National Institute for Health and Care Excellence

Final

Venous thromboembolic diseases: diagnosis, management and thrombophilia testing

[B] Evidence review for the use of the pulmonary embolism rule-out criteria for diagnosis of pulmonary embolism

NICE guideline NG158

Evidence review underpinning recommendation 1.1.16 in the guideline

March 2020

Final version

This evidence review was developed by the NICE Guideline Updates Team

Disclaimer

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

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

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN: 978-1-4731-3735-6

Contents

Pulmonary embolism rule-out criteria for suspected pulmonary embolism (PE)

Review question

In people with suspected pulmonary embolism (PE), what is the diagnostic accuracy of the pulmonary embolism rule-out criteria (PERC)?

Introduction

The pulmonary embolism rule-out criteria (PERC) are a set of clinical criteria that have been designed to allow people with a low probability of pulmonary embolism (PE) to be discharged without further testing. The NICE guideline on the management of venous thromboembolism (VTE) does not currently recommend the use of PERC in the diagnostic pathway. Since the previous guideline was published, new evidence on the diagnostic accuracy and clinical effectiveness of PERC has become available, and this evidence may have an impact on the guideline recommendations. This update reviews the diagnostic accuracy and clinical effectiveness of using the pulmonary embolism rule-out criteria as part of the diagnostic pathway for suspected PE. It identified studies that fulfilled the conditions listed in Table 1. For full details of the review protocol, see appendix A.

Table 1 PICO table for PERC for suspected PE

Population	Adults (aged 18+) with clinically suspected PE
Intervention	**Diagnostic accuracy studies:** • Pulmonary embolism rule-out criteria (PERC) **Test and Treat RCTs:** • Pulmonary embolism rule-out criteria (PERC)
Comparator	**Diagnostic accuracy studies:** • Reference standards (in order of prioritisation): – VTE event during 3 months of follow up – CT Pulmonary angiography – V/Q scan – MRI scan – Pulmonary angiography. **Test and treat RCTs:** Usual care: PERC not used
Outcomes	**Diagnostic accuracy studies:** • Diagnostic accuracy metrics: Sensitivity/specificity, Positive and negative likelihood ratios **Test and treat RCTs:** • All-cause mortality • VTE-related mortality • Diagnostic strategy failure – defined as occurrence of VTE at follow up in patients for which VTE was ruled out • Length of hospital stay • Quality of life

- Adverse events

Methods and process

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

Subgroup analyses were carried out for people with low, intermediate and high pre-test probability of having PE. The classification of results into these groups was based on the definition used in the individual papers. This varied from clinician judgement based on a number of factors (for example, patient characteristics, diagnostic hypotheses in Penazola 2012) to the use of established scoring systems such as the revised Geneva score (Hugli 2011).

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

Protocol deviation

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

Clinical evidence

Included studies

A systematic search was carried out for this review question to identify diagnostic accuracy studies, test-and-treat randomised controlled trials and systematic reviews of these studies, which found 664 references (see appendix C for literature search strategy). Based on title and abstract, 641 references were excluded and 23 references were ordered for screening based on their full texts.

Of the 23 references screened as full texts, 7 references were included based on their meeting the inclusion criteria specified in the review protocol (appendix A). Several systematic reviews were identified as being relevant to the review question. These were used for reference searching to identify primary studies but were not included in the review. Data was extracted directly from the primary papers instead. The clinical evidence study selection is presented as a diagram in appendix D.

A second set of searches, using the original search strategies, were conducted at the end of the guideline development process to capture papers published whilst the guideline was being developed. These searches returned 6,272 references in total for all the questions included in the update, and these were screened on title and abstract. 5 references were included for full text screening, but no additional relevant references were found for this review question.

For the full evidence tables and full GRADE profiles for included studies, please see appendix E and appendix G. The references of individual included studies are given in appendix K.

Excluded studies

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

Summary of clinical studies included in the evidence review

The tables below (Table 2, Table 3) provide a summary of the randomised controlled studies and diagnostic accuracy studies included in this review.

Table 2 Randomised controlled study

Author (year)	Study details	Intervention	Control
Freund (2018)	Cluster RCT across 14 emergency departments 1916 participants	6-month periods in which each cluster used PERC	6-month period in which each cluster did not use PERC

Table 3 Diagnostic accuracy studies

Author (year)	Study details	Index test	Reference standard
Crichlow (2012)	Prospective cohort study 152 participants	PERC	Composite (CTPA or event in 90 days follow-up if negative CTPA)
Hogg (2005)	Prospective cohort study 425 participants	PERC	Composite (D-dimer tests plus low clinical probability, V-Q scan, CTPA or digital subtraction PA)
Hugli (2011)	Prospective cohort study 425 participants	PERC	Composite (MDCT, PA, V-Q scan, CUS or event in 3-month follow-up)
Kline (2004)	Prospective cohort study 1427 participants	PERC	Composite (D-dimer test, CTPA, CT angiography-venography or VQ-scan with selected use of venous ultrasonography if D-dimer positive).
Penaloza (2012)	Prospective cohort study 959 participants	PERC	Composite (event in 3-month follow-up, PE-related death or PE diagnosis at end of initial diagnostic work-up [unclear which tests were used])
Penaloza (2017)	Prospective cohort study 1052 participants	PERC	Composite (highly sensitive age adjusted D-dimer, CTPA, leg ultrasonography or V-Q scan)

See appendix E for full evidence tables.

Quality assessment of clinical studies included in the evidence review

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

Economic evidence

Included studies

A systematic search was carried out for this review question to identify relevant economic analyses. This search returned 192 records, all of which were excluded on title and abstract.

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

Economic model

For this review question, the committee indicated that, alongside test accuracy data for PERC, recommendation making would be facilitated by information on absolute numbers of patients with each test outcome (i.e. true positives, false negatives, true negatives, and false positives), as well as estimates of costs involved in the diagnostic pathway. To provide this information, we developed a simple cost-consequences analysis, comparing outcomes with and without PERC as an initial step in the diagnostic pathway in people at low risk of PE. In the base case, the prevalence of PE was assumed to be 7.3% (Goekoop 2007).

A full cost-utility analysis was felt to be inappropriate for this review question, as cost effectiveness is likely to be heavily dependent on the long-term health outcomes and costs associated with false negative results (patients who have a PE but are incorrectly diagnosed). Since randomised evidence of sufficient quality on the consequences of an intentionally untreated PE is unlikely to exist, such an analysis would not be feasible without substantial speculation on the downstream outcomes for these patients.

The base-case results of the cost-consequences analysis in terms of the test outcomes and cost per 1000 people are presented in Table 4. A more detailed description of the model is provided in appendix I.

Table 4 Test outcomes and total costs for the entire diagnostic pathway for "PERC" and "No PERC" strategies per 1,000 patients (base case)

	PERC	No PERC	Difference (95% CrIs)
Testing outcome			
True positive	59	61	-3 (-5 to -1)
False negative	14	11	3 (1 to 5)
True negative	904	899	5 (2 to 11)
False positive	23	29	-5 (-11 to -2)
Costs			
Total	£77,970	£96,292	-£18,322 (-£29,486 to -£9,416)

A sensitivity analysis was undertaken using diagnostic accuracy data for PERC from studies in people with a low pre-test probability of PE only but still assuming the base case prevalence of 7.3%. Results of this sensitivity analysis are shown in Table 5.

Table 5 Test outcomes and total costs for the entire diagnostic pathway for "PERC" and "No PERC" strategies per 1,000 patients (sensitivity analysis using accuracy data from the low pre-test probability subgroup analysis)

	PERC	No PERC	Difference (95% CrIs)
Testing outcome			
True positive	56	61	-6 (-14 to -1)
False negative	17	11	6 (1 to 14)
True negative	906	899	7 (3 to 13)
False positive	21	29	-7 (-13 to -3)
Costs			
Total	£71,258	£96,292	-£25,033 (-£30,002 to -£20,426)

In an additional sensitivity analysis, a value of 2.1% (from Freund 2018 from the clinical review) was used to represent the prevalence of PE (as opposed to the value of 7.3% in the base case). Results for this scenario were more favourable towards the "PERC" strategy than in the base case. Compared to "no PERC", the "PERC" strategy produced 8 fewer false positive results per 1,000 patients, at the expense of only 2 more false negative results, and a cost saving of £25,636 (see appendix I for further details).

Evidence statements

Clinical evidence statements

Cluster RCT evidence

The format of the evidence statements is explained in the methods in Appendix B.

- Very low to low quality evidence from one cluster RCT containing data on 1,916 participants **could not differentiate** all-cause mortality or diagnostic strategy failure when PERC was used at the start of the diagnostic pathway to rule-out PE compared with when it was not.
- Moderate quality evidence from 1 cluster RCT containing data on 1,916 participants **could not estimate** an effect on VTE-related mortality or major bleeding as both arms reported 0 events.
- Low quality evidence from one cluster RCT containing data on 1,916 participants **found a significantly shorter** median length of emergency department stay when PERC was used at the start of the diagnostic pathway to rule-out PE compared with when it was not (median difference 37 minutes shorter, interquartile range 4 minutes to 1hr 8 minutes shorter).

Diagnostic test accuracy evidence

The format of the evidence statements is explained in the methods in Appendix B.

- The evidence suggests that a **negative** PERC result indicates a **moderate decrease** in the probability that a person with clinically suspected pulmonary embolism has a pulmonary embolism (LR-=0.21 [0.14 to 0.30]). This is also the case when the analysis is restricted to evidence for people with a low pre-test probability of PE (LR-=0.33 [0.14 to 0.77]) or intermediate test-probability (LR- = 0.22 [0.02, 2.63]) or high pre-test probability (LR- = 0.36 [0.07, 1.73]). (Very low to low quality evidence from up to 6 prospective studies comprising up to 5,690 participants)

- The evidence suggests that a **positive** PERC result indicates a **slight increase** in the probability that a person with clinically suspected pulmonary embolism has a pulmonary embolism (LR+=1.24 [1.11 to 1.45]). This is also the case when the analysis is restricted to evidence for people with a low pre-test probability of PE (LR+=1.30 [1.23 to 1.38]) or intermediate test-probability (LR+ =1.04 [1.02, 1.06]) or high pre-test probability (LR+ =1.02 [0.98, 1.08]. (Very low to low quality evidence from up to 6 prospective studies comprising up to 5,690 participants)

- The sensitivity of PERC was 0.95 (0.91-0.98) and the specificity was 0.23 (0.12-0.37). (Evidence from 6 prospective studies comprising 5,690 participants)

Sensitivity analyses removing studies at high risk of bias (main analysis only)

- The evidence suggests that a **negative** PERC result indicates a **moderate decrease** in the probability that a person with clinically suspected pulmonary embolism has a pulmonary embolism (LR-=0.22 [0.14 to 0.32]) and that a **positive** PERC indicates a **slight increase** in probability (LR+=1.22 [1.11 to 1.38]). (Very low to moderate quality evidence from 4 prospective studies comprising up to 4,304 participants).

Economic evidence statements

- A *de novo* cost-consequences model developed for this review question found that using PERC at the start of the diagnostic pathway for patients at low risk of PE is likely to produce cost savings in the diagnostic testing pathway due to fewer false positive results but leads to an increase in false negative results. It was not possible to fully capture all of the downstream health consequences and costs associated the trade-off between the reduction in false positive results and the increase in false negative results.

The committee's discussion of the evidence

Interpreting the evidence

The outcomes that matter most

The committee discussed the impact that true positive, false positive, true negative and false negative PERC results have on patients. People with true positive results go on to further diagnostic tests to confirm PE diagnosis and then receive appropriate anti-coagulation therapy, people with false positive results undergo unnecessary further testing which poses an unnecessary radiation risk (in the case of chest imaging) and healthcare expense, as well as unnecessary anxiety. People with true negative results are correctly discharged and reassured that they do not have PE, and people with false negative results are incorrectly discharged and go untreated with the risk of disease progression and complications, including death.

The committee noted that PERC is a rule out test and its use is being proposed at the very beginning of the diagnostic pathway. There were no studies that allowed us to compare test accuracy (sensitivity and specificity) of PERC to nothing because in practice, the clinician still needs to make a judgement whether or not to refer people for more testing. The committee agreed it was therefore necessary to consider the whole diagnostic pathway when analysing the effectiveness and cost-effectiveness of PERC.

To interpret the clinical evidence, the committee agreed that likelihood ratios were useful if they had to make a decision about the value of PERC alone as a rule out test because likelihood ratios correspond to specific categories for the change in probability of having the disease associated with a given test result (see Table 8). They therefore agreed that assessment of the quality of the findings and decision- making could be carried out for the likelihood ratios rather than using sensitivity and specificity because these measures lacked an equivalent test to be compared to at that point in the diagnostic pathway.

For evidence from test-and-treat randomised trials, the committee considered that the available evidence came from a single test-and-treat RCT and that no robust conclusions could be made for any of the outcomes examined due to the low quality of evidence available. As a result, the committee did not prioritise any outcomes from this study.

The quality of the evidence

Evidence from diagnostic accuracy studies was of low to very low quality. The review was restricted to prospective studies, and although data was gathered prospectively in all of the studies that were included, the PERC criteria were assessed retrospectively in many studies. Several studies were at high risk of bias because of the lack of blinding of people assessing the index test or reference standard. However, a sensitivity analysis excluding studies at high risk of bias showed very similar results to the main analysis, indicating that this may not have had a large impact on the diagnostic accuracies reported by the studies. Composite reference standards were used in all studies, but the details of the reference standards used were often unclear and varied across studies. There was high heterogeneity in the analyses, particularly for specificity and negative likelihood ratios which decreased the certainty in the overall effect and might have occurred because of differences in the reference standards used. However, the committee agreed that pooling was appropriate given no clear criteria could be identified to explain the heterogeneity.

The committee agreed that assessment of the quality of the findings for each outcome and the corresponding evidence statements should use likelihood ratios as this helped to quantify how a positive or negative PERC score affects the likelihood of having a PE, and this method is useful for evaluating the utility of PERC.

The committee noted that low-risk pre-test probability of PE was defined in the studies based on clinician judgement and that this included the results of a D-dimer test in some studies, but not in others.

A single cluster-randomised crossover trial was also included. Evidence from this trial was low to very-low quality. The evidence had serious risk of bias because there was some evidence that inclusion bias might have occurred; the baseline characteristics of people in the PERC periods were different from those in the control periods, with lower baseline pre-test probability of PE in the PERC periods, which might have occurred because clinicians were not blinded. There was also very serious imprecision associated with the majority of outcomes because of the very low event rates in both PERC and control groups. The

committee also noted that the prevalence of PE in the participants in the study was very low – around 3% in the control group.

Benefits and harms

The committee considered evidence from diagnostic accuracy studies that showed that a negative PERC result was associated with a moderate decrease in probability that the person has PE. This result was also found for people with a low, moderate and high pre-test probabilities of having PE when subgroup analyses were carried out. The negative LR results reflected the high sensitivity of PERC, meaning that there would be few false negative results if the test was used in a population with low prevalence.

A cluster randomised controlled trial on people with very low pre-test probability of PE could not differentiate diagnostic strategies that did and did not use PERC in terms of mortality, diagnostic strategy failure and major bleeding events, though the committee noted that this evidence was very uncertain. This, coupled with evidence from the economic model (discussed in the section on cost effectiveness and resource use below) and the diagnostic test accuracy results mentioned above, supported a recommendation to consider using PERC at the beginning of the diagnostic pathway for people who had a low risk of having PE. The committee agreed that a low risk of PE refers to a situation where the person's probability of PE is estimated to be less than 15% based on an unstructured clinical gestalt assessment and where a differential diagnosis is possible. They based this definition on the included studies, many of which used clinician judgment to identify those people at low risk of PE who would be suitable for the PERC. Some studies used a threshold probability of less than 15% before the use of PERC was appropriate (for example, Freund, 2018). The committee agreed that people who are judged clinically suitable for PERC and in whom PERC rules out a PE could be discharged without further testing. The committee chose not to make a stronger recommendation for PERC because of the very low quality of the majority of the evidence informing the recommendation.

The committee did not make a recommendation to use PERC in people who the clinician thought were more likely to have PE (with intermediate and high pre-test probabilities) because the use of PERC in these people would not influence clinical practice by preventing the use of downstream tests. The committee agree that people at higher risk of PE would be investigated for PE irrespective of a PERC result and the use of PERC in these people would therefore be unnecessary and a waste of time and resources. The committee agreed that these people should be diagnosed following the existing pathway in the guideline. The committee noted that the studies evaluating PERC all took place in emergency departments, but they could see no reason why its use should be limited to this setting or why the diagnostic accuracy of PERC would differ in other settings.

Cost effectiveness and resource use

The committee considered the evidence from the *de novo* cost-consequences model developed for this review question, and noted that using PERC at the start of the diagnostic pathway produces a cost saving due to fewer false positive results, but at the expense of more false negative results, compared to the diagnostic pathway without PERC. The committee discussed the relative gravity of false negative and false positive results. They acknowledged that false negative results – leaving a patient with a PE untreated – is likely to lead to serious detrimental health effects and potentially substantial downstream costs. However, the committee also felt that, at the end of the diagnostic pathway, false positive results can also have relatively severe consequences, since inappropriately providing

anticoagulation to patients without a PE can result in adverse health consequences, such as bleeding events, as well as incurring anxiety, and unnecessary drug and monitoring costs. Furthermore, it is likely that patients with a false positive result actually have another underlying health problem, which remains untreated if PE is diagnosed.

This conclusion differs from the discussion of the results of the cost-consequences model developed for the D-dimer review questions, where the committee felt that false negative results were more serious than false positives. This is because the D-dimer cost consequences model only assesses outcomes resulting from a D-dimer test, rather than considering the entire pathway from start to finish. That is to say, patients with a false positive result will receive a chest scan, after which most will be correctly diagnosed. In contrast, model outcomes for this review question represent the final diagnosis of the testing pathway, and therefore determine the treatment that patients actually receive.

On balance, the committee felt that it would be inappropriate to make a strong recommendation either for or against PERC, considering the analysis could not fully capture the trade-off between the downstream consequences of false negative and false positive results. However, the model showed that in all scenarios, PERC is likely to result in cost savings to the diagnostic pathway and the scenario that tilted the balance of benefits to harms most in favour of using PERC was based on the test accuracy data from the subgroup analysis of studies in people with a low pre-test probability and a prevalence of PE of 2.1%, as reported in Freund 2018. In this scenario, PERC resulted in a reduction in the number of false positive results of 8 per 1,000 patients, a small increase in the number of false negative results of 2 per 1,000 patients and a cost saving of £25,636 per 1,000 patients.

