outcomes
	Use of Department of Health risk tool in March 2012 in achieving ≥90% VTE risk assessment	Gender (male to female ratio): Not reported UK	days): VTE code was listed in the first position of the death certificate, thus was considered the direct cause of death. VTE: defined using ICD10 codes - specified by the NHS-Outcome Framework 2013/14: I260, I269, I800, I801, I802, I803, I808, I809, I821, I822, I823, I829, O082, O223, O229, O870, O871, O879, O882
Roberts 2013 [162]	Before and after study Before: Department of Health risk tool (April 2010-March 2011). After: Department of Health risk tool (April 2011-March 2012) use to achieve sustained improvement in risk assessment on the incidence of VTE and the proportion of events attributable to inadequate prophylaxis The cut-point for comparison was delayed for 3 months following achievement of 90% risk assessment to account for potential lag in outcome improvement and the definition of VTE, including events occurring up to 90 days post-discharge.	n=302,057 All patients admitted to one hospital. Age: Not reported Gender (male to female ratio): Not reported UK	VTE (90 days): any new episode of VTE, diagnosed during hospitalisation or within 90 days of discharge following an inpatient stay of at least 2 days, or a surgical procedure under general or regional anaesthesia. Identified from screening radiology reports of CT pulmonary angiogram, ventilation/perfusion scans, upper and lower limb venous compression ultrasound, primary or secondary discharge diagnoses of VTE identified from ICD10 codes I80.0-80.9, I26.0-26.9 or O22.2, O22.3, O87.0 or O87.1, post-mortem reports, and death certificates with VTE listed as a primary cause of death PE (90 days): definition not reported. DVT (90 days): definition not reported.

 ISBN 978 – 1 – 4731 – 2871 - 2

5.4.3 General medical points

5.4.3.1 Department of Health risk tool versus no risk tool

Table 40: Clinical evidence summary: Department of Health risk tool versus no risk tool for general medical patients

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with No Department of Health risk tool	Risk difference with Department of Health risk tool (95% CI)
Mortality, VTE-related	100000 (1 study) 90 days	VERY LOW[a,b] due to risk of bias, imprecision	Rate ratio 0.92 (0.39 to 2.15)	0 per 1000 [c]	0 fewer per 1000 [c] (from 0 fewer to 0 more)
Readmission, VTE-related	100000 (1 study) 30 days	VERY LOW[a] due to risk of bias	Rate ratio 0.99 (0.82 to 1.19)	1 per 1000 [c]	0 fewer per 1000 [c] (from 0 fewer to 0 more)
Readmission, VTE-related	100000 (1 study) 90 days	VERY LOW[a] due to risk of bias	Rate ratio 1.02 (0.88 to 1.19)	2 per 1000 [c]	0 fewer per 1000 [c] (from 0 fewer to 0 more)

a Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs.

c – Anticipated absolute effects could not be calculated accurately as only rate ratio was reported

 ISBN 978 – 1 – 4731 – 2871 - 2

5.4.3.2 Department of Health risk tool: achieving >90% of admissions assessed using Department of Health risk tool versus achieving <90% assessed using risk tool

Table 41: Clinical evidence summary: Department of Health risk tool: achieving >90% of admissions assessed using Department of Health risk tool versus achieving <90% assessed using risk tool for general medical patients

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with No Department of Health risk tool	Risk difference with Department of Health risk tool (95% CI)
Mortality, VTE-related post-discharge – length of stay >3 days	2 590 547 (1 study) 90 days	VERY LOW[a] due to risk of bias	RR 0.96 (0.81 to 1.14)	-	- [c]
Mortality, VTE-related post-discharge - length of stay <4 days	10 719 502 (1 study) 90 days	VERY LOW[a,b] due to risk of bias, imprecision	RR 0.74 (0.6 to 0.92)	-	- [c]
Mortality, primary VTE-related post-discharge - length of stay >3 days	2 590 547 (1 study) 90 days	VERY LOW[a,b] due to risk of bias, imprecision	RR 0.89 (0.71 to 1.1)	-	- [c]
Mortality, primary VTE-related post-discharge - length of stay <4 days	10 719 502 (1 study) 90 days	VERY LOW[a,b] due to risk of bias, imprecision	RR 0.62 (0.47 to 0.81)	-	- [c]
DVT	302057 (1 study) 90 days	VERY LOW[a,b] due to risk of bias, imprecision	RR 0.95 (0.83 to 1.09)	3 per 1000	0 fewer per 1000 (from 0 fewer to 0 fewer)
PE	302057 (1 study) 90 days	VERY LOW[a,b] due to risk of bias, imprecision	RR 0.79 (0.67 to 0.94)	11 per 1000	2 fewer per 1000 (from 1 fewer to 4 fewer)
VTE	302057 (1 study) 90 days	VERY LOW[a,b] due to risk of bias, imprecision	RR 0.88 (0.79 to 0.98)	1 per 1000	0 fewer per 1000 (from 0 fewer to 0 fewer)

a Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
b Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs.

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with No Department of Health risk tool	Risk difference with Department of Health risk tool (95% CI)
c - Could not be calculated as control group risk was not reported appropriately					

5.4.3.3 Padua prediction score versus no risk tool

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Padua prediction score versus no risk tool (95% CI)
DVT	628 (1 study)	VERY LOW[a,b] due to risk of bias, imprecision	RR 0.55 (0.34 to 0.88)	155 per 1000	70 fewer per 1000 (from 19 fewer to 102 fewer)
PE	628 (1 study)	VERY LOW[a,b] due to risk of bias, imprecision	Peto OR 14.47 (0.25 to 830.93)	0 per 1000	- c
Fatal PE	628 (1 study)	VERY LOW[a,b] due to risk of bias, imprecision	Peto OR 14.47 (0.25 to 830.93)	0 per 1000	- c
Major bleeding	628 (1 study)	VERY LOW[a,b] due to risk of bias, imprecision	OR 0.2 (0.01 to 3.55)	5 per 1000	4 fewer per 1000 (from 5 fewer to 13 more)
All cause mortality	628 (1 study)	VERY LOW[a,b] due to risk of bias, imprecision	RR 1.11 (0.32 to 3.91)	15 per 1000	2 more per 1000 (from 10 fewer to 44 more)
a Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias					

 ISBN 978 – 1 – 4731 – 2871 - 2

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Padua prediction score versus no risk tool (95% CI)
b Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. c Absolute effects could not be calculated due to zero events in control arm					

5.4.4 Surgical patients

5.4.4.1 Caprini risk tool versus no risk tool

Table 42: Caprini risk tool versus no risk tool for surgical patients

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with No Caprini risk tool	Risk difference with Caprini risk tool (95% CI)
DVT	2892 (1 study) 30 days	VERY LOW[a,b] due to risk of bias, indirectness,	RR 0.11 (0.04 to 0.32)	23 per 1000	20 fewer per 1000 (from 15 fewer to 22 fewer)
PE	2892 (1 study) 30 days	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	RR 0.49 (0.2 to 1.17)	11 per 1000	6 fewer per 1000 (from 9 fewer to 2 more)
a Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias b Downgraded by 1 increment as the study was conducted in the USA, there are differences in clinical practice c Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs.					

