Supplementary Material

Comparative diagnostic accuracy of pre-test clinical probability scores for the risk stratification of patients with suspected pulmonary embolism: a systematic review and Bayesian network meta-analysis

Table of Contents

Appendix 1: Detailed description of eligibility criteria

Participants and setting

Primary studies reporting on general adult populations with sample sizes of >30 patients suspected of pulmonary embolism based on their clinical signs and symptoms were included in this review. Study populations that were recruited based on the presence of significant comorbidity (e.g., neoplasms, chronic kidney disease, myocardial infarction, or COVID) or a certain novel test out of the scope of current management guidelines (e.g., novel blood markers) were excluded. Studies that recruited based on the results of the pre-test clinical probability scores, e.g. only recruited high-risk patients, were excluded. However, as an exception, we included studies that recruited low-risk patients for the evaluation of PERC score, as this test is usually applied to low-risk patients in practice.

Retrospective studies that recruited based on the record availability of a singular test (e.g. those that recruited solely based on available CT pulmonary angiograms or d-dimer records) were excluded, as these studies discarded a portion of general adult patients suspected of pulmonary embolism (i.e., those that were considered high-risk or low-risk by the responsible physician and didn't undergo d-dimers or CT pulmonary angiograms, respectively). However, retrospective studies that recruited based on clinical suspicion, or the availability of the complete spectrum of tests on both low-risk and high-risk patients with a suspicion of PE were included (e.g., recruitment of patients that underwent the clinical pathway for PE and had either d-dimers or imaging results available). The setting was not limited to the emergency department and primary studies that recruited patients from the emergency department along with those reporting on primary care, in-patient or out-patient departments, or a combination thereof were included.

Index and reference tests

The following pre-test clinical probability scores for PE were included in the present review: 1. Wells score (including simplified, three- or two-tier versions); 2. Geneva score (including original, revised, simplified, three- or two-tier versions); 3. PERC (applied to the entire study population or the low-risk subgroup); 4. YEARS. Studies that reported enough diagnostic test accuracy indices to allow the calculation of true positives, true negatives, false positives, and false negatives for at least one of the mentioned index tests were included. Studies that only evaluated a modified version of the mentioned index tests were excluded.

Only studies where the diagnosis or ruling out of pulmonary embolism (PE) was conducted on all patients using one of the following methods were included: adherence to diagnostic pathways described in the relevant guidelines (References) and exclusive use, or a combination of any of ventilation/perfusion lung scan, digital subtraction pulmonary angiography, computed tomography pulmonary angiography (CTPA), pulmonary magnetic resonance angiography, autopsy, or clinical follow-up.

Study Characteristics

Peer-reviewed diagnostic test accuracy studies, cross-sectional studies, clinical trials, and prospective and retrospective cohort studies (along with cohort studies based on clinical trials) were included. Primary studies that didn't undergo peer review, including letters, conference papers, and abstracts were excluded. Additionally, animal, best practice, cost-effectiveness, quality improvement, and case-control studies along with guidelines, and systematic and narrative reviews were excluded. Given the large amount of available evidence, only studies published in the English language were included.

Appendix 2- Search strategy

Searches were run on June 16th, 2023 on our main databases, PubMed, Embase, and Web of Science. Google Scholar was searched on July 7th, 2023, as a supplementary database. Only the first 200 relevant records of Google Scholar were extracted for further screening. Search strategies were developed by evaluating prior relevant systematic reviews and through discussion between authors. The strategies were validated by authors experienced in designing search strategies before execution.

The records were extracted and deduplicated using the Relaxed option of the Deduplicator section of the Systematic Review Accelerator (SRA). The deduplicated records were then uploaded to abstrackr for screening. Afterward, the references of relevant systematic reviews up to April 1st, 2024 were manually searched for any missed relevant articles. The references from these reviews were extracted using Citationchaser.

PubMed	
Limits and filters	None
Search Strings	
#1	"Wells"[TW] OR ("YEARS score"[tw] OR "YEARS algorithm"[TW] OR "YEARS diagnostic"[tw] OR "Years clinical decision"[TW] OR "Years Criteria"[TW] OR "Years item*"[TW]) OR ("Geneva"[TW] OR "RGS"[TW]) OR "PEGED"[TW] OR (PERC[TW] OR "Pulmonary Embolism Rule out Criteria"[TW] OR "PE Rule out Criteria"[TW])
#2	("Decision Support Systems, Clinical"[Mesh] OR "Clinical Decision Rules"[Mesh] OR "Decision Support Techniques"[Mesh] OR "Decision Trees"[Mesh] OR Algorithms[Mesh]) OR (Algorithm*[tw] OR "Decision Tree*"[tw] OR "Decision Support Technique*"[tw] OR "clinical probabilit*"[tw] OR "Clinical decision support system*"[tw] OR "Clinical Decision Rule*"[tw] OR "clinical prediction rule*"[tw])
#3	("Pulmonary Embolism"[Mesh] OR Pulmonary Embol*[TW] OR Pulmonary Thrombo*[TW] OR “lung embol*”[TW] OR “lung thrombo*”[TW])
#4	"Animals"[Mesh] NOT "Humans"[Mesh]
#5	(#1 OR #2) AND #3 NOT #4
Total Results	2,868