The committee discussed the potential resource impact of their recommendation and determined that it is likely to produce a cost saving, due to fewer patients undergoing D-dimer tests and chest scans, and fewer patients without PEs being inappropriately anticoagulated. This cost saving may be slightly offset by the additional false negative results, but this impact is expected to be minimal when PERC is only applied to patients at low risk of PE. Overall, the committee felt that the resource impact is unlikely to be substantial.

Appendices

Appendix A – Review protocol

Field (based on PRISMA-P)	Content
Review question	In people with suspected PE, what is the diagnostic accuracy of the pulmonary embolism rule-out criteria (PERC)?
Type of review question	Diagnostic
Objective of the review	To assess the suitability of PERC as a rule-out strategy for people with suspected PE.
Eligibility criteria – population/disease	Adults (18+ years) with clinically suspected PE Clinically suspected is defined as exhibiting signs or symptoms of PE
Eligibility criteria – intervention(s)	**Diagnostic accuracy studies:** **Index tests** Pulmonary embolism rule-out criteria (PERC) **Test and Treat RCTs:** **Intervention:** Pulmonary embolism rule-out criteria (PERC)
Eligibility criteria – comparator(s)/control or reference (gold) standard	**For diagnostic accuracy studies:** Reference standards (in order of prioritisation): VTE event during 3 months of follow up CT Pulmonary angiography V/Q scan MRI scan Pulmonary angiography. **Test and treat RCTs:** **Comparator:**

	Usual care: PERC not used

Outcomes and prioritisation	**For diagnostic accuracy studies:**
	• Diagnostic test accuracy measures (sensitivity, specificity, positive and negative likelihood ratios)
	
	For test and treat RCTs:
	
	• All-cause mortality
	• VTE-related mortality
	• Diagnostic strategy failure – defined as occurrence of VTE at follow up in patients for which VTE was ruled out
	• Length of hospital stay
	• Quality of life
	  ○ Generic and disease-specific measures will be reported
	  ○ Overall score will be reported (data on subscales will not be reported)
	• Adverse events
	  ○ Total serious adverse events (as defined by the European medicines agency) will be reported if data is available.

Eligibility criteria – study design	• Prospective diagnostic accuracy studies • Randomised controlled trials
Other inclusion exclusion criteria	• English language papers included only. • Studies that do not report sufficient information to allow a 2*2 table (TP, FP, TN, FN) to be constructed will be excluded • Studies where performance of index test depends of the result of the reference test (or vice versa) will be excluded. (e.g. where only patients with positive PERC receive the reference standard) • Retrospective studies • Studies using different reference standards across participants • Case-controlled studies

Proposed sensitivity/sub-group analysis	• Delayed onset • Previous presentation of PE • Analysis will be stratified by pre-test probability (e.g. in groups categorised by Well's score) where data is available. • People with cancer. • People who have restricted movement. • People with chronic infection / HIV • People with previous VTE • People with delayed clinical presentation (7 days or more) • People with obesity III (a BMI of 40 kg/m^2 or more). • People who have stage 3 to 5 chronic kidney disease.
Selection process – duplicate screening/selection/analysis	10% of the abstracts were reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. If meaningful disagreements were found between the different reviewers, a further 10% of the abstracts were reviewed by two reviewers, with this process continued until agreement is achieved between the two reviewers. From this point, the remaining abstracts will be screened by a single reviewer. This review made use of the priority screening functionality with the EPPI-reviewer systematic reviewing software. See Appendix B for more details.
Data management (software)	See appendix B

Information sources – databases and dates	• Sources to be searched
	  ○ Clinical searches - Medline, Medline in Process, PubMed, Embase, Cochrane CDSR, CENTRAL, DARE (legacy records) and HTA.
	  ○ Economic searches - Medline, Medline in Process, PubMed, Embase, NHS

	EED (legacy records) and HTA, with economic evaluations and quality of life filters applied.
	• Supplementary search techniques
	  ○ None identified
	• Limits
	  ○ Studies reported in English
	  ○ Study design RCT, SR and Observational filter will be applied (as agreed)
	  ○ Animal studies will be excluded from the search results
	  ○ Conference abstracts will be excluded from the search results

Identify if an update	This is a new question for the update of the guideline, therefore no previous search has been undertaken for this question.
Author contacts	https://www.nice.org.uk/guidance/indevelopment/gid-ng10087
Highlight if amendment to previous protocol	For details please see section 4.5 of Developing NICE guidelines: the manual
Search strategy – for one database	For details please see appendix C of the evidence review
Data collection process – forms/duplicate	A standardised evidence table format will be used and published as appendix E (clinical evidence tables) or I (economic evidence tables) of the evidence review (where relevant).
Data items – define all variables to be collected	For details please see evidence tables in appendix E (clinical evidence tables) or I (economic evidence tables) of the evidence review (where relevant).
Methods for assessing bias at outcome/study level	See appendix B
Criteria for quantitative synthesis (where suitable)	See appendix B
Methods for analysis – combining studies and exploring (in)consistency	See appendix B

Meta-bias assessment – publication bias, selective reporting bias	See appendix B
Assessment of confidence in cumulative evidence	See appendix B
Rationale/context – Current management	For details please see the introduction to the evidence review.
Describe contributions of authors and guarantor	A multidisciplinary committee developed the guideline. The committee was convened by the NICE Guidelines Updates Team and chaired by Susan Bewley in line with section 3 of Developing NICE guidelines: the manual. Staff from the NICE Guidelines Updates Team undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the guideline in collaboration with the committee. For details please see the methods section of the evidence review.
Sources of funding/support	The NICE Guideline Updates Team is an internal team within NICE.
Name of sponsor	The NICE Guideline Updates Team is an internal team within NICE.
Roles of sponsor	The NICE Guideline Updates Team is an internal team within NICE.

Appendix B – Methods

Priority screening

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

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

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

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

Incorporating published systematic reviews

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

Quality assessment

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

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

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

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

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

Using systematic reviews as a source of data

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

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

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

Evidence of effectiveness of interventions

Quality assessment

Individual RCTs and quasi-randomised controlled trials were quality assessed using the Cochrane Risk of Bias Tool. Other study was quality assessed using the ROBINS-I tool. Each individual study was classified into one of the following three groups:

- Low risk of bias – The true effect size for the study is likely to be close to the estimated effect size.
- Moderate risk of bias – There is a possibility the true effect size for the study is substantially different to the estimated effect size.
- High risk of bias – It is likely the true effect size for the study is substantially different to the estimated effect size.

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

- Direct – No important deviations from the protocol in population, intervention, comparator and/or outcomes.
- Partially indirect – Important deviations from the protocol in one of the populations, intervention, comparator and/or outcomes.
- Indirect – Important deviations from the protocol in at least two of the following areas: population, intervention, comparator and/or outcomes.

Methods for combining intervention evidence

Meta-analyses of interventional data were conducted with reference to the Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al. 2011).

Where different studies presented continuous data measuring the same outcome but using different numerical scales (e.g. a 0-10 and a 0-100 visual analogue scale), these outcomes were all converted to the same scale before meta-analysis was conducted on the mean differences. Where outcomes measured the same underlying construct but used different instruments/metrics, data were analysed using standardised mean differences (Hedges' g).

A pooled relative risk was calculated for dichotomous outcomes (using the Mantel–Haenszel method). Both relative and absolute risks were presented, with absolute risks calculated by applying the relative risk to the pooled risk in the comparator arm of the meta-analysis.

Fixed- and random-effects models (der Simonian and Laird) were fitted for all syntheses, with the presented analysis dependent on the degree of heterogeneity in the assembled evidence. Fixed-effects models were the preferred choice to report, but in situations where the assumption of a shared mean for fixed-effects model were clearly not met, even after appropriate pre-specified subgroup analyses were conducted, random-effects results are presented. Fixed-effects models were deemed to be inappropriate if one or both of the following conditions was met:

- Significant between study heterogeneity in methodology, population, intervention or comparator was identified by the reviewer in advance of data analysis. This decision was made and recorded before any data analysis was undertaken.
- The presence of significant statistical heterogeneity in the meta-analysis, defined as $I^2 \geq 50\%$.

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

Meta-analyses were performed in Cochrane Review Manager v5.3.

Minimal clinically important differences (MIDs)

The Core Outcome Measures in Effectiveness Trials (COMET) database was searched to identify published minimal clinically important difference thresholds relevant to this guideline. MIDs were assessed to ensure they had been developed and validated in a methodologically rigorous way, and were applicable to the populations, interventions and outcomes specified in this guideline. No MIDs were identified through this process. In addition, the Guideline Committee were asked to prospectively specify any outcomes where they felt a consensus MID could be defined from their experience. The committee agreed that any difference in mortality would be clinically meaningful, and therefore the line of no effect was used as an MID. The committee chose not to specify any other MIDs by consensus.

For continuous outcomes expressed as a mean difference where no other MID was available, an MID of 0.5 of the median standard deviations of the comparison group arms was used (Norman et al. 2003). For continuous outcomes expressed as a standardised mean difference where no other MID was available, an MID of 0.5 was used. For relative risks where no other MID was available, a default MID interval for dichotomous outcomes of 0.8 to 1.25 was used. For hazard ratios where no other MID was available, no MIDs were set and the line of no effect was used to assess meaningful differences. However, the committee agreed that any difference in mortality would be clinically meaningful, and therefore the line of no effect was used as an MID.

The ‘Evidence to Recommendations’ section of each review makes explicit the committee’s view of the expected clinical importance and relevance of the findings. In particular, this includes consideration of whether the whole effect of a treatment (which may be felt across multiple independent outcome domains) would be likely to be clinically meaningful, rather than simply whether each individual sub outcome might be meaningful in isolation.

GRADE for pairwise meta-analyses of interventional evidence

GRADE was used to assess the quality of evidence for the selected outcomes as specified in ‘Developing NICE guidelines: the manual (2014)’. Data from all study designs was initially rated as high quality and the quality of the evidence for each outcome was downgraded or not from this initial point, based on the criteria given in Table 7.

Table 7: Rationale for downgrading quality of evidence for intervention studies

GRADE criteria	Reasons for downgrading quality
Risk of bias	Not serious: If less than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from studies at high risk of bias, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies at high and low risk of bias.
Indirectness	Not serious: If less than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from indirect studies, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between direct and indirect studies.
Inconsistency	Concerns about inconsistency of effects across studies, occurring when there is unexplained variability in the treatment effect demonstrated across studies (heterogeneity), after appropriate pre-specified subgroup analyses have been conducted. This was assessed using the I^2 statistic. N/A: Inconsistency was marked as not applicable if data on the outcome was only available from one study. Not serious: If the I^2 was less than 33.3%, the outcome was not downgraded. Serious: If the I^2 was between 33.3% and 66.7%, the outcome was downgraded one level. Very serious: If the I^2 was greater than 66.7%, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies with the smallest and largest effect sizes.
Imprecision	If an MID other than the line of no effect was defined for the outcome, the outcome was downgraded once if the 95% confidence interval for the effect size crossed one line of the MID, and twice if it crosses both lines of the MID. If the line of no effect was defined as an MID for the outcome, it was downgraded once if the 95% confidence interval for the effect size crossed the line of no effect (i.e. the outcome was not statistically significant), and twice if the sample size of the study was sufficiently small that it is not plausible any realistic effect size could have been detected. Outcomes meeting the criteria for downgrading above were not downgraded if the confidence interval was sufficiently narrow that the upper and lower bounds would correspond to clinically equivalent scenarios.

Publication bias

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

Evidence statements

For outcomes with a defined MID, evidence statements were divided into 4 groups as follows:

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

For outcomes without a defined MID or where the MID is set as the line of no effect (for example, in the case of mortality), evidence statements are divided into 2 groups as follows:

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

Diagnostic test accuracy evidence

In this guideline, diagnostic test accuracy (DTA) data are classified as any data in which a feature – be it a symptom, a risk factor, a test result or the output of some algorithm that combines many such features – is observed in some people who have the condition of interest at the time of the test and some people who do not. Such data either explicitly provide, or can be manipulated to generate, a 2x2 classification of true positives and false negatives (in people who, according to the reference standard, truly have the condition) and false positives and true negatives (in people who, according to the reference standard, do not).

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

- **Positive likelihood ratios** describe how many times more likely positive features are in people with the condition compared to people without the condition. Values greater than 1 indicate that a positive result makes the condition more likely.
  - LR^+ = (TP/[TP+FN])/(FP/[FP+TN])

- **Negative likelihood ratios** describe how many times less likely negative features are in people with the condition compared to people without the condition. Values less than 1 indicate that a negative result makes the condition less likely.
  - LR^- = (FN/[TP+FN])/(TN/[FP+TN])
- **Sensitivity** is the probability that the feature will be positive in a person with the condition.
  - sensitivity = TP/(TP+FN)
- **Specificity** is the probability that the feature will be negative in a person without the condition.
  - specificity = TN/(FP+TN)

Interpretation of diagnostic accuracy measures

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

Table 8: Interpretation of likelihood ratios

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

The schema above has the effect of setting a minimal important difference for positive likelihoods ratio at 2, and a corresponding minimal important difference for negative likelihood ratios at 0.5. Likelihood ratios (whether positive or negative) falling between these thresholds were judged to indicate no meaningful change in the probability of disease.

Quality assessment

Individual studies were quality assessed using the QUADAS-2 tool, which contains four domains: patient selection, index test, reference standard, and flow and timing. Each individual study was classified into one of the following two groups:

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

Each individual study was also classified into one of three groups for directness, based on if there were concerns about the population, index features and/or reference standard in the

study and how directly these variables could address the specified review question. Studies were rated as follows:

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

Methods for combining diagnostic test accuracy evidence

Meta-analysis of diagnostic test accuracy data was conducted with reference to the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (Deeks et al. 2010).

Where applicable, diagnostic syntheses were stratified by:

- Presenting symptomatology (features shared by all participants in the study, but not all people who could be considered for a diagnosis in clinical practice).
- The reference standard used for true diagnosis.

Where five or more studies were available for all included strata, a bivariate model was fitted using the `mada` package in R v3.4.0, which accounts for the correlations between positive and negative likelihood ratios, and between sensitivities and specificities. Where sufficient data were not available (2-4 studies), separate independent pooling was performed for positive likelihood ratios, negative likelihood ratios, sensitivity and specificity, using Microsoft Excel. This approach is conservative as it is likely to somewhat underestimate test accuracy, due to failing to account for the correlation and trade-off between sensitivity and specificity (see Deeks 2010).

Random-effects models (der Simonian and Laird) were fitted for all syntheses, as recommended in the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (Deeks et al. 2010).

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

Modified GRADE for diagnostic test accuracy evidence

GRADE has not been developed for use with diagnostic studies; therefore a modified approach was applied using the GRADE framework. GRADE assessments were only undertaken for positive and negative likelihood ratios, as the MIDs used to assess imprecision were based on these outcomes, but results for sensitivity and specificity are also presented alongside those data.

Cross-sectional and cohort studies were initially rated as high-quality evidence if well conducted, and then downgraded according to the standard GRADE criteria (risk of bias, inconsistency, imprecision and indirectness) as detailed in Table 9 below.

Table 9: Rationale for downgrading quality of evidence for diagnostic questions

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

Publication bias

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

Evidence statements

Evidence statements were written for positive and negative likelihood ratios and indicate the magnitude of effect on the probability of having a PE (based on the categories in Table 8) associated with a positive test result or a negative test result with a quality rating for each finding. The evidence for sensitivity and specificity is presented for the main analysis only and does not contain a quality rating as this has been assessed at the LR level.

Appendix C – Literature search strategies

Searches were run on 16th May 2018 in Medline, Medline in Process, Medline Epub Ahead of Print, Embase (all Ovid platform), Cochrane Database of Systematic Reviews, CENTRAL and DARE (all Wiley platform). The searches were re run on 4th April 2019.

The Medline strategy is presented below. This was translated for the other databases.

1 exp pulmonary embolism/
2 ((pulmonary or lung) adj4 (embol* or thromboembo* or microembol*)).tw.
3 (pulmonary adj infarction).tw
4 or/1-3
5 (perc or "pulmonary embolism rule-out criteria" or "pulmonary embolism rule out criteria" or "pe rule-out criteria" or "pe rule out criteria").tw.
6 "clinical decision".tw
7 ((clinical or decision) adj2 (tool* or criteria or rule*)).tw.
8 or/5-7
9 4 and 8 (439)
10 animals/ not humans/
11 9 not 10
12 limit 11 to english language

Searches to identify economic evidence were run on 18th May 2018 in in Medline, Medline in Process, Econlit and Embase (all va the Ovid platform), NHS EED and the Health Technology Database (via the Wiley platform. NICE inhouse economic evaluation and Quality of Life filters were attached to the Medline and Embase strategies of the above search. A single search to identify economic evidence across all questions was re run on 9th April 2019.The Medline versions of the filters is displayed below

Economic evaluations

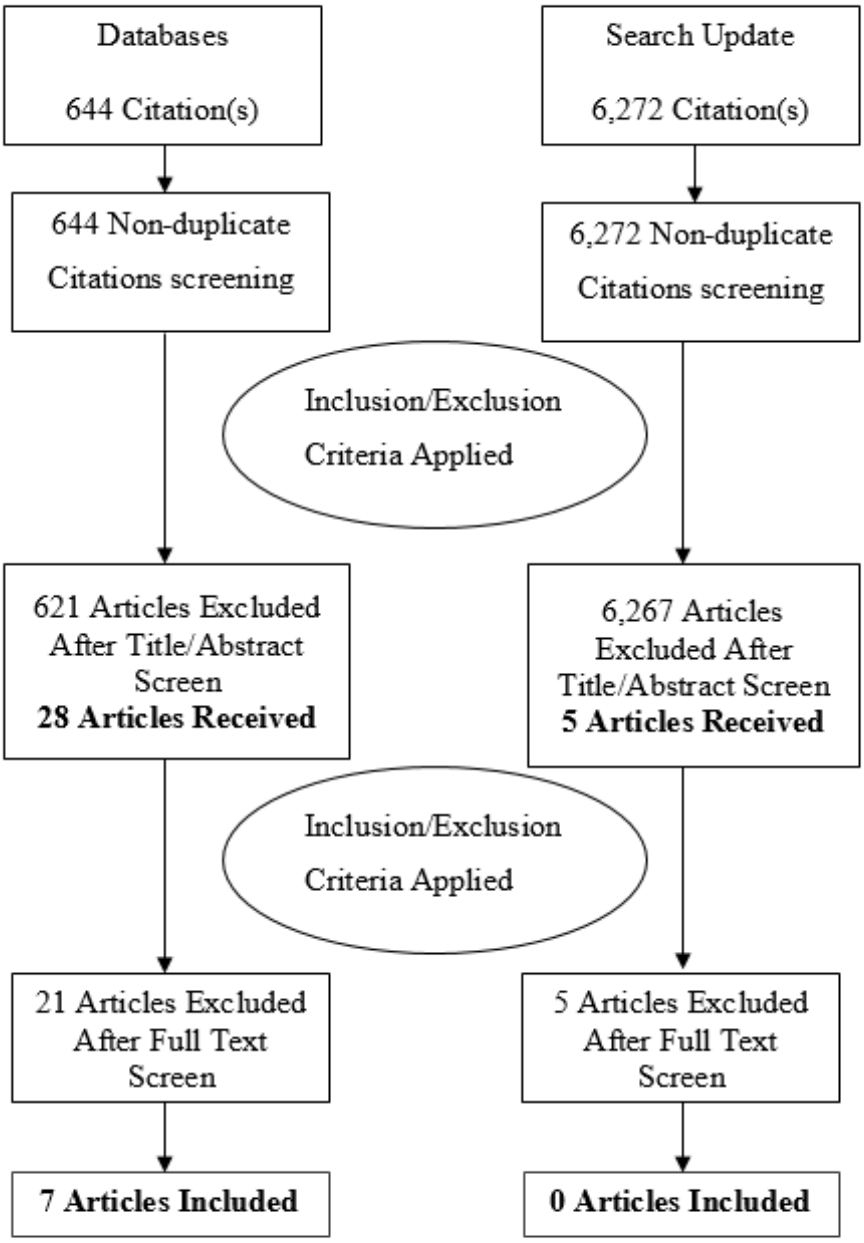
1 Economics/
2 exp "Costs and Cost Analysis"/
3 Economics, Dental/
4 exp Economics, Hospital/
5 exp Economics, Medical/
6 Economics, Nursing/
7 Economics, Pharmaceutical/
8 Budgets/
9 exp Models, Economic/
10 Markov Chains/
11 Monte Carlo Method/
12 Decision Trees/
13 econom$.tw.
14 cba.tw.
15 cea.tw.
16 cua.tw.
17 markov$.tw.
18 (monte adj carlo).tw.
19 (decision adj3 (tree$ or analys$)).tw.