 ISBN 978 – 1 – 4731 – 2871 - 2

5.4.4.2 Department of Health risk tool: achieving >90% of admissions assessed using Department of Health risk tool versus achieving <90% assessed using risk tool

Table 43: Department of Health risk tool: achieving >90% of admissions assessed using Department of Health risk tool versus achieving <90% using risk tool for surgical patients

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with No Department of Health risk tool	Risk difference with Department of Health risk tool (95% CI)
Mortality, VTE-related post-discharge - length of stay >3 days	1 550 794 (1 study) 90 days	VERY LOW[a,b] due to risk of bias, imprecision	RR 0.73 (0.46 to 1.16)	-	- c
Mortality, VTE-related post-discharge- length of stay <4 days	2 851 838 (1 study) 90 days	VERY LOW[a,b] due to risk of bias, imprecision	RR 0.82 (0.65 to 1.03)	-	- c
Mortality, primary VTE-related post-discharge- length of stay >3 days	1 550 794 (1 study) 90 days	VERY LOW[a,b] due to risk of bias, imprecision	RR 0.62 (0.44 to 0.89)	-	- c
Mortality, primary VTE-related post-discharge - length of stay <4 days	2 851 838 (1 study) 90 days	VERY LOW[a,b] due to risk of bias, imprecision	RR 0.57 (0.3 to 1.06)	-	- c

a Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
b Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs.
c - Could not be calculated as control group risk was not reported appropriately

 ISBN 978 – 1 – 4731 – 2871 - 2

5.4.5 Economic evidence

Published literature

Two health economic studies were identified with the relevant comparisons and have been included in this review.[99, 115] These are summarised in the health economic evidence profiles below (Table 44 and Table 45) and the health economic evidence table in appendix J.

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

New economic analysis

A cost impact analysis was also undertaken to aid the committee's decision making. In this analysis, with support from committee members, the speciality codes for general medical patients were identified. Using NHS Digital, Hospital Episode Statistics (HES) for 2015/16, the number of bed days for people who stayed in hospital as general medical patients for more than 3 days was identified (18.8 million).

The committee members advised that the National risk assessment tool used currently results in 80% of people having pharmacological prophylaxis. It is anticipated that the IMPROVE risk assessment tool would result in around 40% of people having prophylaxis; in line with the intermediate eligibility group in the Miller study.[115] The cost of prophylaxis per bed day is £3.03. The difference in the number of bed days at 80% and 40% prophylaxis was multiplied by the cost per day. This was then adjusted for an increase in costs due to increased cases of DVT and PE using Millar 2016. [115] The net saving from this reduction in prophylaxis is estimated to be around £22.3 million.

 ISBN 978 – 1 – 4731 – 2871 - 2

Table 44: Health economic evidence profile: Risk assessment tools vs no risk assessment tool

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost-effectiveness	Uncertainty
Lecumberri 2011[99] [Spain]	Partially applicable [(a)]	Potentially serious limitations[(b)]	**-Population:** All hospitalised adult inpatients (medical and surgical) at the University Clinic of Navarra. The population also included pregnant women but very small percentage ranging between 3.2 to 4.4% across the follow-up periods. **-Study design:** cost-consequences analysis based on a before-and-after cohort study. **-Interventions:** **Intervention 1:** No e-alert system to stratify patients' risk of thrombosis. **Intervention 2:** E-alert software to identify hospitalised patients at risk of VTE. The risk assessment scoring systems used were: PRETEMED scale (a validated risk stratification tool) for medical patients) and ACCP guidelines for surgical patients.	2 vs 1 Saves £6 per patient	2 vs 1: **VTE events:** 1 to 2 fewer VTE events per 1000 patients **Major bleeding:** 10 fewer major bleeding events per 1000 patients	Using risk assessment tools is dominant	None of the sensitivity analyses results in a change of the conclusion regarding dominance of the intervention.

Abbreviations: VTE: venous thromboembolism

(a) The risk assessment tools used are different from those included in the clinical review. QALYs are not used as measure of outcome. Uncertainty regarding the applicability of costs and resource use from the Spanish health care system in 2011 to current NHS perspective.

(b) The economic analysis is conducted alongside a single observational study, so by definition does not reflect all evidence in this area. Short follow-up period, so long terms and consequences have not been included. Unit costs are based on local rather than national sources; hence it is not clear if these are generalisable.

Table 45: Health economic evidence profile: prophylaxis based on risk stratification using individual risk factors vs no prophylaxis

Study	Applicability	Limitations	Other comments	Cost	Effects	Incremental costs	Incremental effects	Cost effectiveness	Uncertainty
Millar 2016 [115] ([Australia])	Partially applicable[(a)]	Potentially serious limitations [(b)]	• **Study design:** Cost consequences analysis using Decision tree model based on the results of a single RCT (the PREVENT trial) • **Population**: adult internal medicine patients admitted to all Australian hospitals • **Interventions:** 1. No prophylaxis 2. VTE prophylaxis using LMWH (Enoxaparin 40 mg/day). Three levels of eligibility for prophylaxis were examined: 2.a. restricted[(d)] (25% of all admissions), 2.b. intermediate[(c)] (40% of all admissions) and 2.c. broad[(e)] (80% of all admissions)	1. £29	1. 4.3 DVTs, 2.3 PEs, 0.4 deaths per 1000	DVT: No prophylaxis: dominated Restricted eligibility: baseline Intermediate eligibility: extendedly dominated Broad eligibility: £29,861 per DVT averted PE: No prophylaxis: dominated Restricted eligibility: baseline Intermediate eligibility: extendedly dominated Broad eligibility: £170,827 per PE averted Deaths: No prophylaxis: £30,000 per death averted Restricted eligibility: baseline Intermediate eligibility: dominated Broad eligibility: dominated			A range of sensitivity analyses were conducted including changing baseline VTE risk, fatality rate for PE and major bleeding and assumptions regarding VTE risk in non-eligible patients.
				2.a. £26	2.a. 2.5 DVTs, 2 PEs, 0.5 deaths per 1000				
				2.b. £30	2.b. 2.4 DVTs, 1.99 PE, 0.6 deaths				
				2.c. £39	2.c. 2.1 DVTs, 1.93 PEs, 0.9 deaths per 1000				

Abbreviations: DVT: deep vein thrombosis; ICER: incremental cost-effectiveness ratio; LMWH: low molecular weight heparin; PE: pulmonary embolism; QALY: quality-adjusted life years; RCT: randomised controlled trial; VTE: venous thromboembolism.

(a) Some uncertainty regarding the applicability of resource use and cost data from Australia in 2014 to current NHS context. Discounting was used only for health outcomes and the rate used is different from that recommended in the NICE Reference Case. QALYs are not used as an outcome measure.

(b) The model has a short time horizon that covers only the duration of the hospital stay, hence, does not capture long term costs. Only symptomatic events are included in the model. The source of baseline risk and relative treatment effects is based on a single trial and is not reflective of the total body of evidence. The results of the costs and outcomes are not presented as means per patient.

(c) Restricted: where only patients with strongest risk factors were given prophylaxis (malignancy, especially with chemotherapy, previous history of VTE, some rarer high risk conditions such as inflammatory bowel disease. (~ 25% of all inpatient admissions)

(d) Intermediate: where patients with strong and moderate risk factors, such as cardiac or respiratory failure, sepsis or inflammation, are given prophylaxis (~ 40% of all inpatient admissions)

(e) Broad: where everyone from the intermediate group as well as those satisfying an age criterion (>40 or >60) are given prophylaxis (~80% of all inpatient admissions)

 ISBN 978 – 1 – 4731 – 2871 - 2

5.4.6 Evidence statements

Clinical

For assessing VTE risk in general medical patients, very low quality evidence from one large study (n=100,000) showed no clinical difference in mortality, or 30 and 90 day readmission rates when the Department of Health risk tool was used compared to no risk tool being used. When the quality standard of assessment of 90% of admissions with the Department of Health risk tool had been achieved, very low quality evidence from another large study (n=10,719,502) suggested a clinical benefit for possible VTE-related, and primary VTE-related, mortality post-discharge following a hospital stay of less than 4 days. However the uncertainty around these effects means the estimates could also be consistent with no difference. No clinical difference was found between the ≥90% and <90% DOH assessed groups for the same mortality outcomes in patients whose hospital stay was longer than 3 days, and for VTE, DVT and PE. When general medical patients were risk assessed with the Padua prediction score, very low quality evidence from one study (n=628) suggested a possible clinical benefit for all-cause mortality, DVT and major bleeding, compared to those assessed with clinical-judgment only (no risk tool), although there was large uncertainty around all these estimates.