20 (cost or costs or costing$ or costly or costed).tw.
21 (price$ or pricing$).tw.
22 budget$.tw.
23 expenditure$.tw.
24 (value adj3 (money or monetary)).tw.
25 (pharmacoeconomic$ or (pharmaco adj economic$)).tw.
26 or/1-25

Quality of Life

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

Appendix D – Clinical evidence study selection



Appendix E – Clinical evidence tables

Randomised controlled trial

Author (year)	Title	Study details	Risk of bias
Freund (2018)	Effect of the Pulmonary Embolism Rule-Out Criteria on Subsequent Thromboembolic Events Among Low-Risk Emergency Department Patients: The PROPER Randomized Clinical Trial	**Study type** • Cluster randomised controlled trial *Crossover trial: in each cluster there were 2 6-month periods, separated by a 2 month 'washout' period. PERC was used in 1 of the 2 periods, allocated at random.* **Study details** • Study location *France* • Study setting *14 Emergency departments* • Study dates *Trial recruitment began in August 2015, ended in September 2016, and follow-up ended in December 2016.* • Duration of follow-up *3 months: all patients were interviewed by phone at the end of this period.* • Sources of funding *Programme Hospitalier de Recherche Clinique-PHRC 2014 (Ministère de la Santé, Paris, France).* **Inclusion criteria** • Suspicion of PE *All patients presented to the emergency department with a*	**Random sequence generation** • Low risk of bias *Randomization was computer generated in blocks,* **Allocation concealment** • Unclear risk of bias **Blinding of participants and personnel** • High risk of bias *Participants and personnel not blinded. The number of eligible participants available that were not enrolled was not reported. There were more very low probability patients enrolled in the PERC period, suggesting that there may have been inclusion bias introduced.* **Blinding of outcome assessment** • Low risk of bias *Experts assessing outcomes were blinded to strategy allocation.*

Author (year)	Title	Study details	Risk of bias
		suspicion of PE were eligible for inclusion. • Low clinical probability of PE *Estimated by the treating physician's gestalt as an expectation below 15% probability of PE.* • Symptoms of PE *New onset or worsening of shortness of breath or chest pain* **Exclusion criteria** • Other obvious etiology *For example, pneumothorax or acute coronary syndrome* • Acute severe presentation • Contraindication to CTPA • Pregnancy • Inability to be followed up • Receiving anticoagulant treatment **Sample characteristics** • Sample size *14 emergency departments, including 1916 participants* • Split between study groups *PERC period: 962 Control period: 954* • Loss to follow-up *An intention to treat analysis was reported by the study in which all patients were included. However this analysis was calculated by considering the worst-case scenario - that all patients lost to follow up experienced the event (e.g. mortality, VTE). This has the effect of artificially reducing the*	**Incomplete outcome data** • Low risk of bias *Loss to follow up small and similar across groups.* **Selective reporting** • Low risk of bias **Other sources of bias** • Low risk of bias *Cluster RCT: sequence effect assessed and not present, intracluster correlation reported and low* **Overall risk of bias** • Moderate **Directness** • Directly applicable

Author (year)	Title	Study details	Risk of bias
		confidence intervals by increasing the event rate, therefore we present a different ITT analysis, where participants lost to follow up were assumed not to have experienced the event. 48 participants were lost to follow up - 25 during the PERC period, 23 during the control period [numbers from taken from figure - inconsistent with total number reported in the text]. • %female *PERC: 48% Control: 54%* • Mean age (SD) *PERC: 44 (17) Control: 45(17)* • Active malignancy *PERC: 8 Control: 10* **Interventions** • PERC • Control **Outcome measure(s)** • All cause mortality *Assessed at 3 months follow up* • VTE related mortality *An adjudication committee of 3 experts (blinded to allocation strategy) adjudicated all deaths as to whether or not they were likely to have been related to a PE.* • Diagnostic strategy failure *Defined as the occurrence of a symptomatic thromboembolic event during the 3-month follow up period*	

Author (year)	Title	Study details	Risk of bias
		which was not diagnosed at the time of the inclusion visit. • Length of hospital stay *Length of stay in the emergency department* • Adverse event: Major bleeding	

Diagnostic accuracy studies

Author (year)	Title	Study details	Risk of bias
Crichlow (2012)	Overuse of computed tomography pulmonary angiography in the evaluation of patients with suspected pulmonary embolism in the emergency department	**Study type** • Prospective cohort study **Study details** • Study location *USA* • Study setting *1 hospital emergency department.* • Study dates *December 2009 to May 2010.* • Loss to follow-up *8 participants were lost to follow-up.* • Sources of funding *This study was supported by grant K12HL087064 of the National Heart, Lung, and Blood Institute, National Institutes of Health.* **Inclusion criteria** • Age *18 years or older.* • CT-PA *Underwent CT-PA for suspected PE as part of their ED evaluations.* *Study did not require that participants were at low-risk however 72% were low-risk according to Well's score.	**Patient selection** • Low risk of bias **Index test** • Low risk of bias **Reference standard** • High risk of bias *Unclear if interpretation of reference tests was blinded to index test results in all patients.* **Flow and timing** • Low risk of bias **Overall risk of bias** • Moderate *Downgraded due to unclear blinding status for interpretation of reference standard.* **Directness** • Directly applicable

Author (year)	Title	Study details	Risk of bias
		Exclusion criteria • Previous diagnoses *Previous diagnoses of acute PE or DVT within four weeks of presentation to the ED.* • Contact details *Patients who did not provide contact home, cell, or work phone numbers for the 90-day follow-up.* • Informed consent *Patients unable to provide informed consent.* **Sample characteristics** • Sample size *166 patients were enrolled and 152 patients were analysed.* • %female *112 (73.7%) were female* • Mean age (SD) *46.3 (SD 15.6) years.* • Pre-test probability *110 (72%) patients had a Wells score ≤4 (low risk). Data (n, %) are reported for individual Wells criteria, by PE-negative and PE-positive patients.* • % Cancer *31 (20.4%) patients had active cancer* **Index test(s)** • PERC *NOTE: PERC domains not the same as others? e.g. heart rate*	

Author (year)	Title	Study details	Risk of bias
		>99 vs 110 **Reference standard(s)** • Composite *CT-PA plus all patients with CT-PAs negative for PE were followed-up for 90 days after enrolment.*	
Hogg (2005)	Application of pulmonary embolism rule-out criteria to the UK Manchester Investigation of Pulmonary Embolism Diagnosis (MIOPED) study cohort	**Study type** • Prospective cohort study **Study details** • Study location *UK* • Study setting *Hospital emergency department* • Study dates *February 2002 to May 2003* • Loss to follow-up *There was complete data for all patients.* **Inclusion criteria** • Age *18 years and older* • Clinical presentation *Pleuritic chest pain*	**Patient selection** • Unclear risk of bias *Not enough information provided to assess for low or high risk of bias (article is letter to the Editor, not full article).* **Index test** • High risk of bias *Unclear if interpretation of index test was blinded to reference test results.* **Reference standard** • High risk of bias *D-dimer was used to exclude PE. Unclear if interpretation of reference tests was blinded to index test results in all patients.* **Flow and timing** • Unclear risk of bias

Author (year)	Title	Study details	Risk of bias
		*Study did not require that participants were at low-risk however *88% of patients scored a low Wells' score* **Exclusion criteria** • Clinical presentation *Pneumothorax, electrocardiogram changes of myocardial infarction, ischemia or pericarditis, pregnancy or trauma within 4 weeks.* • Already included *Patients previously recruited to the main cohort study.* **Sample characteristics** • Sample size *425 patients* • %female *51.1% were female.* • Mean age (SD) *38.3 (SD 15.0) years* • Pre-test probability *88% of patients scored a low Wells' score, 8.7% moderate and 3.3% high.* **Index test(s)** • PERC **Reference standard(s)** • Composite	*Not enough information provided to assess for low or high risk of bias (article is letter to the Editor, not full article).* **Overall risk of bias** • High *Downgraded due to risk of bias from potential unblinded interpretation of reference and index tests.* **Directness** • Partially directly applicable *Only included patients with chest pain and not all patients with suspected PE.*

Author (year)	Title	Study details	Risk of bias
		IL D-dimer test combined with low clinical probability, ventilation-perfusion scan with/without low clinical probability, CT pulmonary angiograph or digital subtraction pulmonary angiography. All patients were followed-up clinically for 3 months.	
Hugli (2011)	The pulmonary embolism rule-out criteria (PERC) rule does not safely exclude pulmonary embolism. [Erratum appears in J Thromb Haemost. 2012 Apr;10(4):740]	**Study type** • Prospective cohort study *PERC rule was applied retrospectively.* **Study details** • Study location *Switzerland, France and Belgium.* • Study setting *Hospital emergency departments.* • Study dates *1 January 2005 to 30 August 2006.* • Loss to follow-up *1 patient was lost to follow-up.* **Inclusion criteria** • Age *Adults.* • Clinical presentation *Treated in the emergency department with a clinical suspicion of PE.*	**Patient selection** • Low risk of bias **Index test** • High risk of bias *PERC calculated retrospectively from prospectively collected data.* **Reference standard** • Low risk of bias **Flow and timing** • Low risk of bias **Overall risk of bias** • Moderate *Downgraded due to risk of bias from lack of blinding of reference standard results when*

Author (year)	Title	Study details	Risk of bias
		*Study did not require that participants were at low risk. 35% of participants were low risk according to Well's score. **Exclusion criteria** • Previous diagnoses *Previous documented diagnosis of PE or receiving anticoagulant therapy at presentation.* • Contraindication *Contraindication to helical multidetector computed tomography (i.e. allergy to iodine contrast agents, creatinine clearance < 30 mL min)1 or pregnancy)* • Terminal illness *terminal illness with an expected survival of < 3 months.* **Sample characteristics** • Sample size *1693 patients were included in the per-protocol analysis of the original clinical trial analysis. 1675 patients were included in the present analysis. 17 patients were excluded as the PERC rule could not be assessed and 1 patient was lost to follow-up. All patients: 1675 Pretest probability subgroups: Low: 587 (35%) Intermediate: 1038 (62%) High: 50 (3%)* • %female *All patients: 917 (54.7%) were female. Sex was not reported for pretest probability subgroups.* • Mean age (SD) *All patients, median (IQR): 61 (45 to 76) years. Age was not reported for pretest probability subgroups.* • Pre-test probability	*interpreting index test results.* **Directness** • Directly applicable

Author (year)	Title	Study details	Risk of bias
		Clinical pretest probability based on the revised Geneva score. Low: 587 (35%) Intermediate: 1038 (62%) High: 50 (3%) • % Cancer *Active malignancy: 126 (7.5%).* **Index test(s)** • PERC **Reference standard(s)** • Composite *MDCT, pulmonary angiography, ventilation/perfusion lung scan or a proximal deep vein thrombosis (DVT) documented by compression ultrasonography. A 3-month follow-up was conducted to establish subsequent diagnoses of venous thromboembolic events after discharge.*	
Kline (2004)	Clinical criteria to prevent unnecessary diagnostic testing in emergency department patients with suspected pulmonary embolism	**Study type** • Prospective cohort study *NOTE to Sarah - I have only extracted data for the 'validation' part of the study, and only for the 'low risk' population - the other part of the study doesn't match the protocol and neither does the very low risk population (as PE was not initially suspected and our population is people with suspected PE)* **Study details** • Study location *USA.*	**Patient selection** • Unclear risk of bias *Not enough information reported to judge as high or low risk of bias.* **Index test** • High risk of bias *PERC calculated retrospectively from prospectively collected data.*

Author (year)	Title	Study details	Risk of bias
		• Study setting *2 hospital emergency departments.* • Study dates *1 January 2001 to 30 June 2003.* • Loss to follow-up *There were no [patients lost to follow-up.* **Inclusion criteria** • Clinical presentation *Patients who underwent evaluation for possible pulmonary embolism. Judged by the clinician to be of low enough risk to justify exclusion of pulmonary embolism on the basis of a negative D-dimer.* **Exclusion criteria** • None reported **Sample characteristics** • Sample size *Low-risk patients: 1427* • %female *Low-risk patients: 852 (60%) were female.* • Mean age (SD) *Low-risk patients: 47 (SD 17) years.* • Pre-test probability *Patients were defined as low-risk according the clinicians' belief that patients were at low enough risk to justify exclusion of*	**Reference standard** • Low risk of bias **Flow and timing** • Low risk of bias **Overall risk of bias** • Moderate *Downgraded due to unclear risk of bias from patient selection and possible lack of blinding of reference standard results when interpreting index test results.* **Directness** • Directly applicable

Author (year)	Title	Study details	Risk of bias
		pulmonary embolism on the basis of a negative D-dimer. A quantitative measure of pretest probability was not used. • % Cancer *Low-risk patients: 176 (12%) had prior or current malignancy.* **Index test(s)** • PERC **Reference standard(s)** • Composite *D-dimer. CT angiography, CT angiography-venography or ventilation-perfusion lung scanning with selected use of venous ultrasonography was performed if D-dimer testing was abnormal. Each patient was followed-up for 90 days after enrolment using the combination of telephone follow-up, examination of medical records and contact with the patient's personal physician.*	
Penaloza (2012)	Performance of the Pulmonary Embolism Rule-out Criteria (the PERC rule) combined with low clinical probability in high prevalence population	**Study type** • Prospective cohort study *Retrospective analysis of a prospective cohort.* **Study details** • Study location *France and Belgium* • Study setting *117 emergency departments (116 in France and 1 in Belgium).* • Loss to follow-up	**Patient selection** • Low risk of bias **Index test** • High risk of bias *PERC calculated retrospectively for prospectively collected data.*

Author (year)	Title	Study details	Risk of bias
		Follow-up data was not obtained for 55 patients. **Inclusion criteria** • Clinical presentation *Patients suspected of PE.* *Study did not require that participants were at low-risk. Risk was assessed according to clinician gestalt, which estimated that 41.6% of participants were at low risk of PE. **Exclusion criteria** • Previous diagnoses *The diagnosis of thromboembolic disease was documented before admission. Patients who were anticoagulated for an initial diagnosis of DVT without PE.* • Testing cancelled *Diagnostic testing was cancelled for ethical reasons, because of rapid death, or because the patient decided to leave the hospital against medical advice or declined testing.* • Hospital stay *PE was suspected during a hospital stay of more than 2 days' duration.* **Sample characteristics** • Sample size *Original trial: 1529. All analysed patients: 959. Subgroups: Low gestalt clinical probability: 399 (41.6%) Moderate gestalt clinical*	**Reference standard** • Unclear risk of bias *Unclear which diagnostic test was used.* **Flow and timing** • High risk of bias *Approximately one third of the original sample size was excluded as information on PERC, risk score or clinical gestalt was not available. No further information on these patients was given. Not enough information reported to determine whether all patients received a reference standard.* **Overall risk of bias** • High *Downgraded due to risk of bias from lack of blinding when interpreting the index test and from the exclusion of a large number of patients due to missing data.* **Directness** • Directly applicable

Author (year)	Title	Study details	Risk of bias
		probability: 326 (34.0%) High gestalt clinical probability: 234 (24.4%) • %female *All patients analysed: 595 (62%) were female. Data were not reported by pretest probability subgroups.* • Mean age (SD) *All patients analysed: 63.9 (SD 0.6) years. Data were not reported by pretest probability subgroups.* • Pre-test probability *All patients analysed: Low gestalt clinical probability: 399 (41.6%) Moderate gestalt clinical probability: 326 (34.0%) High gestalt clinical probability: 234 (24.4%)* • % Cancer **Index test(s)** • PERC **Reference standard(s)** • Composite *Not clear which diagnostic test was used. The authors considered a final diagnosis of PE: i) a PE diagnosis ruled in at the end of the initial diagnostic work-up; ii) a thromboembolic event (PE or deep vein thrombosis) occurring during follow-up (3 months) among patients in whom the diagnosis of PE was initially ruled out or iii) death adjudicated as related or possibly related to PE.*	

Author (year)	Title	Study details	Risk of bias
Penaloza (2017)	Pulmonary embolism rule-out criteria (PERC) rule in European patients with low implicit clinical probability (PERCEPIC): a multicentre, prospective, observational study	**Study type** • Prospective cohort study **Study details** • Study location *France and Belgium.* • Study setting *12 emergency departments.* • Study dates *May 1, 2015, and April 30, 2016* • Loss to follow-up *2 patients were lost to follow-up.* • Sources of funding *The study was supported by a grant from the French Ministry of Health (PHRC 2014 API14/A/018).* **Inclusion criteria** • Age *18 years or older.* • Clinical presentation *Patients with suspected pulmonary embolism (including dyspnoea, chest pain, and other symptoms like syncope or haemoptysis) without any obvious explanation after clinical examination and first-line non-specific exams (including electrocardiogram, chest x-ray, and blood gases) that led the physician to order pulmonary embolism diagnostic tests.* • Low-risk of PE	**Patient selection** • Low risk of bias **Index test** • Low risk of bias **Reference standard** • High risk of bias *Some patients received only D-dimer as a reference standard.* **Flow and timing** • Low risk of bias **Overall risk of bias** • Moderate *Downgraded due potential risk of bias from use of D-dimer as a reference standard.* **Directness** • Directly applicable

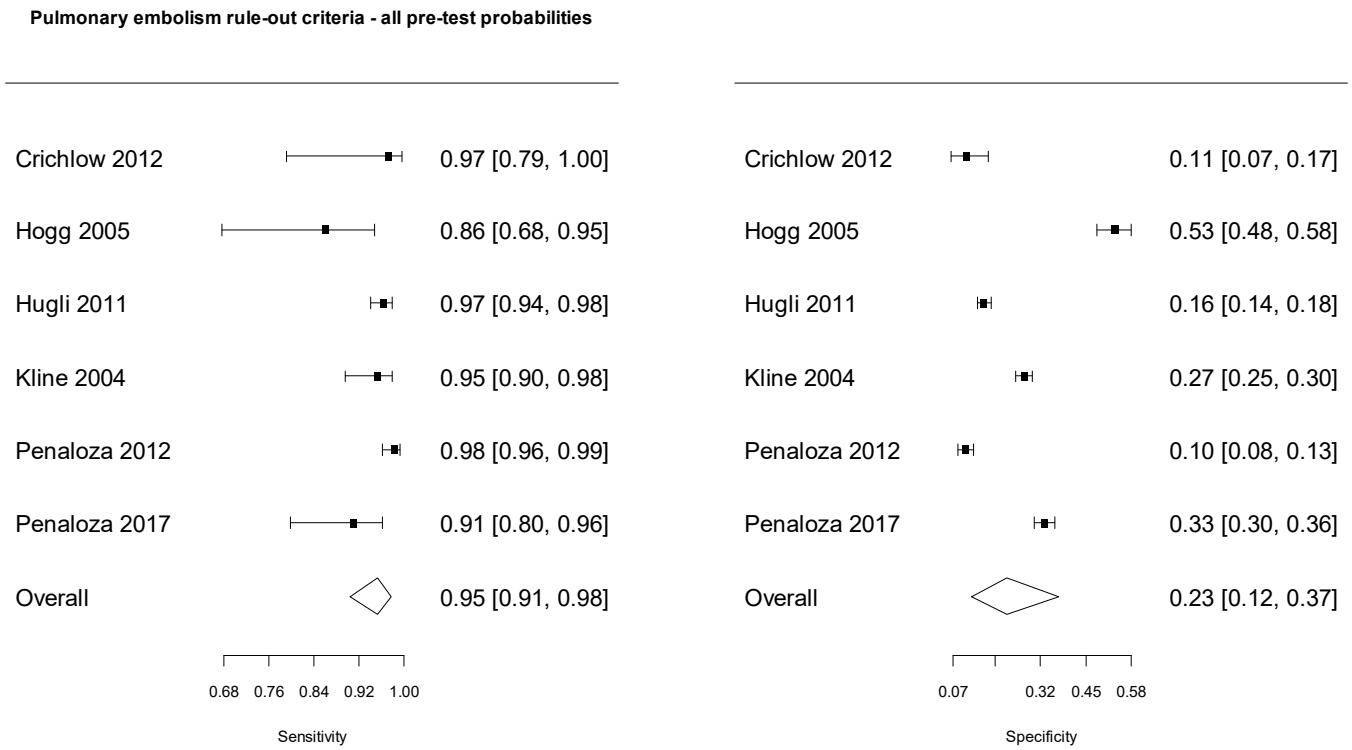
Author (year)	Title	Study details	Risk of bias
		All patients included in the analysis were classed as 'low clinical probability' (gestalt). **Exclusion criteria** • Previous diagnoses *Had a diagnosis of thromboembolic disease documented before admission to the emergency department.* • Informed consent *Refused to give consent or to be contacted by phone after 3 months.* • Hospital stay *Already hospitalised for more than 2 days.* • Coagulation treatment *Had curative anticoagulant therapy in progress for more than 2 days.* • Follow-up *Could not be followed up for 3 months.* **Sample characteristics** • Sample size *1757 patients included in the trial. 1052 patients had low clinical probability and were included in this analysis.* • %female *All patients (n=1757): 1023 (58%) were female. Data were not reported for patients included in the analysis.* • Mean age (SD) *All patients (n=1757): 53 (SD 20) years. Data were not reported for patients included in the analysis.*	

Author (year)	Title	Study details	Risk of bias
		• Pre-test probability *All patients included in the analysis were classed as 'low clinical probability' (gestalt).* • % Cancer *All patients (n=1757): 135 (8%) had cancer. Data were not reported for patients included in the analysis.* **Index test(s)** • PERC *Calculated retrospectively from prospectively collected data using a standardised form designed specifically for a trial to answer this research question.* **Reference standard(s)** • Composite *One or a combination of the following four tests: high sensitivity D-dimer test (interpreted using the age-adjusted threshold), CT pulmonary angiography (CTPA), ventilation perfusion scan, or leg ultrasonography. All patients were followed up for 3 months.*	

Appendix F – Forest plots

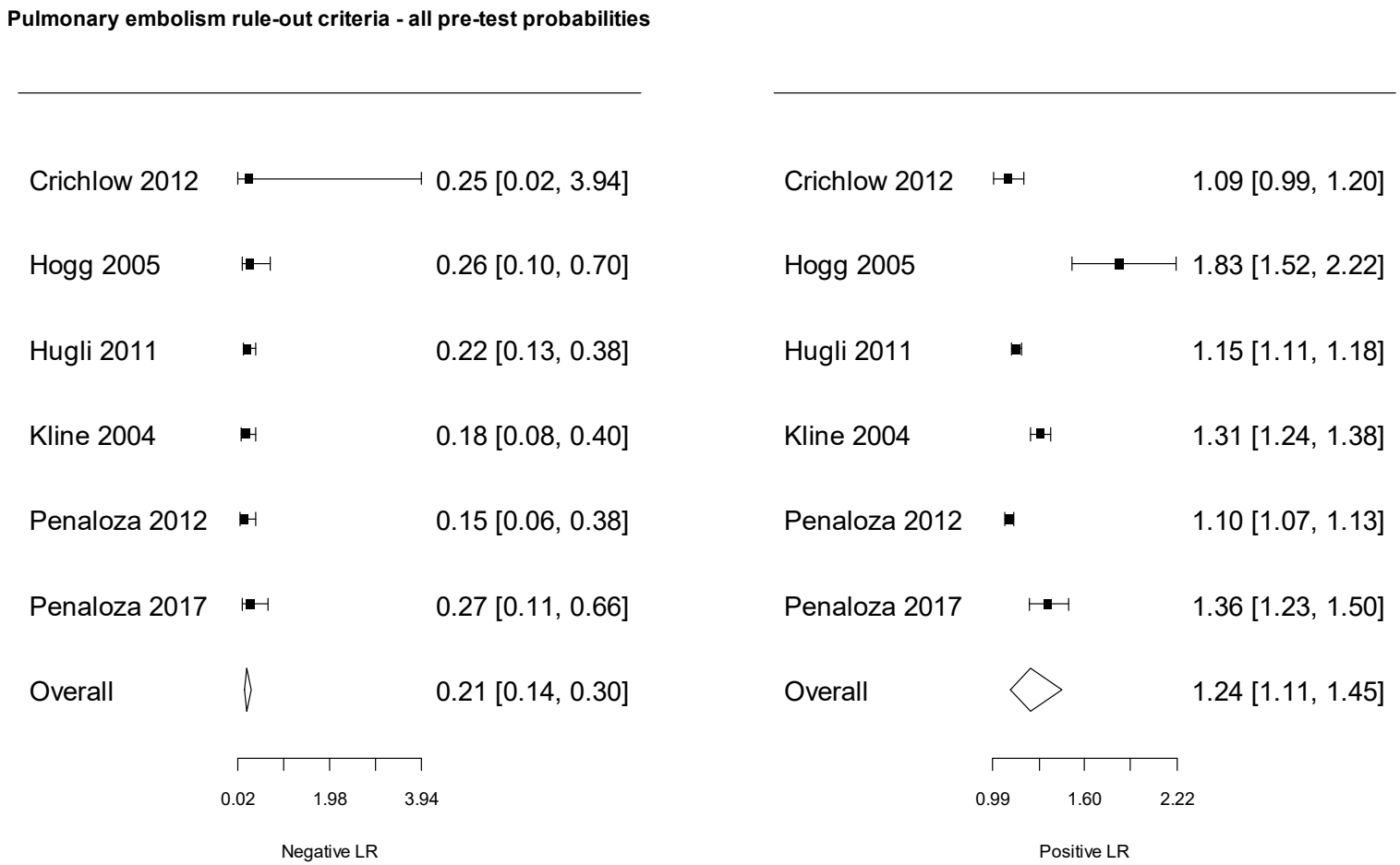
Main analysis

Figure 1: Sensitivity and specificity for pulmonary embolism rule-out criteria (all pre-test probabilities)



I^2 sensitivity=31.7% I^2 specificity=98.3%

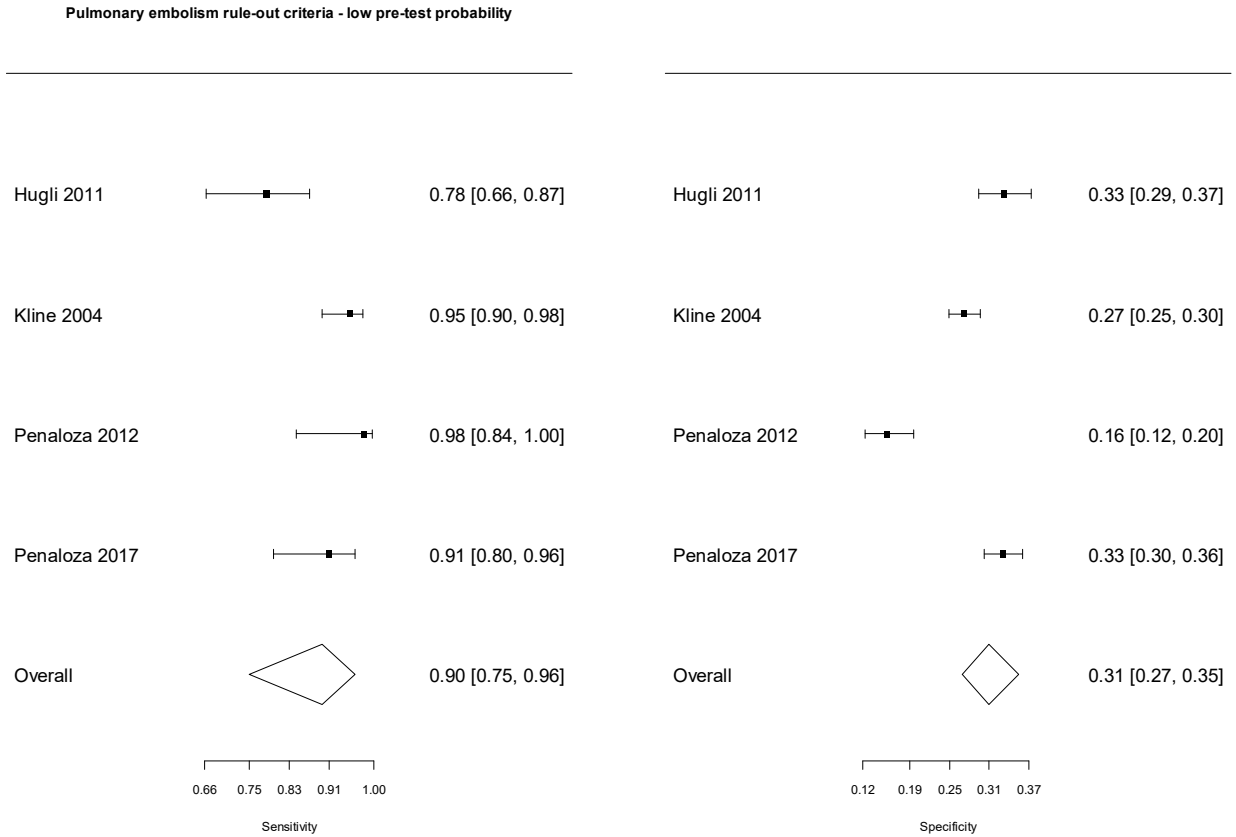
Figure 2: Likelihood ratios for pulmonary embolism rule-out criteria (all pre-test probabilities)



I^2 –ve LR= 0% I^2 +ve LR=92.6%

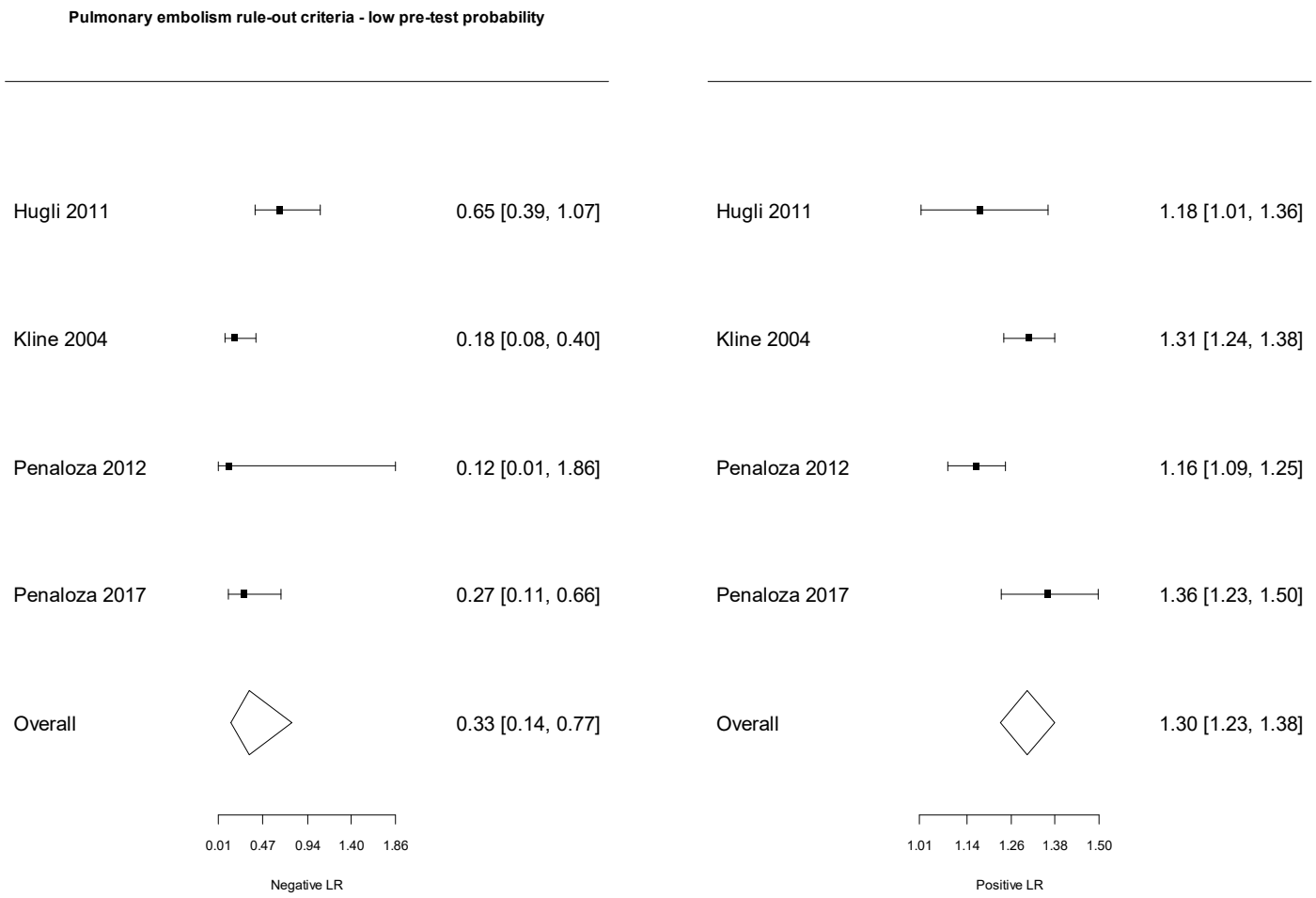
Subgroup and sensitivity analyses

Figure 3: Subgroup analysis: sensitivity and specificity for pulmonary embolism rule-out criteria (low pre-test probabilities only)