For assessing VTE risk in surgical patients, very low quality evidence from one study (n=2892) showed a clinically important reduction in DVT when assessing surgical patients with the Caprini risk tool compared to no risk tool. Very low quality evidence from the same study also suggested a lower PE rate in those assessed with the Caprini risk tool; however uncertainty around the PE estimate is also consistent with no difference. When the quality standard of assessment of 90% of admissions with the Department of Health risk tool had been achieved, very low quality evidence from another large study (n=1,550,794) suggested a clinical benefit for possible VTE-related, and primary VTE-related, mortality post-discharge following a hospital stay of more than 3 days, and primary VTE-related, mortality post-discharge following a hospital stay of less than 4 days. However the uncertainty around these effects means the estimates could also be consistent with no difference.

Economic

- One cost-effectiveness analysis found that in people admitted to hospital risk assessment using PRETEMED scale (a validated risk stratification tool) for medical patients and ACCP guidelines for surgical patients was dominant (less costly and more effective) compared to no risk assessment. This study was assessed as partially applicable with potentially serious limitations.
- One cost-consequences analysis found that in adults admitted to internal medicine department restricting eligibility for prophylaxis to the top 25% based on risk assessment using individual risk factors was dominant (less costly and more effective) compared to no prophylaxis. This study was assessed as partially applicable with potentially serious limitations.

5.5 Risk assessment for people having day procedures

Accuracy of risk assessment tools for VTE for day procedures

5.5.1 Review question: What is the accuracy of individual risk assessment or prediction tools in predicting the likelihood of VTE in patients who are having day procedures (including surgery and chemotherapy) at hospital?

For full details see review protocol in appendix C.

 ISBN 978 – 1 – 4731 – 2871 - 2

Table 46: PICO characteristics of review question

Question	What is the accuracy of individual risk assessment or predication tools in predicting the likelihood of VTE in patients who are having day procedures (including surgery and chemotherapy) at hospital?
Population	Adults and young people (aged 16 or over) who are having day procedures (including surgery and chemotherapy)
Risk tool	Derived and validated risk tools identified in literature
Target condition(s)	• VTE (symptomatic or asymptomatic) (7- 90 days; up to 180 days for people having cancer treatment) • VTE-related mortality (7- 90 days; up to 180 days for people having cancer treatment) • DVT alone (7- 90 days; up to 180 days for people having cancer treatment) • PE alone (7- 90 days; up to 180 days for people having cancer treatment)
Outcomes (in terms of predictive test accuracy, calibration)	Statistical outputs may include: • Discrimination (sensitivity, specificity, predictive values) • Area under the ROC curve (c-statistic) • Predicted risk versus observed risk (calibration) • Reclassification • Other statistical measures: for example, D statistic, R^2 statistic and Brier score
Study types	Prospective and retrospective cohort Exclusions: derivation studies

5.5.2 Clinical evidence

Seven studies evaluating 2 risk tools were included in the review, [9, 17, 27, 91, 148, 186, 193] these are summarised in Table 47 below. See also the study selection flow chart in appendix E, study evidence tables in appendix H, and excluded studies list in appendix N. Full details of the tools included in this review are provided in the clinical evidence tables in appendix H.

Five of the papers explored the predictive ability of the Khorana Score in a range of cancer patients, one explored an unnamed risk tool for cancer patients and the seventh paper explored an unnamed risk tool for surgical outpatients.

Table 47: Summary of studies included in the review

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
People undergoing cancer treatment					
Ay 2010[9]	Khorana score	n=819 People with cancer undergoing chemotherapy, radiotherapy and/or surgery Primary site of cancer: Breast 17.1% Lung 15.3% Stomach 4.4% Colorectal 13.7%	VTE (180 days): no routine screening for VTE. When a patient developed symptoms of VTE, objective imaging methods were performed to confirm or exclude the diagnosis. Duplex sonography or venography were applied for diagnosis of deep vein thrombosis (DVT) and computerized	n= 61 (7.4%)	Prospective cohort

 ISBN 978 – 1 – 4731 – 2871 - 2

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
		Pancreas 5.7% Kidney 2.9% Prostate 13.7% Brain (high-grade glioma) 13.1% Lymphoma 11.8% Multiple myeloma 2.2% Austria	tomography or ventilation/perfusion lung scan for diagnosis of pulmonary embolism (PE) Sensitivity Specificity NPV PPV		
Bezan 2017 [17]	Unnamed risk stratification model	n=349 People with testicular germ cell tumours Seminoma 56.8% Non-seminoma 43.2% Stage IA-B 64.8% Stage IS 2.6% Stage II1-IIC 14.3% Stage IIIA-C 18.3% Switzerland	VTE (12 months): not defined C-statistic	n=18 (5.2%)	Retrospective cohort
Cella 2017 [27]	Khorana score	n=843 People with active cancer undergoing chemotherapy, endocrine therapy, radiotherapy, target therapy and/or surgery in combination or alone. Primary tumour site: Breast 37% Gastroenteropancreatic 30% Genito/urinary tract 13% Lung 4% Metastatic disease 55% Other 16.5% Italy and Germany	VTE (12 months)L defined by Doppler ultrasound and CT Sensitivity Specificity C-statistic	n=73 (8.6%)	Retrospective cohort
Khorana 2008 [91]	Khorana score	n=1365 People with cancer undergoing chemotherapy	VTE (timepoint unclear): not defined Sensitivity Specificity	n=28 (2.1%)	Prospective cohort

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
		Primary site of cancer: Breast 34.6% Lung 17.3% Lymphoma 13.5% Colorectal 11.9% Gynaecologic 10.40% Gastric and pancreatic 1.4% Age: <65 years 62.3%; ≥65 years 37.7% Gender (male to female ratio): 1:2 USA	NPV PPV C-statistic Hosmer-Lemeshow test		
van Es 2017 [186]	Khorana Score	n=876 Ambulatory cancer patients with solid tumours Age, mean (SD): 64 (11) years 56% male Tumour type Lung 26% Oesophagus 19% Colorectal 18% Pancreas 12% Breast 9% Prostate 5% Gastric 5% Ovarian 5% Bladder 1% The Netherlands, Italy, France and Mexico	VTE (6 months): objectively confirmed symptomatic PE and DVT C-statistic	n=53 (6.1%)	Prospective cohort
Wang 2017 [193]	Khorana Score	n=270 People with hepatocellular carcinoma (HCC) Age, mean (range): 58.5 (26-80)	VTE (time point not defined) based on radiographic examinations using compression ultrasound, contrast-enhanced CT, and pulmonary angiogram	n=16 (5.93%)	Retrospective cohort