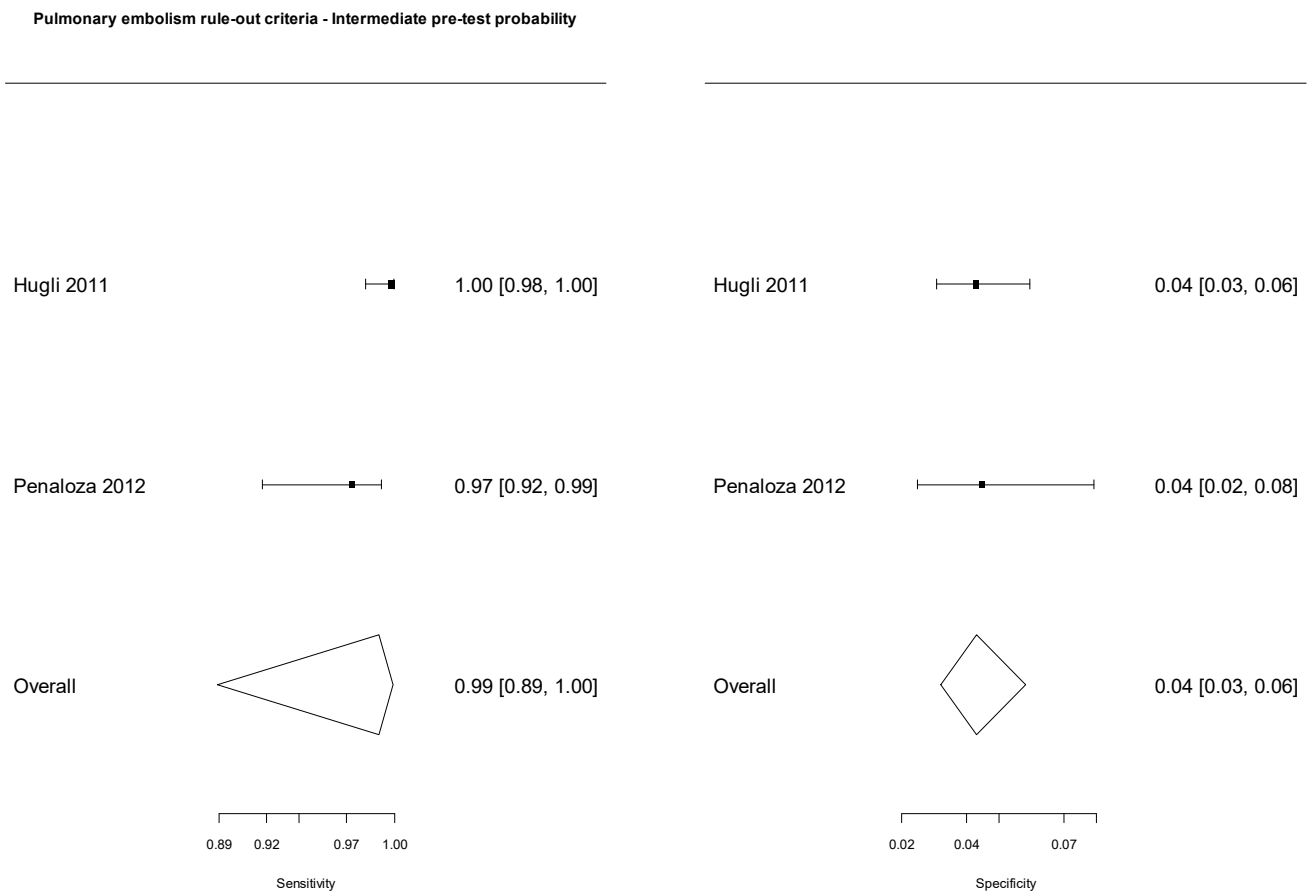
I^2 sensitivity=80.9%, I^2 specificity=83.0%

Figure 4: Subgroup analysis: Likelihood ratios for pulmonary embolism rule-out criteria (low pre-test probabilities only)



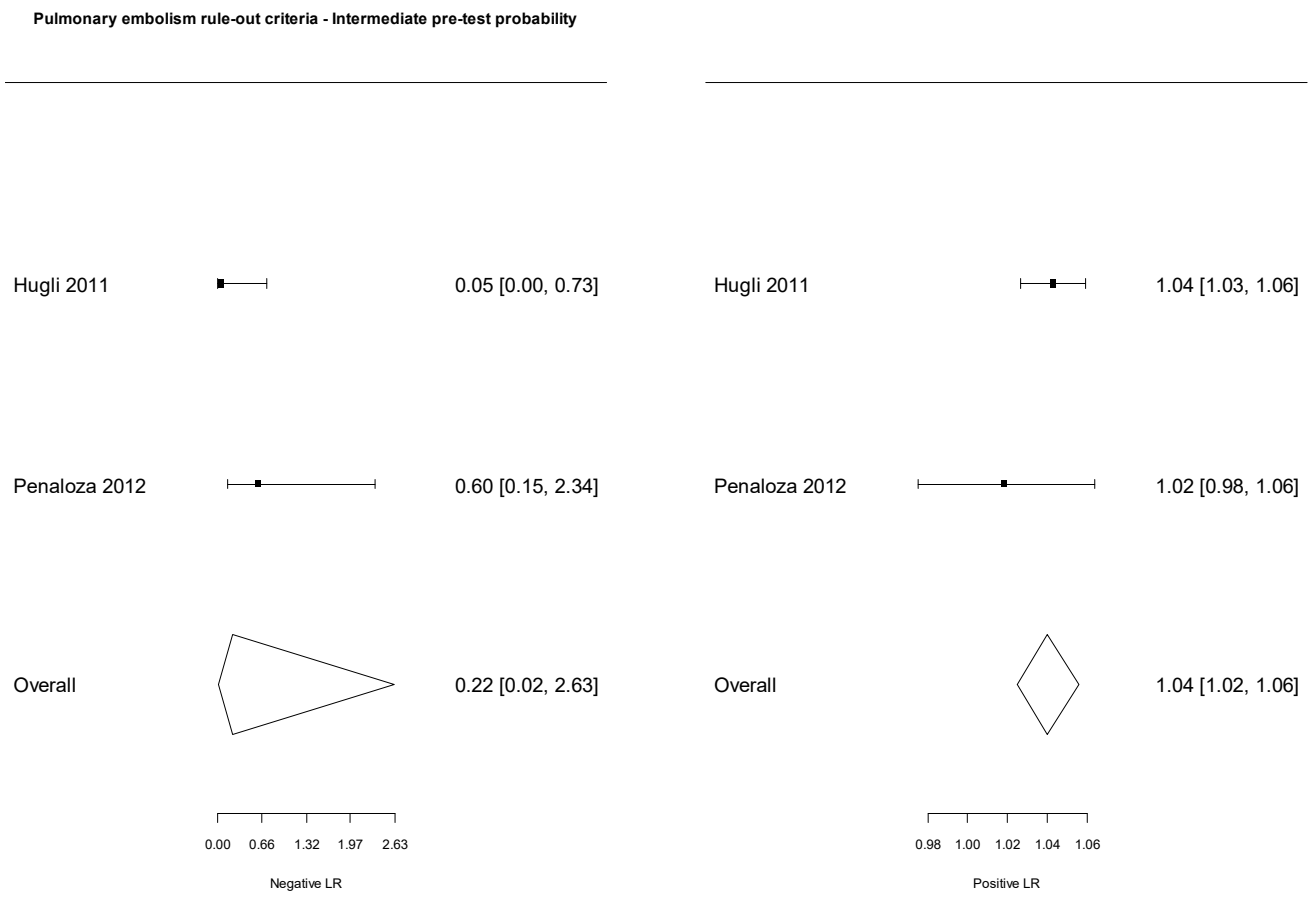
I^2 –ve LR=75.2%, I^2 +ve LR=24.0%

Figure 5: Subgroup analysis: sensitivity and specificity for pulmonary embolism rule-out criteria (intermediate pre-test probabilities only)



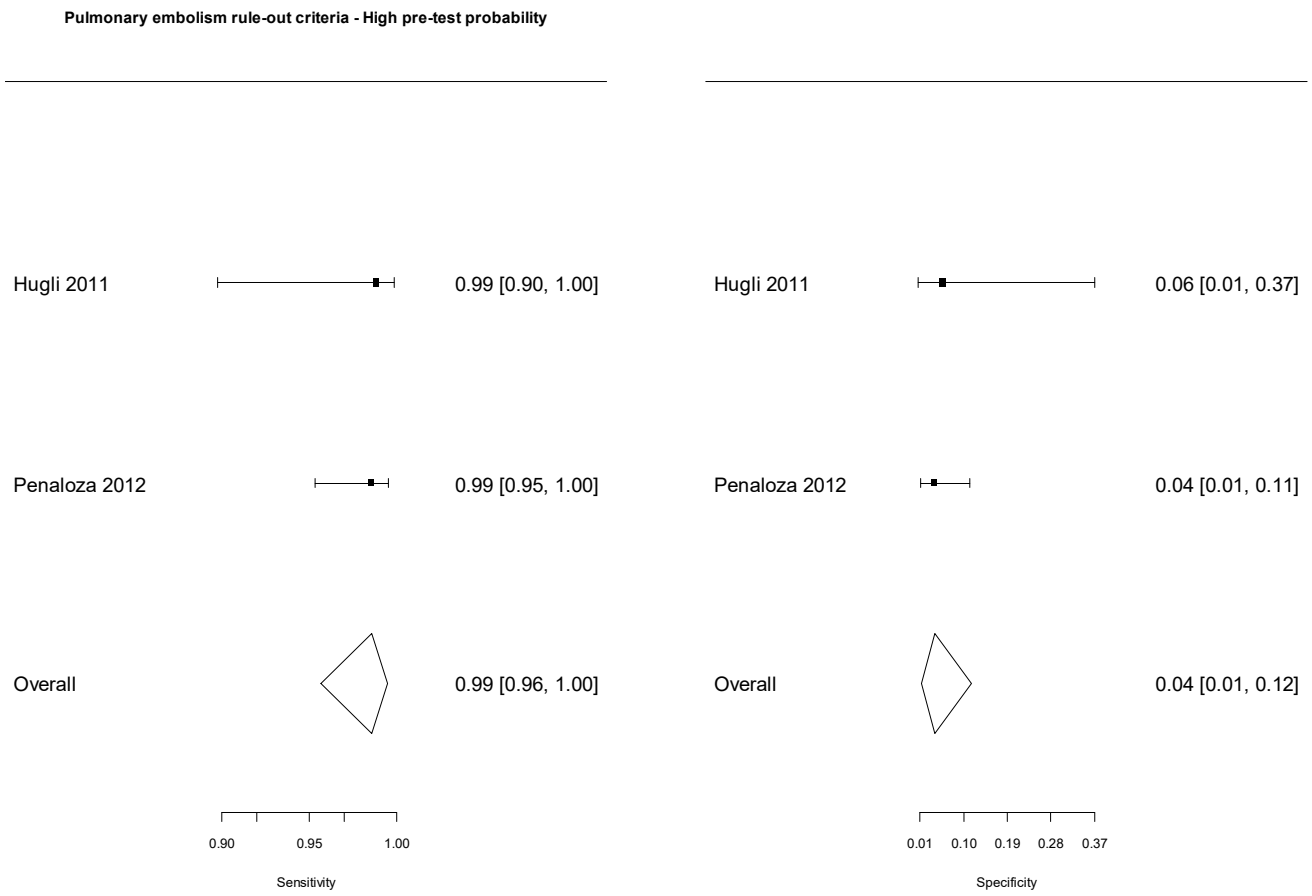
I^2 sensitivity=65.8%, I^2 specificity=0.0%

Figure 6: Subgroup analysis: likelihood ratios for pulmonary embolism rule-out criteria (intermediate pre-test probabilities only)



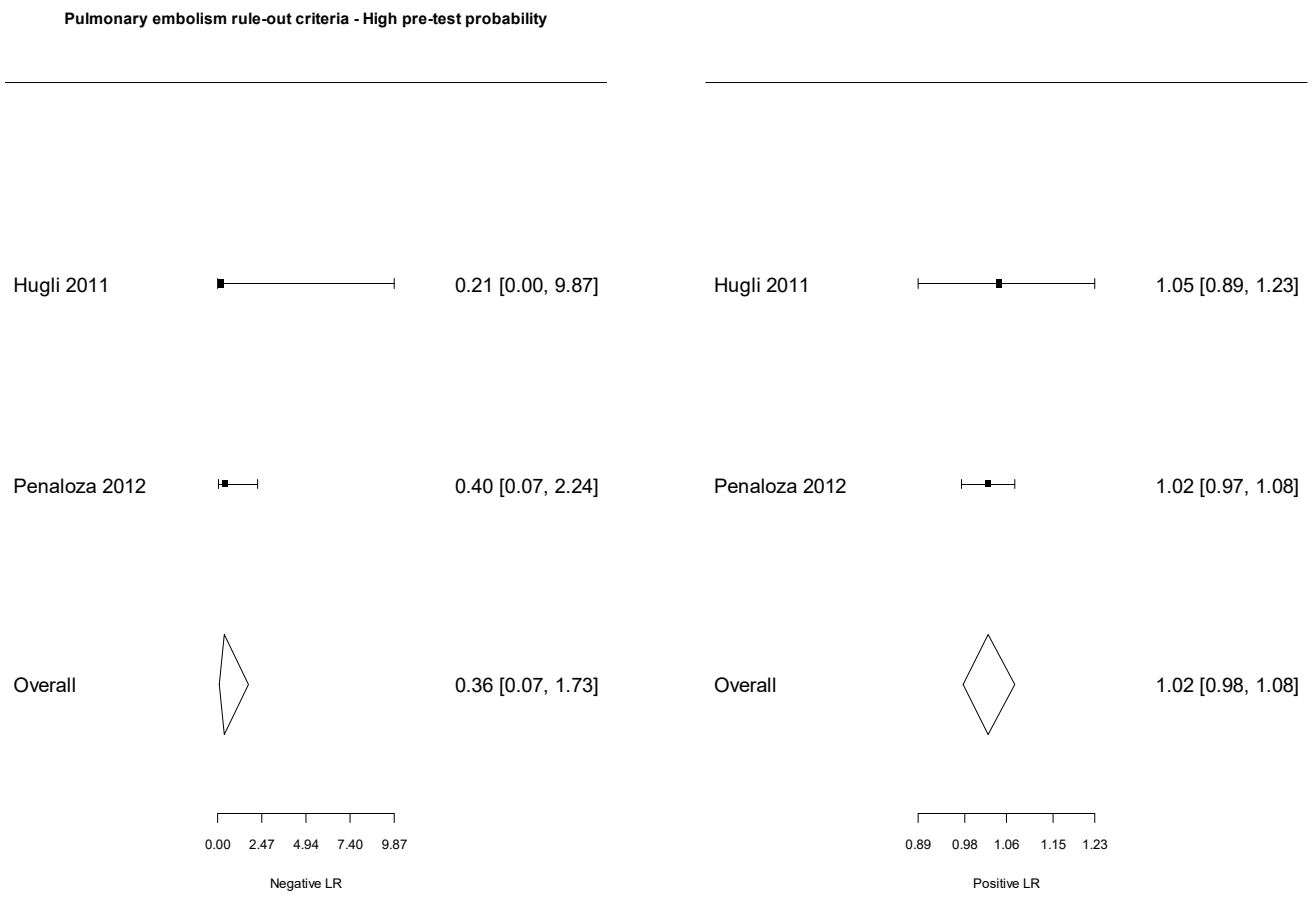
I^2 –ve LR=62.7%, I^2 +ve LR=0.0%

Figure 7: Subgroup analysis: sensitivity and specificity for pulmonary embolism rule-out criteria (high pre-test probabilities only)



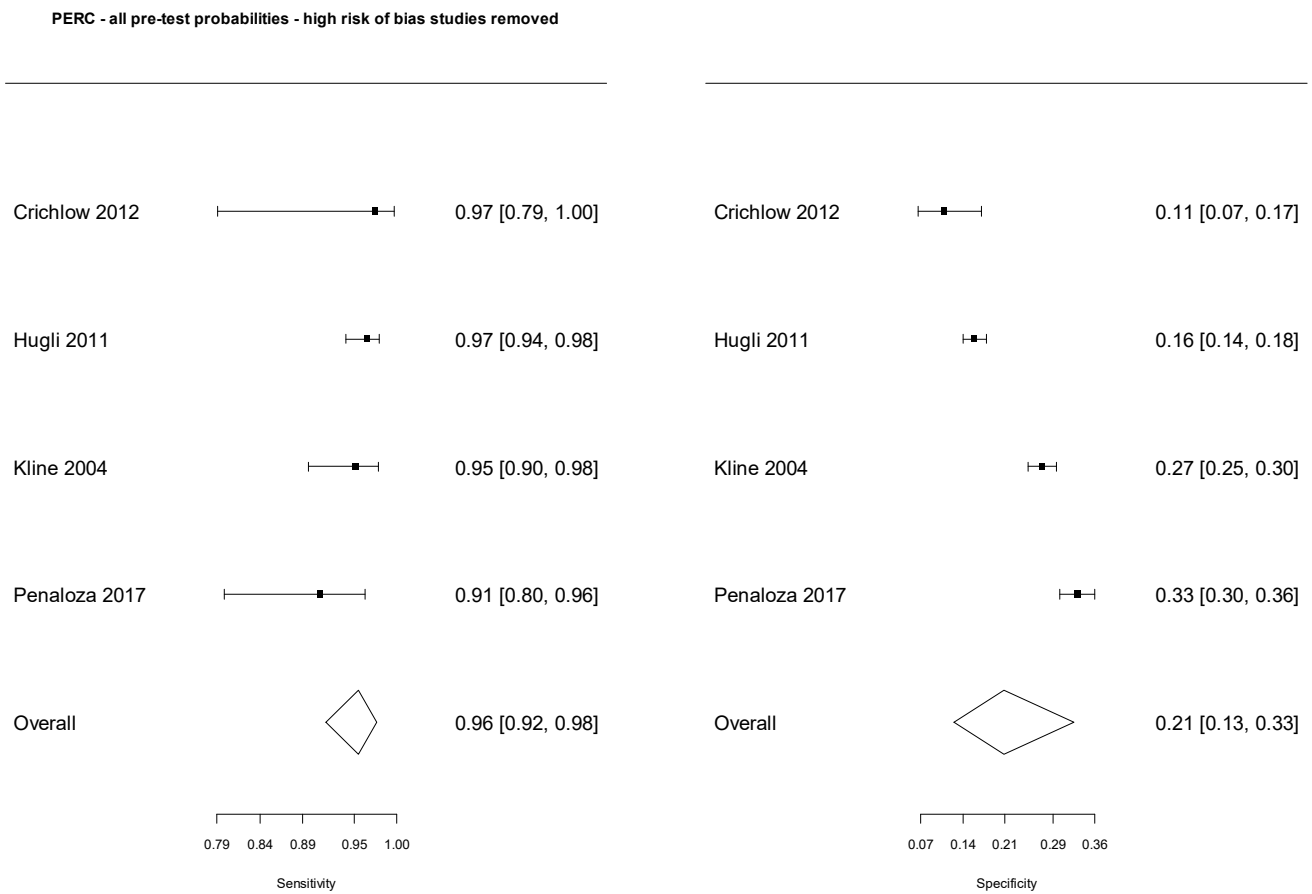
I^2 sensitivity= 0.0%, I^2 specificity=0.0%

Figure 8: Subgroup analysis: likelihood ratios for pulmonary embolism rule-out criteria (high pre-test probabilities only)