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
		Gender (M/F): 50/220 HCC with Barcelona stage 0-A 42.6% Advanced HCC with Barcelona stage C or D 57.4% USA	Sensitivity Specificity		
People undergoing surgery					
Pannucci 2012[148]	Unnamed (Pannucci 2012)	n=85,730 Surgical outpatients Herniorrphaphy:33% Integument: 22% Liver, biliary system, and pancreas: 13% Musculoskeletal: 9.1% Arteries and veins: 6.4% Hindgut (small bowel, large bowel, rectum and anus): 4.7% Endocrine: 3% Genital system (male or female): 2% Foregut (stomach, including gastric bypass procedure): 1.6% Head and neck, oesophagus: 1.5% Urinary system: 1.2% Hemic and lymphatic system, mediastinum and diaphragm: 0.9% Miscellaneous peritoneal procedures: 0.9% Nervous system structures: 0.5% Respiratory and cardiovascular: 0.1% Age (derivation and validation cohort): < 40 years 18.5%; 40-59 years 45.5%; 60 years 36% Gender (male to female ratio) (derivation and	VTE (30 days): DVT and/or PE. DVT is considered to be a new thrombus within the venous system that is confirmed using an objective imaging method (e.g. duplex ultrasound or computed tomography scan). PE is defined as an obstructing thrombus within the pulmonary arterial system. PE requires confirmation using an objective imaging method (e.g. computed tomography scan or arteriogram) C-statistic Hosmer-Lemeshow test	DVT: n=87 (0.10%) PE: n=37 (0.043%)	Prospective cohort

 ISBN 978 – 1 – 4731 – 2871 - 2

Study	Risk tool	Population	Outcomes	No of events (%)	Study design
		validation cohort): 1:1.4 USA			

 ISBN 978 – 1 – 4731 – 2871 - 2

5.5.3 Discrimination

5.5.3.1 People undergoing surgery

Table 48: Clinical evidence profile: risk tools for predicting VTE in people undergoing surgical day procedures

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic	Quality
Unnamed (Pannucci 2012)	1	85,730	Serious risk of bias[a]	No serious inconsistency[b]	No serious indirectness[c]	No serious imprecision[d]	-	-	0.78 (0.72 - 0.84)	MODERATE

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

(d) Imprecision was assessed according to the range of point estimates of the primary measure. The evidence was downgraded by 1 increment when there was a 20%-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

5.5.3.2 People having cancer treatment

Table 49: Clinical evidence profile: risk tools for predicting VTE in people having cancer day treatment

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic median (range)	Quality
Khorana Score										

ISBN 978 – 1 – 4731 – 2871 - 2

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic median (range)	Quality
Khorana score (≥3) *Pooled estimate*	5	4173	Very serious risk of bias[a]	Serious inconsistency[b]	No serious indirectness[c]	Very serious imprecision[d]	15.99% (1-55)	95.80% (82-99)	0.583 (0.47-0.70)	VERY LOW
Unnamed tools										
Unnamed risk stratification model (Bezan 2017)	1	349	Very serious risk of bias[a]	No serious inconsistency[b]	Very serious indirectness[c]	Not estimable	-	-	0.84	LOW

The assessment of the evidence quality was conducted with emphasis on test specificity as this was the primary measure discussed in decision making.

(a) Risk of bias was assessed using the PROBAST checklist items relating to risk of bias

(b) Inconsistency was assessed by inspection of the sensitivity/specificity (based on the primary measure) forest plots, or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the committee (the threshold above which would be acceptable to recommend a test). For example, the committee might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by one increment if the individual studies varied across two areas (eg. 50-90% and 90-100%) and by two increments if the individual studies varied across three areas (eg. 0-50%, 50-90% and 90-100%).

(c) Indirectness was assessed using the PROBAST checklist items relating to applicability

(d) Imprecision was assessed according to the range of point estimates of the primary measure. The evidence was downgraded by 1 increment when there was a 20-40% range of the confidence interval around the point estimate, and downgraded by 2 increments when there was a range of >40%.

ISBN 978 – 1 – 4731 – 2871 - 2

5.5.4 Calibration

5.5.4.1 People undergoing surgery

Table 50: Clinical evidence profile: risk tools for predicting VTE in people undergoing surgical day procedures

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	R^2 (95%CI)	Hosmer-Lemeshow test (p-value)	Brier score (95%CI)	D statistic	Quality
Unnamed (Pannucci 2012)	1	85,730	Serious risk of bias[a]	No serious indirectness[b]	Not estimable	-	0.826	-	-	MODERATE

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision making.
(a) Risk of bias was assessed using the PROBAST checklist.
(b) Indirectness was assessed using the PROBAST checklist items relating to applicability

5.5.4.2 People having cancer treatment

5.5.5People having cancer treatment

Table 51: Clinical evidence profile: risk tools for predicting VTE in people having cancer day treatment

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	R^2 (95%CI)	Hosmer-Lemeshow test (p-value)	Brier score (95%CI)	D statistic	Quality
Khorana score Khorana 2008	1	1365	Serious risk of bias[a]	Serious indirectness[b]	Not estimable	-	0.15	-	-	LOW

The assessment of the evidence quality was conducted with emphasis on test specificity as this was the primary measure discussed in decision making.
(a) Risk of bias was assessed using the PROBAST checklist.
(b) Indirectness was assessed using the PROBAST checklist items relating to applicability

ISBN 978 – 1 – 4731 – 2871 - 2

5.5.6 Economic evidence

Published literature

No relevant health economic studies were identified.

See also the economic article selection flow chart in appendix F.

5.5.7 Evidence statements

Clinical

Moderate quality evidence from a single study (n=85,730) suggested moderate discrimination for an unnamed tool at predicting risk of VTE for people undergoing surgical day procedures with calibration data of 0.826. No further discrimination data was reported.

Very low quality evidence from a diagnostic meta-analysis of 5 papers (n=4173) showed sensitivity of 15.99% (1-55) and specificity of 95.80% (82-99) for the Khorana Score at predicting VTE based on a high-risk cut-off of ≥3. There was very serious uncertainty around the estimate for sensitivity. This sensitivity was far below the pre-specified threshold set by the committee. Three of the five papers presented c-statistics which ranged from 0.47 to 0.70 with a median poor discrimination of 0.583.

Economic

No relevant economic evaluations were identified.

5.6 Accuracy of risk assessment tools for bleeding for day procedures

5.6.1 Review question: What is the accuracy of individual risk assessment or prediction tools in predicting the likelihood of major bleeding or the risk of bleeding in patients who are having day procedures (including surgery and chemotherapy) at hospital?

For full details see review protocol in appendix C.

Table 52: PICO characteristics of review question

Question	What is the accuracy of individual risk assessment or predication tools in predicting the likelihood of major bleeding or the risk of bleeding in patients who are having day procedures (including surgery and chemotherapy) at hospital?
Population	Adults (aged 16 or over) who are having day procedures (including surgery and chemotherapy)
Risk tool	Derived and (externally or temporally) validated risk tools identified in literature
Target condition(s)	Major bleeding (up to 90 days)
Outcomes (in terms of predictive test accuracy, calibration)	Statistical outputs may include: • Discrimination (sensitivity, specificity, predictive values) • Area under the ROC curve (c-statistic) • Predicted risk versus observed risk (calibration) • Reclassification • Other statistical measures: for example, D statistic, R^2 statistic and Brier score
Study types	Prospective and retrospective cohort Exclusions: derivation studies

 ISBN 978 – 1 – 4731 – 2871 - 2

5.6.2 Clinical evidence

No studies evaluating risk tools for predicting major bleeding associated with VTE in people having day procedures were included in the review. See the study selection flow chart in appendix E, study evidence tables in appendix H, and excluded studies list in appendix N. Full details of the tools included in this review are provided in the clinical evidence tables in appendix H.

5.6.3 Discrimination

No relevant studies were identified.

5.6.4 Calibration

No relevant studies were identified.

5.6.5 Economic evidence

Published literature

No relevant health economic studies were identified.

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

5.6.6 Evidence statements

Clinical

No relevant studies were identified.

Economic

No relevant economic evaluations were identified.

5.7 Effectiveness of risk assessment tools for day procedures

5.7.1 Review question: How clinically and cost effective are risk assessment tools at reducing the rate of VTE in patients who are having day procedures (including surgery and chemotherapy) at hospital?

For full details see review protocol in appendix C.

Table 53: PICO characteristics of review question

Population	Adults (aged 16 or over) who are having day procedures (including surgery and chemotherapy)
Intervention(s)	Derived and validated risk tool for predicting the risk of VTE/DVT/PE/major bleeding The Department of Health risk tool (not validated)
Comparison(s)	No risk tool, other risk tools
Outcomes	Critical: • All-cause mortality (up to 90 days from hospital discharge) • VTE (symptomatic or asymptomatic) (up to 90 days from hospital discharge) • DVT (symptomatic or asymptomatic) (up to 90 days from hospital discharge)

 ISBN 978 – 1 – 4731 – 2871 - 2

	• Pulmonary embolism (7- 90 days from hospital discharge) • Fatal pulmonary embolism (up to 90 days from hospital discharge) • Major bleeding (up to 90 days from hospital discharge) • Quality of life (validated scores) (up to 90 days from hospital discharge) Important: • Fatal bleeding (up to 90 days from hospital discharge) • Clinically relevant non-major bleeding (up to 45 days from hospital discharge) • Heparin-induced thrombocytopenia (up to 90 days from hospital discharge) • Hospital length of stay (up to 90 days from hospital discharge) • Unplanned readmission (up to 90 days from hospital discharge) • Haemorrhagic stroke (up to 90 days from hospital discharge)
Study design	Systematic reviews of RCTs or RCTs. If no RCTs are identified, consider observational studies (including before and after studies)

5.7.2 Clinical evidence

No relevant clinical studies were identified that compared validated risk tools with other or no risk tools, which predicted the risk of VTE, DVT, PE or major bleeding in people having day procedures. See the study selection flow chart in appendix E and excluded studies list in appendix N.

5.7.3 Economic evidence

Published literature

No relevant health economic studies were identified.

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

5.7.4 Evidence statements

Clinical

No relevant clinical studies were identified.

Economic

No relevant economic evaluations were identified.

5.8 Recommendations and link to evidence

Recommendations	**Risk assessment** **1.1.1 Assess all patients to identify the risk of venous thromboembolism (VTE) and bleeding (see recommendations 1.1.2, 1.1.5, 1.1.9, 1.4.17 and 1.4.23)** **People admitted to hospital**

 ISBN 978 – 1 – 4731 – 2871 - 2

	Medical patients **1.1.2 Assess all medical patients to identify the risk of VTE and bleeding:** • **as soon as possible after admission to hospital or by the time of the first consultant review** • **using a tool published by a national UK body, professional network or peer-reviewed journal. The most commonly used risk assessment tool for medical patients is the Department of Health VTE risk assessment tool[mmm] (See Appendix T) . [2018]** **1.1.3 Balance the person's individual risk of VTE against their risk of bleeding when deciding whether to offer pharmacological thromboprophylaxis to medical patients. [2018]** **1.1.4 If using pharmacological VTE prophylaxis for medical patients, start it as soon as possible and within 14 hours of admission, unless otherwise stated in the population-specific recommendations (see chapters 9-13). [2018]**
Research recommendation	**1. What is the accuracy of individual risk assessment tools in predicting the risk of VTE and risk of bleeding in medical patients admitted to hospital?**
Relative values of different outcomes	<u>Predictive accuracy of VTE and bleeding risk tools</u> The committee was interested in the prognostic accuracy of risk assessment tools for medical patients admitted to hospital or who are in hospital having day procedures. A risk assessment tool would be used to identify people with an increased risk of VTE who would benefit from having VTE prophylaxis, or identify people with an increased risk of major bleeding in order to determine appropriate prophylaxis strategies, for example not giving pharmacological prophylaxis to people who are at a high risk of bleeding. The committee agreed that sensitivity was more important than specificity in medical patients because people who are at higher risk of VTE could be identified for potential VTE prophylaxis treatment (fewer false negatives). The committee set thresholds for the acceptability of a test; for the populations noted here, these were ≥80% sensitivity and ≥60% specificity. Some studies only reported a C-statistic. The committee acknowledged that this metric was important for comparing the overall accuracy of the tools, but in itself was unlikely to provide enough information on which to base a recommendation as it does not indicate the number of false positives and negatives of the tool. Therefore, the committee decided against recommending a tool without sensitivity and specificity data. <u>Clinical effectiveness of risk tools for reducing VTE</u> For the review of the clinical effectiveness of risk tools, the committee considered all-cause mortality, VTE (symptomatic or asymptomatic), DVT (symptomatic or asymptomatic), PE, fatal PE, major bleeding and quality of life as critical outcomes. The time points for these outcomes were up to 90 days from hospital discharge. The committee considered fatal bleeding, clinically relevant non-major bleeding, heparin-induced thrombocytopenia, hospital length of stay, unplanned readmission and haemorrhagic stroke as important outcomes. The time points for these outcomes were up to 90 days, apart from clinically relevant non-major bleeding up to 45 days

mmm Reproduced with the permission of the Department of Health and Social Care under the Open Government Licence.

 ISBN 978 – 1 – 4731 – 2871 - 2

	from hospital discharge. Please see section 4.4.3 in the methods chapter for further detail on prioritisation of the critical outcomes.
Quality of the clinical evidence	Predictive accuracy of VTE and bleeding risk tools Fourteen studies were identified looking at risk tools for predicting VTE in medical patients. Eight papers featured people admitted to hospital and six featured those having day procedures, all of whom were people coming into hospital to receive cancer treatment. One study was identified looking at a risk tool to predict the risk of major bleeding in hospitalised medical patients. PROBAST was used to assess the risk of bias. All these studies were at a high or very high risk of bias. Common reasons for this were papers only supplying retrospective validation, papers not reporting a clear definition or method of confirmation for the target condition (VTE, DVT, PE or major bleeding), papers not reporting the time-point for the target condition measurement, or unclear flow and timing between when the risk score was calculated and when the outcome was measured. There were also very low event rates in many of the studies and therefore not a reasonable number of outcome events compared to the number of factors in the risk tool. Many papers also failed to report all the relevant performance measures (sensitivity and specificity). The committee were concerned about the applicability of some risk tools for UK practice due to the setting the tool was originally derived in as well as the location of the validation studies. The committee noted the differences in care settings and medical practices in the US and decided to downgrade any papers from a US setting for indirectness (see further detailed discussion on this in the following section). Clinical effectiveness of risk tools for reducing VTE No randomised controlled trials were identified, therefore observational studies were considered for inclusion in this review. Four observational studies were included in this review (one retrospective cohort study and three before-and-after studies). Two of the studies compared use of a risk tool versus with no risk tool (the National VTE Risk Assessment Tool [otherwise known as the Department of Health tool, please see the other considerations section for further detail] and the Padua Prediction Score); and two studies compared achieving the quality standard of 90% of admissions being assessed with the National VTE Risk Assessment Tool with not achieving the quality standard. The committee discussed the need for caution when evaluating evidence from quality standard cohort papers and before-and-after studies due to the risk of bias inherent in these designs. The four observational studies provided evidence of very low quality due to risk of bias, primarily based on selection bias and incomplete outcome data; and imprecision around the effect estimates.
Trade-off between clinical benefits and harms	There is no established definition of medical patients, and the papers included in this review cover different groups of people including acutely ill medical patients, people who have had acute stroke and people with cancer; all with different associated thrombotic and bleeding risks. The rate of VTE identified in the evidence ranged from 0.5-4.5%. This large disparity is due to a number of factors, including: the heterogeneous group of patients; different study designs including RCT, prospective and retrospective cohorts and database/registry studies; and different definitions of the VTE endpoint (asymptomatic or symptomatic). Of the 18 studies reporting on risk tools in medical patients only three of these were undertaken in the UK NHS context. All three of these looked only at the National VTE Risk Assessment Tool (hereafter referred to as the National Tool) but none were designed specifically to validate whether this tool can adequately predict risk of VTE or risk of bleeding the UK population. Evidence was identified for a number of VTE risk assessment tools for medical patients including the Padua prediction score, the Kucher score, the Intermountain score and the IMPROVE tool. Evidence was also identified for a bleeding risk version of the IMPROVE tool. The committee discussed these tools at length including the

 ISBN 978 – 1 – 4731 – 2871 - 2

various risk factors that went into them and whether these were weighted or not. The committee noted that the National Tool and Intermountain score performed more like a checklist as they are not weighted tools but instead involve an in-or-out decision. The committee determined that none of the tools demonstrated sufficiently accurate performance for predicting VTE or bleeding risk based on the evidence, with none reaching the committee's pre-specified sensitivity and specificity thresholds and many reporting only poor discrimination.