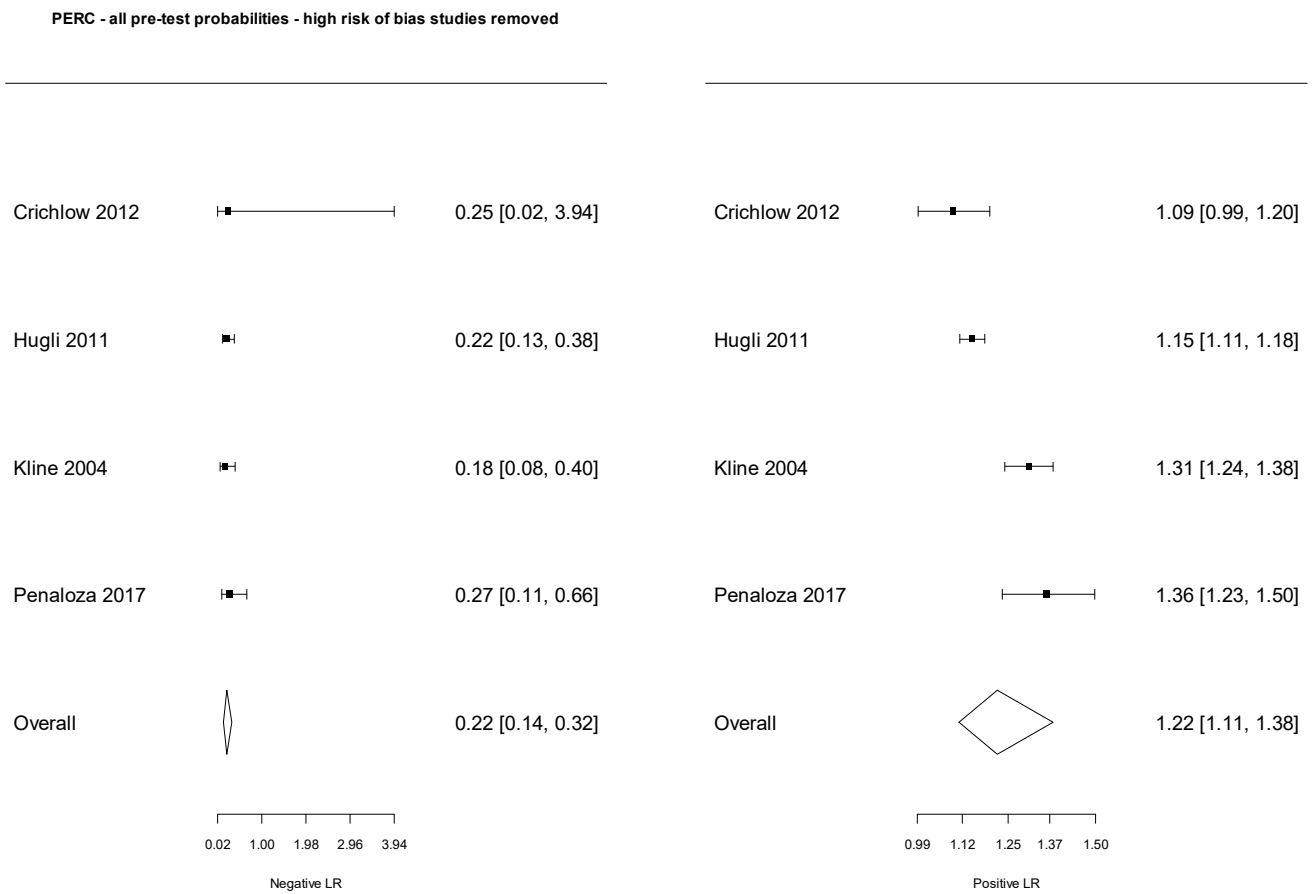
I^2 –ve LR=0.0%, I^2 +ve LR=0.0%

Figure 9: Sensitivity analysis (high risk of bias studies removed): sensitivity and specificity for pulmonary embolism rule-out criteria (all pre-test probabilities)



I^2 sensitivity=0.0% I^2 specificity=97.2%

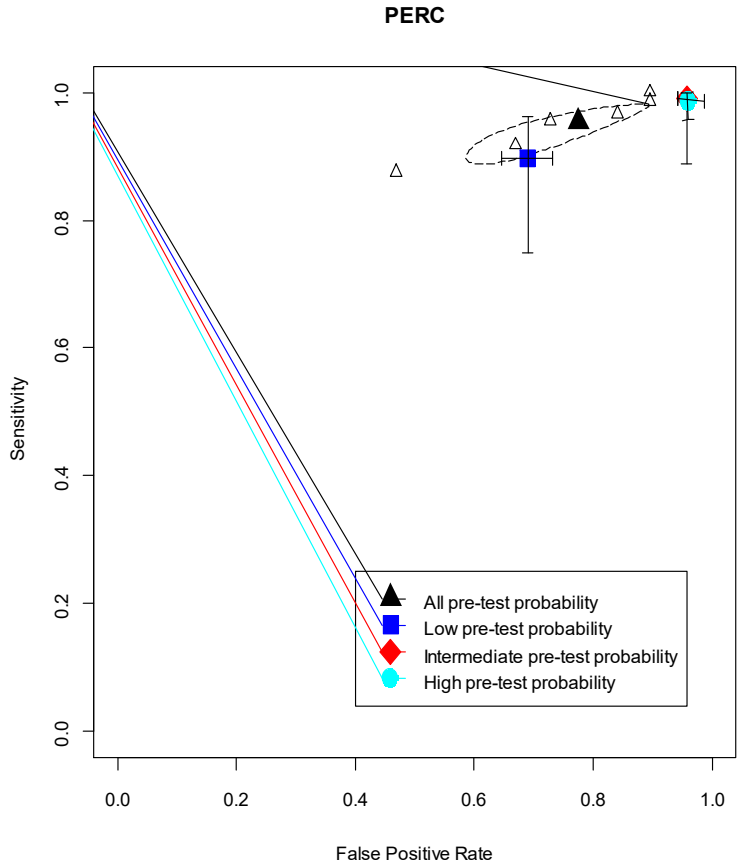
Figure 10: Sensitivity analysis (high risk of bias studies removed): likelihood ratios for pulmonary embolism rule-out criteria (all pre-test probabilities)



I^2 –ve LR=0.0%, I^2 +ve LR=90.4%

ROC plot – Main analysis and analysis by pre-test probability

Figure 11: ROC plot – main analysis and analysis by pre-test probability



Note that the main analysis (all pre-test probabilities) used a bivariate analysis, and 95% confidence intervals are shown as an ellipse. Subgroup analyses used a univariate model as insufficient data was available to allow a bivariate analysis, and confidence intervals are shown by error bars for sensitivity and false positive rate (1-specficity)

Appendix G – GRADE tables

Randomised controlled trial – PERC vs Control

Quality assessment						No of patients		Effect			Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	PERC	Control	Relative (95% CI)	Absolute: control	Absolute intervention (PERC)	
All cause mortality (3 months) (follow-up 3 months)											
1 (Freund 2018)	RCT	Serious[1]	N/A	Not serious	Serious[2]	3/962	2/954	RR 1.49 (0.25 to 8.88)	0.21 per 100	0.31 per 100 (0.05, 0.86)	Low
VTE-related mortality (3 months) (follow-up 3 months)											
1 (Freund 2018)	RCT	Serious[1]	N/A	Not serious	Not estimable[6]	0/962	0/954	Not estimable[6]	Not estimable[6]	Not estimable[6]	Moderate
Diagnostic strategy failure (thromboembolic events in participants diagnosed as not having PE) (follow-up 3 months)[5]											
1 (Freund 2018)	RCT	Serious[1]	N/A	Not serious	Very serious[3]	1/962	0/954	RR 2.98 (0.12 to 72.94)[5]	Not estimable[7]	Not estimable[7]	Very low
Length of hospital stay (hours in emergency department) (Better indicated by lower values)											
1 (Freund 2018)	RCT	Serious[1]	N/A	Not serious	Serious[4]	962	954	-	Median 0.62 lower (IQR 0.06 to 1.13 lower, p<0.001)	-	Low
Adverse event: Major bleeding (follow-up 3 months)											
1 (Freund 2018)	RCT	Serious[1]	N/A	Not serious	Not estimable[6]	0/962	0/954	Not estimable[6]	Not estimable[6]	Not estimable[6]	Moderate

1. Crossover cluster trial: unblinded - more participants judged at very low risk of PE included in the PERC periods suggesting presence of inclusion bias.
2. 95% confidence interval crosses the line of no effect.
3. 95% confidence interval crosses 2 MIDs (0.8, 1.25)
4. 95% confidence intervals not reported, though effect was statistically significant (p<0.001)
5. An intention to treat analysis was reported by the study in which all patients were included. However, this analysis was calculated by considering the worst-case scenario - that all patients lost to follow up experienced the event (e.g. mortality, VTE). This has the effect of artificially reducing the confidence intervals

by increasing the event rate, therefore we present a different ITT analysis, where participants lost to follow up were assumed not to have experienced the event.

6. Not estimable as there were 0 events in both arms.
7. Not estimable as there were 0 events in the control arm.

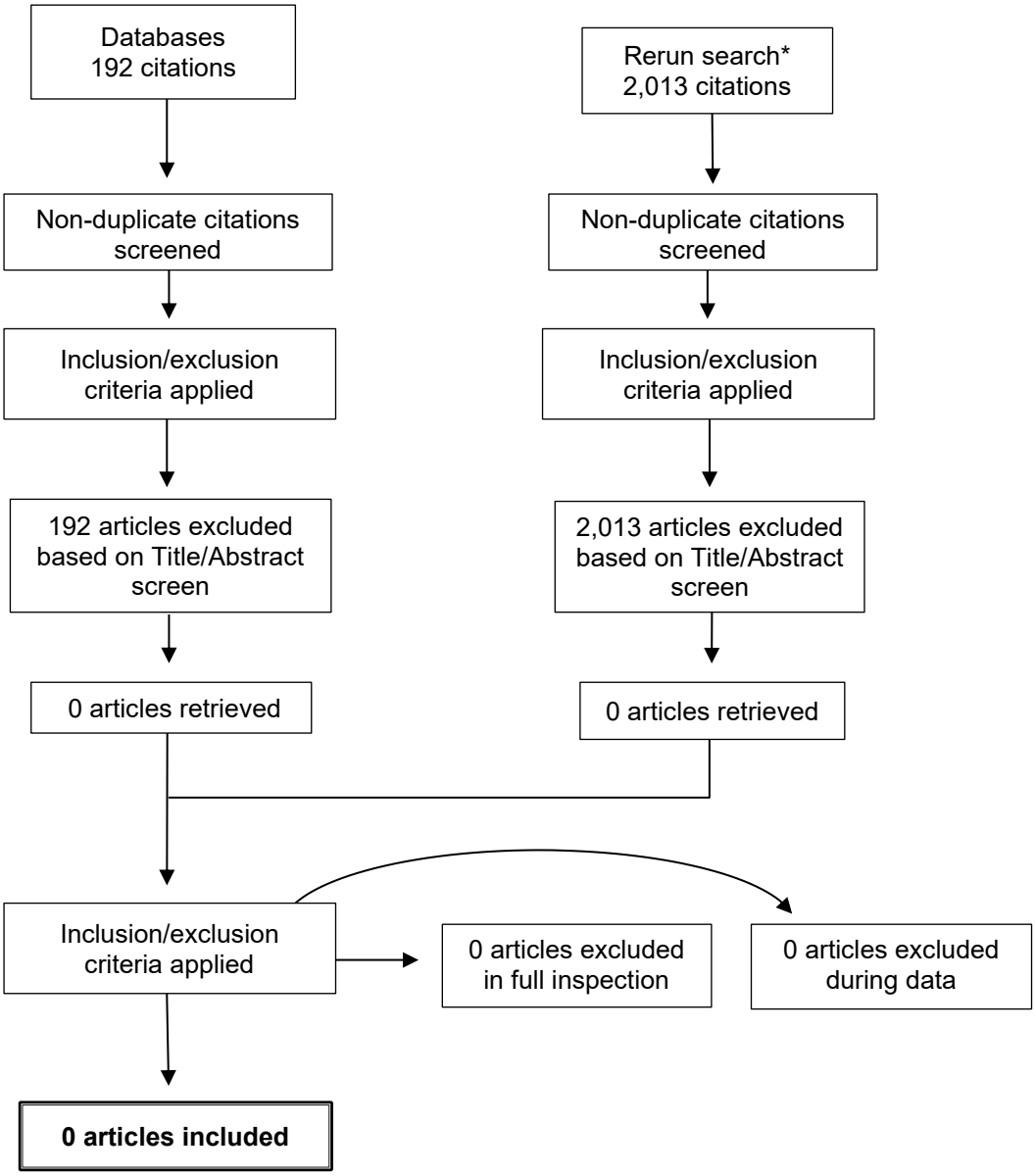
Diagnostic accuracy studies

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	Quality
Main analysis: All Pre-test probabilities (Figure 1 and Figure 2)										
6	Prospective diagnostic accuracy	5690	0.95 (0.91, 0.98)	0.23 (0.12, 0.37)	LR+ 1.24 (1.11, 1.45)	Very serious[1]	Not serious	Very serious[2]	Not serious	Very low
					LR- 0.21 (0.14, 0.30)	Very serious[1]	Not serious	Not serious	Not serious	Low
Subgroup analysis: Low pre-test probability (Figure 3 and Figure 4)										
4	Prospective diagnostic accuracy	3463	0.90 (0.75, 0.96)	0.31 (0.27, 0.35)	LR+ 1.30 (1.23, 1.38)	Very serious[1]	Not serious	Not serious	Not serious	Low
					LR- 0.33 (0.14, 0.77)	Very serious[1]	Not serious	Very serious[2]	Serious[3]	Very low
Subgroup analysis: Intermediate pre-test probability (Figure 5 and Figure 6)										
2	Prospective diagnostic accuracy	1366	0.99 (0.89, 1.00)	0.04 (0.03, 0.06)	LR+ 1.04 (1.02, 1.06)	Very serious[1]	Not serious	Not serious	Not serious	Low
					LR- 0.22 (0.02, 2.63)	Very serious[1]	Not serious	Serious[4]	Very serious[5]	Very low
Subgroup analysis: High pre-test probability (Figure 7 and Figure 8)										
2	Prospective diagnostic accuracy	284	0.99 (0.96, 1.00)	0.04 (0.01, 0.12)	LR+ 1.02 (0.98, 1.08)	Very serious[1]	Not serious	Not serious	Serious[6]	Very low
					LR- 0.36 (0.07, 1.73)	Very serious[1]	Not serious	Not serious	Very serious[5]	Very low
Sensitivity analysis (high risk of bias studies removed): (all pre-test probabilities) (Figure 9 and Figure 10)										
4	Prospective diagnostic accuracy	4,304	0.96 (0.92, 0.98)	0.21 (0.13, 0.33)	LR+ 1.22 (1.11, 1.38)	Serious[7]	Not serious	Very serious[2]	Not serious	Very low
					LR- 0.22 (0.14, 0.32)	Serious[7]	Not serious	Not serious	Not serious	Moderate
1. >33.3% of studies were at high risk of bias										

.

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	Quality
2. I^2 >66.6% 3. Confidence intervals cross 1 MID (0.5 or 1) 4. I^2 >33.3% 5.Confidence intervals cross 2 MIDs (0.5 and 1) 6. Confidence intervals cross 1 MID (1 or 2) 7. >33.3% of studies were at moderate risk of bias										

Appendix H – Economic evidence study selection



**Combined search for all questions in the guideline*

Appendix I – Economic model

Background

For this review question, the committee indicated that, alongside testing accuracy data for PERC, recommendation making would be facilitated by information on absolute numbers of patients with each testing outcome (i.e. true positives, false negatives, true negatives, and false positives), as well as estimates of costs involved in the testing process. To provide this information, we developed a simple cost-consequences analysis, comparing outcomes with and without PERC as an initial step in the diagnostic pathway in people at low risk of PE.

A full cost-utility analysis was felt to be inappropriate for this review question, as cost effectiveness is likely to be heavily dependent on the long-term health outcomes and costs associated with false negative results (patients who have a PE but are incorrectly diagnosed). Since randomised evidence of sufficient quality on the consequences of an intentionally untreated PE is unlikely to exist, such an analysis would not be feasible without substantial speculation on the downstream outcomes for these patients.

Methods

Population

People with clinically suspected PE who have a low probability of PE.

Comparators

The model compares outcomes of the following strategies:

- "No PERC" – The diagnostic pathway for PE, as specified in the 2012 update to this guideline
- "PERC" – The diagnostic pathway for PE, as specified in the 2012 update to this guideline, but with PERC as an initial strategy for ruling out further testing.

Perspective, time horizon, and discount rate

This evaluation is conducted from the perspective of the NHS/PSS. The time horizon covers the diagnostic pathway, and therefore only considers short-term costs and outcomes (<48 hours). As the time horizon is less than a year, no discounting of costs or health outcomes is applied.

Model structure

We used a decision tree structure to represent the diagnostic pathway for PE as recommended in the 2012 update to this guideline. The structure of the decision tree for the "No PERC" strategy is shown in Figure 12.

At the start of the tree, patients either have a PE or no PE (although their true status is unknown at this point). All patients undergo a Wells test; those with a "likely" score receive a chest scan (either a computed tomography pulmonary angiogram [CTPA] or a lung ventilation/perfusion [V/Q] scan). Those with an "unlikely" Wells score receive a D-dimer test.

Patients with a positive D-dimer result receive a chest scan, whereas those with a negative result receive no further testing.

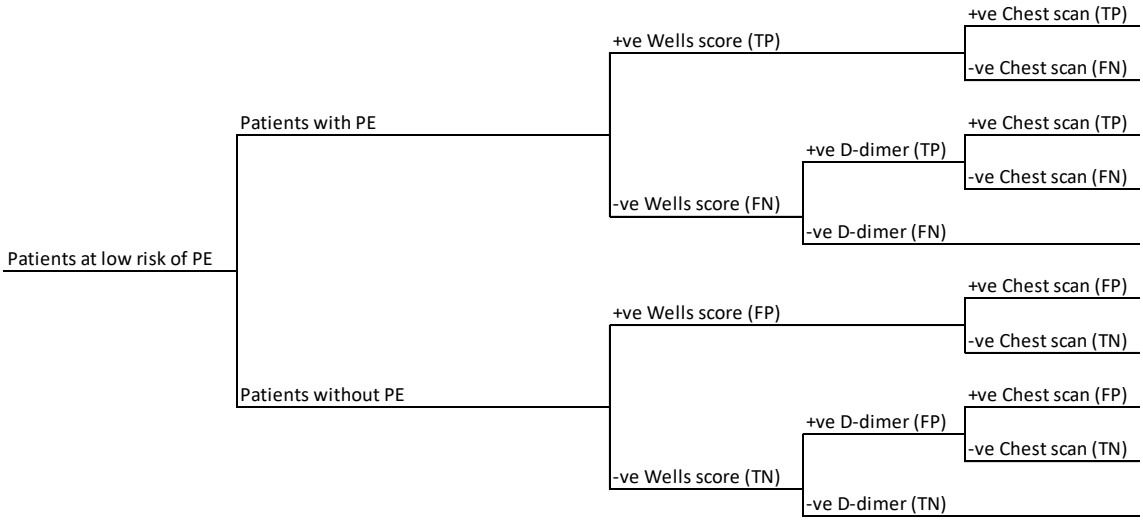


Figure 12 – Decision tree for the "No PERC" strategy

Figure 13The structure of the decision tree for the "PERC" strategy is shown in Figure 13. This structure is the same as the previous decision tree, but all patients receive PERC as an initial test. Those with a "positive" PERC result (i.e. meeting 1 or more criteria) progress through the decision tree as described above, whereas those with a "negative" are ruled out and receive no further testing.