All committee members agreed that risk assessment is a critical part of the pathway for VTE prophylaxis. They also agreed that risk tools are beneficial in this process. However, in the absence of clear evidence there was disagreement about which tool to recommend. Based on its increasing use in the US context, initial discussions considered whether the IMPROVE Tool should be recommended over current practice, which is the National Tool.

There are two different versions of the IMPROVE tool. The 4-factor version of the tool is known as the predictive version because information on all 4 factors the tool measures should be available at admission and are considered to be predictive of VTE during the 3-month period following hospital admission.[179] The 7-factor version of the tool is known as the association version because some of the extra factors will require judgement of in-hospital factors that cannot be known for certain on admission (for example expected number of days the person might be immobilised) that are believed to be associated with an increased risk of VTE during the 3-month period following hospital admission.[179] Evidence included in this review is for the 4-factor version of IMPROVE as this was the only version with an identified validation study that met the inclusion criteria for the review. No validation studies of the 7-factor tool met the criteria in the review protocol.

The committee noted that the National Tool has been embedded in practice for 7 years with a high level of adherence. However, several committee members were of the opinion that the tool leads to over prescribing of prophylaxis in medical patients without clear evidence of benefit, potentially incurring a significant cost to the NHS. Around 73% of medical patients in the UK receive prophylaxis using the National Tool (NHS Safety Thermometer Data – March 2016 to March 2017, published April 12, 2017; accessed 15 August 2017) compared to around 40% of medical patients (in largely US based populations) for other tools. [64] The committee considered the high rate of prophylaxis being given was in part due to the way the National Tool is being used in practice. The National Tool may have become a 'tick-box exercise' where clinicians view it as a unweighted checklist of risk factors; if you tick one box (a single risk factor), that equates to a high VTE risk and this automatically results in prophylaxis being offered. The committee stressed that this has led to a larger number of medical patients receiving VTE prophylaxis than would be expected. Most importantly this fails to highlight the clinical judgement that must come into play in order to consider whether individual risk factors lead to an overall increased risk, and the balance of this with any bleeding risk factors or other contraindications. The committee understood that none of the identified tools, nor the currently practiced National Tool, offer clear guidance on how to balance VTE risk and bleeding risk to come to a decision on whether to offer prophylaxis, and if so what type. While the IMPROVE tool has both a VTE risk and bleeding risk version, both of which are available in online calculator format (beta version and no validation available), these also only provide a percentage risk for each outcome with no guidance on how to balance the two.

The committee also discussed the indirect context of the evidence for the IMPROVE tools (both the VTE risk version and the bleeding risk version). In particular the committee highlighted that in the US a much higher proportion of medical patients are cared for on intensive care wards (ICU), whereas in the UK it is only the very ill (generally those in need of artificial ventilation) who are moved to critical care – so the baseline condition of the two populations would be very different. The 7-factor version of the IMPROVE tool has ICU/CCU stay as a major risk component and this would contribute to different risk assessment interpretations in the UK compared to

 ISBN 978 – 1 – 4731 – 2871 - 2

	the US population in which the tool is validated. The committee also acknowledged that the average length of stay in intensive care is around 7 days in the USA, compared to a shorter stay of approximately 2–3 days in the UK. This is reflected in the National Tool listing mobility significantly reduced ≥3 days as a risk, and the 7-factor IMPROVE tool listing immobilisation ≥7 days as a risk. Factors such as these require the clinician to make judgements about anticipated patient features that cannot be known with certainty at admission. The committee pointed out that tools that require information that may not be available at the point of admission are not practical. Overall, the committee agreed that there is a lack of good quality evidence for any tool. The following options were considered as recommendations for assessing risk in medical patients: (1) use the National Tool (2) use the IMPROVE Tool (3) use either the National Tool or the IMPROVE Tool (4) consider medical patients at risk if immobility was a factor and they have an additional risk factor, with individual risk factors being provided as examples in a box; (5) use an existing derived or validated tool or checklist. After considerable debate a committee meeting consensus was reached to rule out the first 3 options. However, no consensus was reached on whether to recommend options number 4 or 5. The main arguments behind supporting each of these options were: • Those favouring option 4 expressed concerns with recommending option 5. They were concerned about organisational rigour in a resource-stretched NHS and that the decision on which tool to use will be made that may not be in the patient's best interest. A particular tool may be chosen because of potential cost saving benefit and not because it is considered to be more accurate or effective. • Those favouring option 5 believed it better reflects the uncertainty in evidence as there is no clear evidence that one tool is better than another. It allows clinicians to decide which tool to use whereas option 4 seemed too similar to current practice. It would also prompt clinicians to consider that risk assessment for VTE is not just a checklist of risk factors that once ticked automatically mean prophylaxis, it is a balance between VTE risk and bleeding risk which requires clinical judgement before the decision to offer prophylaxis is made. Because of the split decision the committee voted for one of these two options and agreed whichever option had the most votes would determine the recommendation. The vote produced a majority favouring option 5. Following stakeholder consultation the committee also decided to acknowledge in the recommendation that the most commonly used VTE risk assessment tool for hospital patients in the NHS is the National Tool (see appendix T). Reflecting the uncertainty in the evidence for one risk tool over another, the committee prioritised a research recommendation in this area.
Trade-off between net clinical effects and costs	Two economic studies were included. One of the studies compared the use of a risk assessment tool for medical patients based on the PRETEMED scale (a validated risk stratification tool for medical patients) which was integrated in the hospital electronic system in the form of an e-alert system. The second study assessed the impact of restricting the provision of LMWH prophylaxis based on a list of risk factors that allow restricted, intermediate or broad eligibility for prophylaxis in general medical patients admitted to hospital. The committee discussed the two studies and noted that the study that compared using a risk assessment tool to not using one showed that the use of a risk assessment tool was dominant (both more effective

 ISBN 978 – 1 – 4731 – 2871 - 2

	and less costly). The committee acknowledged however that the tool used in this study was not validated and was not one of those identified in the clinical review. The committee highlighted that all the risk tools included in the clinical review are generally not associated with any licencing cost although some may require a specific software installation. However, the committee acknowledged that the prognostic performance of the risk tool, as well as the baseline risk in the target population, would determine the number of individuals who would receive prophylaxis. The choice of a tool that has high specificity would minimise the cost of unnecessary prophylaxis provision. If the specificity of a tool is low, there is a risk that a large number of people will be triggered for further care that they do not require (over-treatment), which would make the tool unlikely to be cost-effective. Conversely, if the tool has low sensitivity then a large number of people will not be identified as being at risk of VTE, and therefore not receive the prophylaxis they could benefit from. The committee determined that the evidence for the prognostic accuracy of the tools identified was inconclusive and does not support recommending one tool over another. This increases the uncertainty in the cost effectiveness of these tools. The committee acknowledged that the use of the National Tool is considered current practice for surgical, medical and trauma patients. Hence, any changes are likely to have cost impact. For medical admissions, the committee discussed the potential of using the IMPROVE tool, both the 4- and 7- factor versions; however there were concerns about the fact that neither has been validated in a UK population. Furthermore, the tool mainly assesses the risk of symptomatic VTE and does not identify patients at risk of developing an asymptomatic DVT. A cost impact analysis was also undertaken to aid the committee's decision making. This analysis showed that using the IMPROVE risk assessment tool would result in around 40% of people having prophylaxis, in line with the intermediate eligibility group in the Miller study. The saving from this reduction in prophylaxis is estimated to be around £22.3 million. However after the extensive discussions and voting process outlined above, it was determined that the evidence underpinning the accuracy and effectiveness of IMPROVE and all the tools considered for medical patients (including the National Tool) did not show that one tool is better than the other and a research recommendation was made to allow for future research to address the uncertainty in this area.
Other considerations	The National VTE Prevention Programme was launched in England in 2010 mandating VTE risk assessment in all adult patients admitted to an acute hospital, using a National VTE risk assessment tool.[161] The committee noted that CG92 and the National Tool were published concurrently in 2010, therefore CG92 did not recommend the National Tool by name. However, it was also noted that the recommendation in CG92 and the National Tool is identical. The initial goal as part of the Commissioning for Quality Innovation (CQuIN) Framework was to set a 90% target of all patients risk assessed for VTE. This was supported by a financial incentive (CQuIN) payment and within 3 years this goal was increased to 95% which has been exceeded in subsequent years.[161] However the committee noted that there have been no published studies examining the long-term impact of the National VTE prevention programme, specifically no research has been conducted validating the National Tool's performance at predicting medical patients' risk of VTE and risk of bleeding. The committee expressed their disappointment in this, especially as this was an area highlighted for further research by the CG92 committee. The committee made a high-priority research recommendation on risk assessment tools; see appendix R for more details. The committee discussed giving guidance on the appropriate time to initiate

 ISBN 978 – 1 – 4731 – 2871 - 2

	pharmacological prophylaxis following completion of the risk assessment. In particular the committee wanted to highlight that, if using pharmacological prophylaxis, it should be given in a timely manner to ensure that people are not left for too long without it if they happened to be admitted shortly after what is usually a set daily time for doses to be given on a ward. The committee recommend a time point that is in line with current NHS policy on time to consultant review of acute inpatients. This standard states that all emergency admissions must be seen and have a thorough clinical assessment by a suitable consultant as soon as possible, but at the latest within 14 hours from the time of admission to hospital.[134] The committee agreed that recommending a similar timeframe within which pharmacological prophylaxis should be given (if indicated by risk assessment) makes logical clinical sense and will ensure clinical care is not delayed.

Recommendations	***Surgical and trauma patients*** **1.1.5 Assess all surgical and trauma patients to identify the risk of VTE and bleeding:** • **as soon as possible after admission to hospital or by the time of the first consultant review** • **using a tool published by a national UK body, professional network or peer-reviewed journal. The most commonly used risk assessment tool for surgical patients is the Department of Health VTE risk assessment tool[nnn] (See Appendix T). [2018]** **1.1.6 Balance the person's individual risk of VTE against their risk of bleeding when deciding whether to offer pharmacological thromboprophylaxis to surgical and trauma patients. [2018]** **1.1.7 If using pharmacological VTE prophylaxis for surgical and trauma patients, start it as soon as possible and within 14 hours of admission, unless otherwise stated in the population-specific recommendations (see chapters 9-13). [2018]**
Research recommendation	**1. What is the accuracy of individual risk assessment tools in predicting the risk of VTE and risk of bleeding in surgical and trauma patients admitted to hospital?**
Relative values of different outcomes	**Predictive accuracy of VTE and bleeding risk tools** The committee was interested in the prognostic accuracy of risk assessment tools for surgical and trauma patients admitted to hospital or who are in hospital having day-case surgery. A risk assessment tool would be used to identify people with an increased risk of VTE who would benefit from having VTE prophylaxis, or identify people with an increased risk of major bleeding in order to determine appropriate prophylaxis strategies, for example not giving pharmacological prophylaxis to people who were at a high risk of bleeding. The committee agreed that sensitivity was more important than specificity in surgical patients because people who are at higher risk of VTE could be identified for potential VTE prophylaxis treatment (fewer false negatives). The committee set thresholds for the acceptability of a test; for the populations noted here, these were

nnn Reproduced with the permission of the Department of Health and Social Care under the Open Government Licence.

 ISBN 978 – 1 – 4731 – 2871 - 2

	≥80% sensitivity and ≥60% specificity. Some studies only reported a C-statistic. The committee acknowledged that this metric was important for comparing the overall accuracy of the tools, but in itself was unlikely to provide enough information on which to base a recommendation as it does not indicate the number of false positives and negatives of the tool. Therefore, the committee decided against recommending a tool without sensitivity and specificity data. **Clinical effectiveness of risk tools for reducing VTE** For the review of clinical effectiveness of risk tools, the committee considered all-cause mortality, VTE (symptomatic or asymptomatic), DVT (symptomatic or asymptomatic), PE, fatal PE, major bleeding and quality of life as critical outcomes. The time points for these outcomes were up to 90 days from hospital discharge. The committee considered fatal bleeding, clinically relevant non-major bleeding, heparin-induced thrombocytopenia, hospital length of stay, unplanned readmission and haemorrhagic stroke as important outcomes. The time points for these outcomes were up to 90 days, apart from clinically relevant non-major bleeding up to 45 days from hospital discharge. Please see section 4.4.3 in the methods chapter for further detail on prioritisation of the critical outcomes.
Quality of the clinical evidence	**Predictive accuracy of VTE and bleeding risk tools** Fifteen studies were identified looking at risk tools for predicting VTE in surgical or trauma patients. Fourteen papers featured people admitted to hospital (10 for surgery, 3 for trauma and 1 for burn injuries) and one featured people in hospital for day-case surgery. No studies were identified looking at risk tools to predict the risk of major bleeding in surgical or trauma patients. PROBAST was used to assess the risk of bias. All of these studies were at a high or very high risk of bias. Common reasons for this were papers only supplying retrospective validation, papers not reporting a clear definition or method of confirmation for the target condition (VTE, DVT, PE or major bleeding), papers not reporting the time-point for the target condition measurement, or unclear flow and timing between when the risk score was calculated and when the outcome was measured. There were also very low event rates in many of the studies and therefore not a sufficient number of outcome events compared to the number of factors in the risk tool. Many papers also failed to report all the relevant performance measures (sensitivity and specificity). The committee were concerned about the applicability of some risk tools for UK practice due to the setting the tool was originally derived in as well as the location of the validation studies. The committee noted the differences in care settings and medical practices in the US and decided to downgrade any papers from a US setting for indirectness. In particular, the committee highlighted that in the US a much higher proportion of surgical patients are cared for on intensive care wards (ICU), whereas in the UK it is only the very ill (generally those in need of artificial ventilation) who are moved to critical care – so the baseline condition of the two populations would be very different. The committee also considered that the average length of stay in intensive care is around 7 days in the US, compared to a shorter stay of approximately 2–3 days in the UK. **Clinical effectiveness of risk tools for reducing VTE** No randomised controlled trials were identified, therefore observational studies were considered for inclusion in this review. Two observational studies were included in this review (one retrospective cohort study and one before-and-after study). One compared use of the Caprini risk assessment model with no risk assessment tool and one study compared achieving the quality standard of 90% of admissions being assessed with the National VTE Risk Assessment Tool (otherwise known as the Department of Health tool, please see the other considerations section for further detail) with not achieving the quality standard.