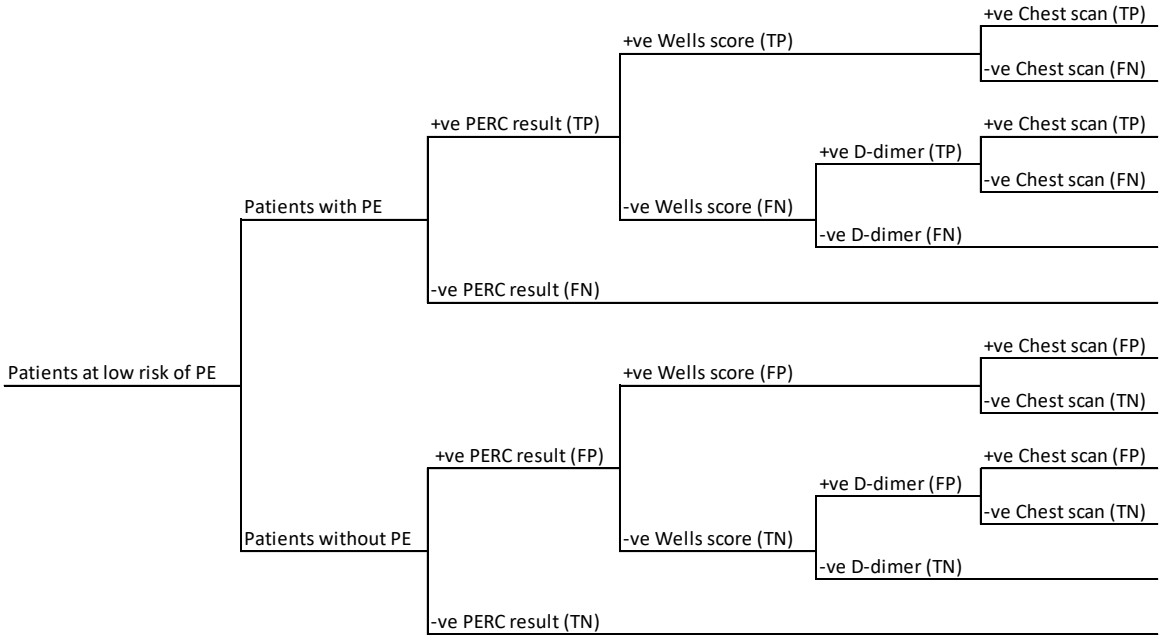


Figure 13 – Decision tree structure for the "PERC" strategy

Model inputs

Probabilities

Probability inputs used in the model (relating to the prevalence of PE and test accuracies) are shown in Table 10.

We calculated the prevalence of PE in patients deemed "low probability" from data on the underlying prevalence of PE in all patients (Goekoop et al., 2007), and data on the accuracy of "clinical gestalt" of hospital residents (postgraduate year 2-3) in determining whether PE was the most likely diagnosis or another diagnosis was more likely (Kabrhel et al., 2005). We achieved this by calculating the proportion of patients who are false negatives and true negatives after using clinical gestalt and, from this, the proportion of all negative results which are false negatives. This provided a prevalence of 7.3% in patients deemed at low-risk of PE.

We obtained data on the sensitivity and specificity of PERC directly from the results of the meta-analysis conducted for the clinical review. We used alternative accuracy data pertaining only to studies in people with a low pre-test probability of PE as a sensitivity analysis. Data on the accuracy of the Wells score, CTPA and V/Q scan were sourced from the literature: Posadas-Martínez et al. (2014), Hogg et al. (2006), and Sostman et al. (2008) respectively. The committee indicated that around 20% of patients requiring a chest scan would receive a V/Q scan, and the remainder a CTPA, so we calculated a weighted average sensitivity and specificity of these two tests to inform the diagnostic accuracy of a chest scan. Data on the accuracy of D-dimer were taken from the results of the meta-analyses conducted for point-of-care and laboratory D-dimer tests in evidence review A. We made the assumption that 50% of patients are tested using a point-of-care D-dimer test, and the remaining 50% with a laboratory test. Using these values, we calculated a weighted average sensitivity and specificity for D-dimer tests.

Table 10 – Probability input parameters

Parameter	Point estimate (95% CIs)	Distribution in PSA*	Source
Prevalence of PE in all presenting patients	12.3% (10.2% to 14.5%)	Beta	Goekoop et al. (2007)
Accuracy of clinical gestalt (resident physician)			
Sensitivity	58% (53% to 65%)	Beta	Kabrhel et al. (2005)
Specificity	75% (70% to 81%)	Beta	Kabrhel et al. (2005)
Prevalence of PE in patients at low risk of PE	7.3%	-	Calculated
Accuracy of PERC			
Sensitivity - all studies	95% (91% to 98%)	Beta	Clinical review
Specificity - all studies	23% (12% to 37%)	Beta	Clinical review
Sensitivity - low pre-test prob studies	90% (75% to 96%)	Beta	Clinical review
Specificity - low pre-test prob studies	31% (27% to 35%)	Beta	Clinical review
Accuracy of Wells score			

Parameter	Point estimate (95% CIs)	Distribution in PSA*	Source
Sensitivity	65% (59% to 72%)	Beta	Posadas-Martínez et al. (2014)
Specificity	81% (77% to 85%)	Beta	Posadas-Martínez et al. (2014)
Accuracy of D-dimer			
Sensitivity - point-of-care test	89% (73% to 96%)	Beta	Clinical review for D-dimer review questions
Specificity - point-of-care test	60% (50% to 69%)	Beta	Clinical review for D-dimer review questions
Sensitivity - lab test	92% (88% to 94%)	Beta	Clinical review for D-dimer review questions
Specificity - lab test	44% (32% to 58%)	Beta	Clinical review for D-dimer review questions
Sensitivity - overall	91%	-	Calculated
Specificity - overall	52%	-	Calculated
Accuracy of chest scans			
Sensitivity - CTPA	89% (83% to 95%)	Beta	Hogg et al. (2006)
Specificity - CTPA	95% (91% to 98%)	Beta	Hogg et al. (2006)
Sensitivity – V/Q scan	77% (70% to 85%)	Beta	Sostman et al. (2008)
Specificity – V/Q scan	98% (96% to 99%)	Beta	Sostman et al. (2008)
Sensitivity - overall	87%	-	Calculated
Specificity - overall	96%	-	Calculated
Testing assumptions			
Proportion of scans which are V/Q rather than CTPA	20%	-	Assumption
Proportion of D-dimer tests which are point-of-care	50%	-	Assumption
Strength of dependence between tests			
Dependence between PERC and Wells score (sensitivity and specificity)	70%	-	Assumption
Dependence between Wells score and D-dimer (sensitivity and specificity)	10%	-	Assumption
Dependence between Wells score and chest scan (sensitivity and specificity)	10%	-	Assumption
Dependence between D-dimer and chest scan (sensitivity and specificity)	10%	-	Assumption

*PSA = probabilistic sensitivity analysis

Since the model simulates a number of sequential tests, we deemed it appropriate to assume that there is some level of conditional dependence between test outcomes. That is

to say, if a test produces a false negative result (for example) for a particular patient, it is reasonable to expect that a second test would also be more likely to produce a false negative result if the two tests measure similar outcomes. This is particularly pertinent for patients who are tested with PERC followed by a Wells score, since many of the criteria for the two tests are the same. Assuming independence between these two tests is likely to favour the "PERC" strategy, since there will be fewer patients with false positive results from both PERC and Wells.

To implement conditional dependence of tests in the model, we used the method described in Gardener et al. (2000). In summary, for each pair of tests, we first calculated a maximum covariance for both sensitivity (γ_{se}) and specificity (γ_{sp}) using the following formulae:

$$\gamma_{se} = MIN(Se_1\ (1 - Se_2\ );\ Se_2\ (1 - Se_1\ ))$$

$$\gamma_{sp} = MIN(Sp_1\ (1 - Sp_2\ );\ Sp_2\ (1 - Sp_1\ ))$$

Where Se_1 = sensitivity of test 1; Se_2 = sensitivity of test 2; Sp_1 = specificity of test 1; Sp_2 = specificity of test 2; and MIN is a function which selects the minimum value between those listed.

Next, we assigned a value between 0 and 1 to each pair of tests to denote the strength of the relationship, based on how similar the pairs of tests were, where 0 = complete independence and 1 = maximum possible co-dependence. We assigned a value of 0.7 to the combination of PERC and Wells, since these tests are very similar, and a value of 0.1 to all other combinations (Wells and D-dimer, Wells and chest scan, D-dimer and chest scan), as these tests measure fundamentally different outcomes. These values were then multiplied by the corresponding maximum covariance for each pair of tests, to calculate the actual conditional covariances used in the model.

Using these covariances and sensitivities and specificities of each test, joint probabilities of obtaining each possible combination of results for patients with and without PE were calculated, using the formulae shown in Table 11.

Table 11 – Formulae for calculating joint testing outcomes

Outcome	Probability
Patients who have PE	
T_1(+ve) AND T_2(-ve)	Se(T_1) x (1 - Se(T_2)) – γ_{se}
T_1(+ve) AND T_2(+ve)	Se(T_1) x Se(T_2) + γ_{se}
T_1(-ve) AND T_2(+ve)	(1 - Se(T_1)) x Se(T_2) – γ_{se}
T_1(-ve) AND T_2(-ve)	(1 - Se(T_1)) x (1 - Se(T_2)) + γ_{se}
Patients who do have PE	
T_1(+ve) AND T_2(-ve)	(1 – Sp(T_1)) x Sp(T_2) - γ_{sp}
T_1(+ve) AND T_2(+ve)	(1 – Sp(T_1)) x (1 - Sp(T_2)) + γ_{sp}
T_1(-ve) AND T_2(+ve)	Sp(T_1) x (1 - Sp(T_2)) - γ_{sp}
T_1(-ve) AND T_2(-ve)	Sp(T_1) x Sp(T_2) + γ_{sp}

Abbreviations: Se = sensitivity; Sp = specificity; T_1 = test 1; T_2 = test 2; γ_{se} = sensitivity covariance; γ_{sp}= specificity covariance

Costs

All costs used in the model are shown in Table 12. We calculated the cost of a point-of-care D-dimer test using a simple mean of all tests listed in the NHS Supply Chain Catalogue. Costs of laboratory D-dimer tests could not be identified in the literature or from standard NHS costing sources, since these values tend to vary regionally depending on the local laboratory service used. Therefore, we obtained costs from the committee, and a mean of these values was taken.

Costs of CTPA and V/Q scan were taken from NHS Reference Costs 2017/18.

We did not model costs for PERC and Wells score, since these tests were assumed to be carried out within the initial consultation, and do not require any additional capital expenditure.

Table 12 – Cost input parameters

Parameter	Point estimate (95% CIs)	Distribution in PSA*	Source
D-dimer testing costs			
Alere Triage (5 pack) - quantitative	£29.22	-	NHS Supply Chain Catalogue
Alere Triage (25 pack) - quantitative	£12.63	-	NHS Supply Chain Catalogue
Roche Cobas (2 pack) - quantitative	£27.37	-	NHS Supply Chain Catalogue
Roche Cobas (10 pack) - quantitative	£9.44	-	NHS Supply Chain Catalogue
Ciga Suresign (10 pack) - qualitative	£8.81	-	NHS Supply Chain Catalogue
Siemens dil pak (5 pack) - qualitative	£6.48	-	NHS Supply Chain Catalogue
Chirus StatusFirst (20 pack) - qualitative	£10.04	-	NHS Supply Chain Catalogue
Mean point-of-care test cost	£14.86 (£7.91 to £21.80)	Gamma	Calculated
Cost of laboratory test	£6.79 (£2.44 to £11.13)	Gamma	Calculated
Chest scan costs			
CTPA	£106.12	-	NHS Reference Costs 2017/18 - Computerised Tomography Scan of One Area, with Post-Contrast Only, 19 years and over
V/Q scan	£311.07	-	NHS Reference Costs 2017/18 - Lung Ventilation or Perfusion Scan, 19 years and over
Mean cost of chest scan	£147.11	-	Calculated

*PSA = probabilistic sensitivity analysis

Uncertainty

Uncertainty in model results was explored via probabilistic sensitivity analysis. Model input parameters were assigned probability distributions reflecting uncertainty surrounding point estimates, defined by standard error/confidence intervals and type of parameter. A random value was drawn from each of these distributions for 1,000 iterations and, for each iteration, model results were recorded for each strategy. This process allowed uncertainty in results to be expressed as 95% credible intervals.

The particular distribution assigned to each type of parameter was chosen to reflect the nature of the data. Probabilities were parameterised using a beta distribution, as these values must lie between 0 and 1. Unit costs were given a gamma distribution, since these values are bound at 0, but theoretically have no upper limit.

We conducted one main deterministic sensitivity analysis using accuracy data for PERC from studies in people with a low pre-test probability of PE, which was reported as a subgroup in the clinical review. In addition, we also conducted four exploratory scenario analysis, in order to test key model assumptions. These were:

1. **Lower prevalence of PE** – a value of 2.1% (taken from Freund 2018 from the clinical review) was used to represent the prevalence of PE (as opposed to the value of 7.3% in the base case). In addition, accuracy data for PERC from the low pre-test probability subgroup analysis (as opposed to the overall population) were used.
2. **High accuracy CTPA** – a sensitivity of 95% and a specificity of 97% was used to inform the accuracy of CTPA, to assess outcomes in a scenario where chest scans are more accurate than in the base case. These data were taken from Qanadli et al. (2000) – one of the studies included in the Hogg et al. (2006) meta-analysis. This study was selected as it has the largest sample size of the studies included in the meta-analysis, as well as providing high estimates of CTPA accuracy.
3. **Test outcomes are completely independent** – strength of dependence for each pair of tests set to 0%.
4. **Test outcomes are maximally co-dependent** – strength of dependence for each pair of tests set to 100%.

Results

Base case analysis

Testing outcomes for PERC alone (i.e. not as part of the full diagnostic pathway) per 1,000 patients are shown in Table 13. These results show that PERC produces 217 negative results; PE is ruled out and no further testing is conducted in these patients. Of these results, 213 are true negatives (patients without PE), and 4 are false negatives (patients with PE).

Table 13 – Testing outcomes of PERC alone per 1,000 patients (base case)

Testing outcome	Number of patients (95% CrIs)
True positive	69 (54 to 87)
False negative	4 (1 to 7)
True negative	213 (111 to 342)
False positive	714 (584 to 817)

Testing outcomes for the whole diagnostic pathway, comparing the "PERC" strategy to the "No PERC" strategy are shown in Table 14. These results show that the "PERC" strategy results in 3 more false negative results, but 5 fewer false positive results.

Table 14 – Testing outcomes for the entire diagnostic pathway for "PERC" and "No PERC" strategies per 1,000 patients (base case)

Testing outcome	PERC	No PERC	Difference (95% CrIs)
True positive	59	61	-3 (-5 to -1)
False negative	14	11	3 (1 to 5)
True negative	904	899	5 (2 to 11)
False positive	23	29	-5 (-11 to -2)

The number of downstream tests (D-dimer and chest scan) per 1,000 patients for the "PERC" and "No PERC" strategies are shown in Table 15. The corresponding costs of these tests are shown in Table 16. These results show that the "PERC" strategy leads to substantially fewer D-dimer tests and chest scans, which produces a cost saving of £16,414 per 1,000 patients.

Table 15 – Numbers of D-dimer tests and chest scans for "PERC" and "No PERC" strategies per 1,000 patients (base case)

Test	PERC	No PERC	Difference (95% CrIs)
D-dimer	572	777	-204 (-326 to -108)
Chest scan	488	597	-110 (-177 to -57)

Table 16 – Diagnostic pathway costs for "PERC" and "No PERC" strategies per 1,000 patients (base case)

Cost category	PERC	No PERC	Difference (95% CrIs)
D-dimer	£6,194	£8,402	-£2,208 (-£3,886 to -£1,007)
Chest scan	£71,775	£87,890	-£16,114 (-£26,018 to -£8,240)
Total	**£77,970**	**£96,292**	**-£18,322 (-£29,486 to -£9,416)**

Sensitivity analysis – accuracy of PERC using low pre-test probability subgroup

Results of the sensitivity analysis using accuracy data for PERC from low pre-test probability studies only are shown in Table 17 to Table 20.These results show that, compared to the base case analysis, the "PERC" strategy produces a greater number of false negative results, but also a greater reduction in false positive results, fewer downstream tests, and greater cost savings. This is a result of the lower sensitivity and higher specificity of PERC in the low pre-test probability subgroup analysis.

Table 17 – Testing outcomes of PERC alone per 1,000 patients (sensitivity analysis using accuracy data from the low pre-test probability subgroup analysis)

Testing outcome	Number of patients (95% CrIs)
True positive	66 (49 to 84)
False negative	7 (2 to 17)
True negative	287 (251 to 325)
False positive	640 (601 to 678)

Table 18 – Testing outcomes for the entire diagnostic pathway for "PERC" and "No PERC" strategies per 1,000 patients (sensitivity analysis using accuracy data from the low pre-test probability subgroup analysis)

Testing outcome	PERC	No PERC	Difference (95% CrIs)
True positive	56	61	-6 (-14 to -1)
False negative	17	11	6 (1 to 14)
True negative	906	899	7 (3 to 13)
False positive	21	29	-7 (-13 to -3)

Table 19 – Numbers of D-dimer tests and chest scans for "PERC" and "No PERC" strategies per 1,000 patients (sensitivity analysis using accuracy data from the low pre-test probability subgroup analysis)

Test	PERC	No PERC	Difference (95% CrIs)
D-dimer	500	777	-277 (-313 to -242)
Chest scan	448	597	-150 (-180 to -121)

Table 20 – Diagnostic pathway costs for "PERC" and "No PERC" strategies per 1,000 patients (sensitivity analysis using accuracy data from the low pre-test probability subgroup analysis)

Cost category	PERC	No PERC	Difference (95% CrIs)
D-dimer	£5,406	£8,402	-£2,996 (-£4,321 to -£1,903)
Chest scan	£65,853	£87,890	-£22,037 (-£26,584 to -£17,843)
Total	**£71,258**	**£96,292**	**-£25,033 (-£30,002 to -£20,426)**

Additional scenario analyses

Key results for the 4 additional scenario analyses are shown in Table 21. Results for scenario assuming a lower prevalence of PE are more favourable towards the "PERC" strategy than in the base case; PERC produces a cost saving of £25,636 and 8 fewer false positive results per 1,000 patients, at the expense of only 2 more false negative results, compared to the "No PERC" strategy.