 ISBN 978 – 1 – 4731 – 2871 - 2

	The committee discussed the need for caution when evaluating evidence from quality standard cohort papers and before-and-after studies due to the risk of bias inherent in these designs. The two observational studies provided evidence of very low quality due to risk of bias, primarily based on selection bias and incomplete outcome data. There was imprecision around the effect estimates, and the evidence on the Caprini risk assessment model was also downgraded for indirectness due to the setting being in the US hospital system where practice differs from the UK context.
Trade-off between clinical benefits and harms	Evidence for risk assessment tools came from a very wide range of surgical populations, including abdominal, colorectal, lung, neuro, oesophageal, plastic, and urological surgery; as well as mixed surgical populations, trauma patients and those undergoing day-case surgery (surgical outpatients); all with different associated thrombotic and bleeding risks. The rate of VTE identified in the evidence ranged from 0.33–27.9%. This very large disparity is due to a number of factors including the heterogeneous group of patients and surgery-associated VTE risk; different study designs including RCT, prospective and retrospective cohorts and database/registry studies; and different definitions of the VTE endpoint (asymptomatic or symptomatic). Of the 17 studies reporting on risk tools in surgical and trauma patients only one was undertaken in the UK NHS context. This UK study looked at the National VTE Risk Assessment Tool (hereafter referred to as the National Tool) but was not designed specifically to validate whether this tool can adequately predict risk of VTE or risk of bleeding in the UK surgical population. Evidence was identified for a number of VTE risk assessment tools including the Caprini risk assessment model, the American College of Surgeons National Surgical Quality Improvement Programme (ACS NSQIP) Universal Surgical Risk Calculator (not specific to the outcome of VTE) and the Trauma Embolic Scoring System (TESS). No tool was identified to assess the risk of bleeding. The majority of the evidence was found for the Caprini risk assessment model, which is a weighted tool made up of an extensive list of risk factors. Low and very low quality evidence from some highly specific surgical populations (lung cancer, oesophageal cancer, and high-risk abdominal and neurosurgical) suggested that the Caprini risk assessment model reached the committee's thresholds for consideration for both sensitivity and specificity when using cut-offs such as ≥9, ≥10.5 and ≥15. The low and very low evidence from these studies suggested the tools showed moderate discrimination for predicting VTE. Very low quality evidence from the clinical effectiveness review also suggested a reduction in DVT rates when using the Caprini risk assessment model compared to using no formal risk assessment. All committee members agreed that risk assessment is a critical part of the pathway for VTE prophylaxis. They also agreed that risk tools are beneficial in this process. Based on the evidence, initial discussions considered whether the Caprini risk assessment model should be recommended over current practice, which is the National Tool. The committee highlighted that there was not thought to be the same issue within the surgical population as that recognised in the medical population (use of the National Tool leading to giving too much prophylaxis). However, they acknowledged that the National Tool has not been validated in any surgical population or in people with trauma. While the evidence suggested the Caprini risk assessment model could be beneficial, the evidence was of low to very low quality and was only validated in highly specific surgical populations and the committee could not be sure that these findings could be generalised to the wider 'mixed' surgical population. There was also concern that the Caprini risk assessment model has almost exclusively been validated only in a US population, and never in the UK population. Following decisions on the recommendation for risk assessment in medical patients, the committee discussed whether it was conceptually feasible to recommend different risk assessment tools for the surgical and trauma patients as for the medical patients. They highlighted that the distinction between these two populations is becoming increasingly blurred in the current UK context as surgical

 ISBN 978 – 1 – 4731 – 2871 - 2

	patients will increasingly be older and/or have more medical comorbidities (increasing rates of life-style diseases such as obesity, non-alcoholic fatty liver disease and diabetes). This was also discussed in the context of day-case or outpatient surgery. This covers a mix of minor procedures and as technology improves, and surgeons have access to innovative technologies, surgical time will be reduced and an increasing amount of surgical procedures will become day cases. For this population the VTE and bleeding risk may not necessarily be related to the surgical procedure, but instead be related to the pre-surgical context (for example their medical status). The committee agreed that it was logical and advisable to have the same risk assessment recommendation for the surgical and trauma population as for the medical population. They also considered that the question of risk assessment tools for the surgical and trauma population was a key priority for future research alongside the research recommendation for risk assessment tools in the medical population. Following stakeholder consultation the committee also decided to acknowledge in the recommendation that the most commonly used VTE risk assessment tool for hospital patients in the NHS is the National Tool (see appendix T).
Trade-off between net clinical effects and costs	One economic study was included. This compared the use of a risk assessment tool based on using ACCP guidelines for surgical patients which were integrated in the hospital electronic system in the form of an e-alert system. The committee discussed the study and noted that, similar to the general medical population in the study, the use of a risk assessment tool for surgical patients was dominant (both more effective and less costly). The committee noted that all the risk tools included in the clinical review are generally not associated with any licencing cost although some may require a specific software installation. However, the committee agreed that the evidence for the tools identified was inconclusive and does not support recommending one tool over another. The committee acknowledged that the use of the National Tool for both surgical and trauma patients is currently embedded in NHS practice. However, in contrast to the case in medical patients, the committee did not feel that this tool led to over-prescribing of prophylaxis in the surgical population given the higher baseline risk of VTE compared to general medical patients. The committee also acknowledged that changing from the use of the National Tool to any other tool is likely to have a cost impact to allow the integration of a new tool into practice, which would require robust evidence in terms of clinical and cost effectiveness to support it. The current status of the retrieved evidence did not offer a strong base for recommending any of the identified tools. The committee discussed the potential of using the Caprini tool, however there were concerns about the fact that it has not been validated in a UK population and also that it has only been validated in a small number of surgical specialities. After the extensive discussions and voting process outlined in the discussion on risk assessment in medical patients, it was determined that the evidence underpinning the accuracy and effectiveness of all the tools considered for the surgical and trauma populations did not show that one tool is better than the other and a research recommendation was made to allow for future research to address the uncertainty in this area.
Other considerations	The National VTE Prevention Programme was launched in England in 2010 mandating VTE risk assessment in all adult patients admitted to an acute hospital, using a National VTE risk assessment tool.[161] The committee noted that CG92 and the National Tool were published concurrently in 2010, therefore CG92 did not recommend the National Tool by name. However, it was also noted that the recommendation in CG92 and the National Tool is identical. The initial goal as part of the Commissioning for Quality Innovation (CQuIN) Framework was to set a 90% target of all patients risk assessed for VTE. This was supported by a financial incentive (CQuIN) payment and within 3 years this goal was increased to 95% which has been exceeded in subsequent years.[161] However the

 ISBN 978 – 1 – 4731 – 2871 - 2

	committee noted that there have been no published studies examining the long-term impact of the National VTE prevention programme, specifically no research has been conducted validating the National Tool's performance at predicting surgical and trauma patients risk of VTE and risk of bleeding. The committee expressed their disappointment in this, especially as this was an area highlighted for further research by the CG92 committee. The committee made a high-priority research recommendation on risk assessment tools; see appendix R for more details. The committee discussed giving guidance on the appropriate time to initiate pharmacological prophylaxis following completion of the risk assessment. In particular the committee wanted to highlight that, if using pharmacological prophylaxis, it should be given in a timely manner to ensure that people are not left for too long without it if they happened to be admitted shortly after what is usually a set daily time for doses to be given on a ward. The committee recommend a time point that is in line with current NHS policy on time to consultant review of acute inpatients. This standard states that all emergency admissions must be seen and have a thorough clinical assessment by a suitable consultant as soon as possible, but at the latest within 14 hours from the time of admission to hospital.[134] The committee agreed that recommending a similar timeframe within which pharmacological prophylaxis should be given (if indicated by risk assessment) makes logical clinical sense and will ensure clinical care is not delayed.

 ISBN 978 – 1 – 4731 – 2871 - 2