The scenario in which optimistic accuracy data are used for CTPA shows a reduction in the incremental false positives produced by PERC, but results are otherwise similar to the base case.

The two scenarios exploring the conditional dependence of tests show that these assumptions are unlikely to materially affect test outcomes. Assuming maximum possible co-dependence of each pair of tests reduces the cost saving produced by PERC somewhat, although it should be noted that this is an extreme scenario.

Table 21 – Key results of scenario analyses per 1,000 patients

Scenario	Incremental false negatives for whole pathway ("PERC" minus "No PERC")	Incremental false positives for whole pathway ("PERC" minus "No PERC")	Incremental cost of whole pathway ("PERC" minus "No PERC")
Lower prevalence of PE	2	-8	-£25,636
High accuracy CTPA	3	-3	-£18,322

Scenario	Incremental false negatives for whole pathway ("PERC" minus "No PERC")	Incremental false positives for whole pathway ("PERC" minus "No PERC")	Incremental cost of whole pathway ("PERC" minus "No PERC")
Test outcomes are completely independent	3	-6	-£20,558
Test outcomes are maximally co-dependent	3	-7	-£13,969

Discussion

The decision of whether or not to use PERC as a ruling out strategy at the start of the diagnostic pathway for PE is a trade-off between a higher number of false negative test results, balanced against lower costs and fewer false positive results. If the detrimental effects of false negative and false positive results were weighted equally, using PERC as an initial rule-out test would be the clearly superior strategy; it produces a cost saving of £18,322 and 5 fewer false positive results, at the expense of only 3 more false negative test results in the base case. However, consideration should be given to the relative severity of false negative and false positive results. False negative results cause a delay in the treatment of people with a PE, which may lead to serious detrimental health effects and substantial downstream costs. Contrastingly, false positive results at the end of the diagnostic pathway lead to people without a PE receiving unnecessary anticoagulation, which produces additional costs and, in some cases, serious side-effects such as bleeding events. Furthermore, patients who are incorrectly diagnosed as having a PE may have another underlying condition, for which they do not receive appropriate treatment.

A full cost-utility analysis would attempt to quantify all downstream cost and QALY outcomes for each testing outcome, and would therefore explicitly weigh up the trade-offs involved in including PERC in the clinical pathway. However, as previously discussed, conducting such an analysis would be impractical, as high-quality evidence on patients with a PE who are intentionally untreated (i.e. the outcome of a false negative result) is unlikely to exist. Therefore, the weighting of the trade-off between false negatives, false positives and costs must fall to the experience of the committee.

Results of the subgroup analysis using diagnostic accuracy data for PERC from studies with a low pre-test probability of PE are more pronounced than those of the base case analysis. The “PERC” strategy produces 7 fewer false positive results, 6 more false negative results and a slightly larger cost saving of £25,033. This is due to the lower sensitivity and higher specificity of PERC compared to the main analysis. It is possible that these results are more reflective of real practice than those using accuracy data from all populations, since, in reality PERC would only be used for patients deemed at low risk of PE. However, it should also be noted that these accuracy data are based on fewer studies, and therefore may be less be precise than the outcomes of the main analysis.

An additional scenario analysis assuming a lower prevalence of PE showed that, compared to the base case analysis, the “PERC” strategy results in a larger reduction in false positive results (8 per 1,000 people) and a smaller increase in the number of false negative results (2 per 1,000 people). This is because the lower prevalence of PE means that there are fewer false negative results overall, and therefore PERC produces a smaller absolute increase in false negative results. For false positive results, the converse is true; “PERC” produces a larger absolute reduction in the number of incorrect PE diagnoses compared to “No PERC”.

In addition, when combined with the higher specificity of the accuracy data for PERC from the low pre-test probability subgroups analysis, this scenario produces slightly higher cost savings compared to the base case.

Other scenario analyses show that model assumptions are unlikely to affect conclusions, although it was noted that using an optimistic estimate of CTPA accuracy somewhat diminishes the reduction in false positives achieved by PERC. The two analyses exploring conditional dependence of tests show that these assumptions are unlikely to materially affect conclusions.

References

Gardner, I.A., Stryhn, H., Lind, P. and Collins, M.T., 2000. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. Preventive veterinary medicine, 45(1-2), pp.107-122.

Goekoop, R.J., Steeghs, N., Niessen, R.W., Jonkers, G.J., Dik, H., Castel, A., Werker-van Gelder, L., Vlasveld, T.L., van Klink, R.C., Planken, E.V. and Huisman, M.V., 2007. Simple and safe exclusion of pulmonary embolism in outpatients using quantitative D-dimer and Wells' simplified decision rule. Thrombosis and haemostasis, 97(01), pp.146-150.

Hogg, K., Brown, G., Dunning, J., Wright, J., Carley, S., Foex, B. and Mackway-Jones, K., 2006. Diagnosis of pulmonary embolism with CT pulmonary angiography: a systematic review. Emergency medicine journal, 23(3), pp.172-178.

Kabrhel, C., Camargo Jr, C.A. and Goldhaber, S.Z., 2005. Clinical gestalt and the diagnosis of pulmonary embolism: does experience matter?. Chest, 127(5), pp.1627-1630.

National Schedule of Reference Costs. NHS Improvement. Available from: https://improvement.nhs.uk/resources/reference-costs/

NHS Supply Chain Catalogue. Available from: https://my.supplychain.nhs.uk/catalogue

Posadas-Martínez, M.L., Vázquez, F.J., Giunta, D.H., Waisman, G.D., de Quirós, F.G.B. and Gándara, E., 2014. Performance of the Wells score in patients with suspected pulmonary embolism during hospitalization: a delayed-type cross sectional study in a community hospital. Thrombosis research, 133(2), pp.177-181.

Qanadli, S.D., Hajjam, M.E., Mesurolle, B., Barré, O., Bruckert, F., Joseph, T., Mignon, F., Vieillard-Baron, A., Dubourg, O. and Lacombe, P., 2000. Pulmonary embolism detection: prospective evaluation of dual-section helical CT versus selective pulmonary arteriography in 157 patients. Radiology, 217(2), pp.447-455.

Sostman, H.D., Stein, P.D., Gottschalk, A., Matta, F., Hull, R. and Goodman, L., 2008. Acute pulmonary embolism: sensitivity and specificity of ventilation-perfusion scintigraphy in PIOPED II study. Radiology, 246(3), pp.941-946.

Appendix J – Excluded studies

Clinical studies (main search)

Author (year)	Title	Reason(s) for exclusion
Beam (2007)	Application of the pulmonary embolism rule-out criteria in a rural population	• Conference abstract
Ceriani (2010)	Clinical prediction rules for pulmonary embolism: a systematic review and meta-analysis	• Systematic review used as source of primary studies
Courtney (2006)	Prospective evaluation of the Pulmonary Embolism Rule-out Criteria (PERC) rule: an 8-variable block rule to identify subjects	• Conference abstract
Freund (2017)	PERC rule to exclude the diagnosis of pulmonary embolism in low-risk emergency patients: a noninferiority randomized controlled trial	• Conference abstract
Kline (2008)	Prospective multicenter evaluation of the pulmonary embolism rule-out criteria	• Reference standard in study does not match that specified in protocol *Reference standard was 45 days follow up*
Kline (2010)	Prospective evaluation of real-time use of the pulmonary embolism rule-out criteria in an academic emergency department	• Reference standard in study does not match that specified in protocol *Reference standard was 14 day follow up* • Not a relevant study design *Retrospective study.*
Kline (2018)	Utility of a Clinical Prediction Rule to Exclude Pulmonary Embolism Among Low-Risk Emergency Department Patients: Reason to PERC Up	• Review article but not a systematic review
Lucassen (2012)	Review: Gestalt or clinical decision rules have limited sensitivity and specificity for detecting acute PE	• Study does not contain any relevant index tests
Rehnberg (2014)	BET 3: Pulmonary embolism rule-out criteria (PERC) for excluding pulmonary embolism	• Review article but not a systematic review
Righini (2005)	More on: clinical criteria to prevent unnecessary diagnostic testing in emergency department patients with suspected pulmonary embolism	• Not a relevant study design *Retrospective study*
Self (2012)	Is "PERC negative" adequate to rule out pulmonary embolism in the emergency department? Evaluating meta-analysis for studies of clinical prediction models	• Review article but not a systematic review *Commentary on systematic review*

Siau (2015)	Use of pulmonary embolism rule-out criteria (PERC) in the emergency department	• Not a relevant study design *Retrospective study and literature review* • Full text paper not available
Singh (2012)	Diagnostic accuracy of pulmonary embolism rule-out criteria: a systematic review and meta-analysis	• Systematic review used as source of primary studies
Singh (2013)	Pulmonary embolism rule-out criteria (PERC) in pulmonary embolism--revisited: a systematic review and meta-analysis	• Systematic review used as source of primary studies
van der Pol (2018)	Combination of Pulmonary Embolism Rule-out Criteria and YEARS Algorithm in a European Cohort of Patients with Suspected Pulmonary Embolism	• Not a relevant study design *Retrospective study*
Wolf (2008)	Assessment of the pulmonary embolism rule-out criteria rule for evaluation of suspected pulmonary embolism in the emergency department	• Not a relevant study design *Retrospective study*

Clinical studies (search update)

Author (year)	Title	Reason(s) for exclusion
Buntine (2019)	Effect of a clinical flowchart incorporating Wells score, PERC rule and age-adjusted D-dimer on pulmonary embolism diagnosis, scan rates and diagnostic yield.	• 2x2 table cannot be calculated
Crane (2018)	Retrospective validation of the pulmonary embolism rule-out criteria rule in 'PE unlikely' patients with suspected pulmonary embolism.	• Not a relevant study design *Retrospective cohort study*
Gorlicki (2019)	Safety of the Combination of PERC and YEARS Rules in Patients With Low Clinical Probability of Pulmonary Embolism: A Retrospective Analysis of Two Large European Cohorts.	• Not a relevant study design *Retrospective cohort study*
Malavolta (2019)	Effect of the Pulmonary Embolism Rule-Out Criteria on subsequent thromboembolic events among low-risk emergency department patients: the PROPER randomized clinical trial.	• Associate paper of included study (no new data)
Penaloza (2017)	Pulmonary embolism rule-out criteria (PERC) rule in European patients with low implicit clinical probability (PERCEPIC): a multicentre, prospective, observational study.	• Duplicate reference

Appendix K – References

Included clinical studies

Crichlow A, Cuker A, and Mills A M (2012) Overuse of computed tomography pulmonary angiography in the evaluation of patients with suspected pulmonary embolism in the emergency department. Academic Emergency Medicine 19(11), 1219-26

Freund Y, Cachanado M, Aubry A, Orsini C, Raynal P A, Feral-Pierssens A L, Charpentier S, Dumas F, Baarir N, Truchot J, Desmettre T, Tazarourte K, Beaune S, Leleu A, Khellaf M, Wargon M, Bloom B, Rousseau A, Simon T, Riou B, and Group Proper Investigator (2018) Effect of the Pulmonary Embolism Rule-Out Criteria on Subsequent Thromboembolic Events Among Low-Risk Emergency Department Patients: The PROPER Randomized Clinical Trial. JAMA 319(6), 559-566

Hogg K, Dawson D, and Kline J (2005) Application of pulmonary embolism rule-out criteria to the UK Manchester Investigation of Pulmonary Embolism Diagnosis (MIOPED) study cohort. Journal of Thrombosis & Haemostasis 3(3), 592-3

Hugli O, Righini M, Le Gal , G , Roy P M, Sanchez O, Verschuren F, Meyer G, Bounameaux H, and Aujesky D (2011) The pulmonary embolism rule-out criteria (PERC) rule does not safely exclude pulmonary embolism.[Erratum appears in J Thromb Haemost. 2012 Apr;10(4):740]. Journal of Thrombosis & Haemostasis 9(2), 300-4

Kline J A, Mitchell A M, Kabrhel C, Richman P B, and Courtney D M (2004) Clinical criteria to prevent unnecessary diagnostic testing in emergency department patients with suspected pulmonary embolism. Journal of Thrombosis & Haemostasis 2(8), 1247-55

Penaloza A, Verschuren F, Dambrine S, Zech F, Thys F, and Roy P M (2012) Performance of the Pulmonary Embolism Rule-out Criteria (the PERC rule) combined with low clinical probability in high prevalence population. Thrombosis Research 129(5), e189-93

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

Excluded clinical studies (main search)

Beam D, Brewer K, and Kline JA (2007) Application of the pulmonary embolism rule-out criteria in a rural population. Ann Emerg Med 50, 132

Ceriani E, Combescure C, Le Gal , G , Nendaz M, Perneger T, Bounameaux H, Perrier A, and Righini M (2010) Clinical prediction rules for pulmonary embolism: a systematic review and meta-analysis. Journal of Thrombosis & Haemostasis 8(5), 957-70

Courtney D M, Pribaz J R, Senh A C, Shumer D E, Pierce L C, and Gorman S M (2006) Prospective evaluation of the Pulmonary Embolism Rule-out Criteria (PERC) rule: an 8-

variable block rule to identify subjects. Academic Emergency Medicine 13(5 Supplement 1), S157

Freund Y, Bokobza J, Cachanado M, Aubry A, Aufredou F-P, Raynal P-A, Simon T, and Riou B (2017) PERC rule to exclude the diagnosis of pulmonary embolism in low-risk emergency patients: a noninferiority randomized controlled trial. Annals of emergency medicine. Conference: American college of emergency physicians, and ACEP 2017 research forum. United States 70(4 Supplement 1), S30

Kline J A, Courtney D M, Kabrhel C, Moore C L, Smithline H A, Plewa M C, Richman P B, O'Neil B J, and Nordenholz K (2008) Prospective multicenter evaluation of the pulmonary embolism rule-out criteria. Journal of Thrombosis & Haemostasis 6(5), 772-80

Kline J A, Peterson C E, and Steuerwald M T (2010) Prospective evaluation of real-time use of the pulmonary embolism rule-out criteria in an academic emergency department. Academic Emergency Medicine 17(9), 1016-9

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

Lucassen W (2012) Review: Gestalt or clinical decision rules have limited sensitivity and specificity for detecting acute PE. Annals of Internal Medicine 156(2), 20120117

Rehnberg J V (2014) BET 3: Pulmonary embolism rule-out criteria (PERC) for excluding pulmonary embolism. Emergency Medicine Journal 31(1), 81-82

Righini M, Le Gal , G , Perrier A, and Bounameaux H (2005) More on: clinical criteria to prevent unnecessary diagnostic testing in emergency department patients with suspected pulmonary embolism. Journal of Thrombosis & Haemostasis 3(1), 188-9; author reply 190-1

Self W H, and Barrett T W (2012) Is "PERC negative" adequate to rule out pulmonary embolism in the emergency department? Evaluating meta-analysis for studies of clinical prediction models. Annals of Emergency Medicine 60(1), 129-131

Siau R, Young S, Blyth M, and Dickson-Jardine K (2015) Use of pulmonary embolism rule-out criteria (PERC) in the emergency department. Clinical Medicine 15 Suppl 3, s1

Singh B, Parsaik A K, Agarwal D, Surana A, Mascarenhas S S, and Chandra S (2012) Diagnostic accuracy of pulmonary embolism rule-out criteria: a systematic review and meta-analysis. Annals of Emergency Medicine 59(6), 517-20.e1-4

Singh B, Mommer S K, Erwin P J, Mascarenhas S S, and Parsaik A K (2013) Pulmonary embolism rule-out criteria (PERC) in pulmonary embolism--revisited: a systematic review and meta-analysis. Emergency Medicine Journal 30(9), 701-6

van der Pol , L M, van der Hulle , T , Mairuhu A T. A, Huisman M V, and Klok F A (2018) Combination of Pulmonary Embolism Rule-out Criteria and YEARS Algorithm in a European Cohort of Patients with Suspected Pulmonary Embolism. Thrombosis & Haemostasis 118(3), 547-552

Wolf S J, McCubbin T R, Nordenholz K E, Naviaux N W, and Haukoos J S (2008) Assessment of the pulmonary embolism rule-out criteria rule for evaluation of suspected pulmonary embolism in the emergency department. American Journal of Emergency Medicine 26(2), 181-5

Excluded clinical studies (search update)

Buntine, P., Thien, F., Stewart, J. et al. (2019) Effect of a clinical flowchart incorporating Wells score, PERC rule and age-adjusted D-dimer on pulmonary embolism diagnosis, scan rates and diagnostic yield. Emergency Medicine Australasia 31(2): 216-224

Crane, S.; Jaconelli, T.; Eragat, M. (2018) Retrospective validation of the pulmonary embolism rule-out criteria rule in 'PE unlikely' patients with suspected pulmonary embolism. European Journal of Emergency Medicine 25(3): 185-190

Gorlicki, J., Penaloza, A., Germeau, B. et al. (2019) Safety of the Combination of PERC and YEARS Rules in Patients With Low Clinical Probability of Pulmonary Embolism: A Retrospective Analysis of Two Large European Cohorts. Academic Emergency Medicine 26(1): 23-30

Malavolta, D., Quatela, V., Moffat, J. et al. (2019) Effect of the Pulmonary Embolism Rule-Out Criteria on subsequent thromboembolic events among low-risk emergency department patients: the PROPER randomized clinical trial. Internal & Emergency Medicine 14(2): 309-310

Penaloza, A., Soulie, C., Moumneh, T. et al. (2017) Pulmonary embolism rule-out criteria (PERC) rule in European patients with low implicit clinical probability (PERCEPIC): a multicentre, prospective, observational study. The Lancet Haematology 4(12): e615-e621