Embase	
Limits and filters	None
Search Strings	
#1	‘wells’:ti,ab,kw OR ('years score':ti,ab,kw OR ‘Years algorithm’:ti,ab,kw OR ‘years diagnostic’:ti,ab,kw OR ‘years clinical decision’:ti,ab,kw OR 'years criteria':ti,ab,kw OR 'years rule':ti,ab,kw OR ‘Years item*’:ti,ab,kw) OR (‘geneva’:ti,ab,kw OR ‘RGS’:ti,ab,kw) OR 'peged':ti,ab,kw OR ('perc':ti,ab,kw OR ‘Pulmonary Embolism Rule out Criteria’:ti,ab,kw OR ‘PE Rule out Criteria’:ti,ab,kw)
#2	('decision support system'/exp OR 'decision tree'/exp OR 'algorithm'/exp) OR ‘clinical decision support system*':ti,ab,kw OR 'decision tree*’:ti,ab,kw OR ‘decision support technique’:ti,ab,kw OR 'clinical probabilit*':ti,ab,kw OR 'clinical decision rule*':ti,ab,kw OR 'clinical prediction rule*':ti,ab,kw OR ‘algorithm*’:ti,ab,kw
#3	('lung embolism'/exp OR 'lung embol*':ti,ab,kw OR 'lung thrombo*':ti,ab,kw OR 'pulmonary embol*':ti,ab,kw OR 'pulmonary thrombo*':ti,ab,kw)
#4	('animal'/exp OR 'nonhuman'/exp) NOT 'human'/exp
#5	(#1 OR #2) AND #3 NOT #4
#6	'conference abstract'/it OR 'conference paper'/it
#7	#5 NOT #6
Total Results	3,501

Web of Science	
Limits and filters	None
Search Strings	
#1	TS =("Wells" OR ("YEARS score" OR "YEARS algorithm" OR "YEARS diagnostic" OR "Years clinical decision" OR "Years Criteria"OR "Years item*") OR ("Geneva" OR "RGS") OR ("PEGED") OR (PERC OR "Pulmonary Embolism Rule out Criteria" OR "PE Rule out Criteria"))
#2	TS = (Algorithm* OR "Decision Tree*" OR "Decision Support Technique*" OR "clinical probability" OR "Clinical decision support system*" OR "Clinical Decision Rule*" OR "clinical prediction rule*")
#3	TS =(”Pulmonary Embolism*” OR “Pulmonary Embol*” OR “Pulmonary Thrombo*” OR “Lung embol*” OR “Lung Thrombo*”)
#4	(#1 OR #2) AND #3
Total Results	2,531

Google Scholar	
Limits and filters	None
Search Strings	
#1	Wells OR "YEARS score" OR "YEARS algorithm" OR "YEARS diagnostic" OR "Years clinical" OR Geneva OR RGS OR PEGED OR PERC OR "Pulmonary Embolism Rule out Criteria" OR "PE Rule out Criteria" "Pulmonary Embolism" -"Systematic review" -"meta analysis" -"COVID"
Total Results*	17,300

*Only the first 200 relevant records were extracted

Appendix 3- Data collection form

Code	Explanation
digitized	Were tools used to extract data from figures?
studyid	Study DOI
title	Study Title
author_year	First author (publication year)
design	Study Design
Cohort	Did the study use another cohort as its sample?
location	Country where the study was conducted
recruit_method	Clinical suspicion/Presence of records
setting	ED, In patient, out-patient or mixed
sample_size	Sample size relevant to our study
pe_prevalence	Prevalence of PE in the study
hx_vte	History of VTE in the study
age	Mean (or median) age
sex (male : female)	Percent (Male:Female)
ref_standard	What was the reference standard in the study?
gold standard for all	Was gold standard applied to all participants?
assessor	Who assessed the CPS?
time_after_presentation (months)	Time between presentation and diagnosis
fu	Months of follow-up
fu_method	How were the participants followed?
lost	Number lost to FU
n_test	How many tests were evaluated in the study
test names	Name of the evaluated tests
cut_off_1	What were the used cut-offs?
low_risk_with_pe_1	Low-risk patients with PE ruled in
low_risk_no_pe_1	Low-risk patients with PE ruled out
moderate_risk_with_pe_1	Moderate-risk patients with PE ruled in
moderate_risk_no_pe_1	Moderate-risk patients with PE ruled out
high_risk_with_pe_1	High-risk patients with PE ruled in
high_risk_no_pe_1	High-risk patients with PE ruled out

Appendix 4- Decision rule to include one record per cohort

The following decision rules are applied in order of priority, from top to bottom.

1. For studies using a single external cohort, possibly shared across multiple studies (i .e., single cohort studies):
    1. Studies referencing insufficiently described unpublished data will be excluded.
    2. If the original referenced cohort fits the criteria of the present review:
        1. We will prioritize that original study over the later studies that used its data.
    3. If the original referenced cohort doesn't fit the eligibility criteria of the present review (e.g., the original study has reported on a modified pretest probability), but more than one secondary study does:
        1. We will prioritize the study with the larger analysis sample size.
1. For studies that used more than one external cohort in their analyses (i.e., multiple cohort studies):
    1. If the outcomes are reported separately for each external cohort:
        1. We will treat each case like those using a single original external cohort described in section 1.
    2. If the outcomes are reported for one combined external cohort:
        1. If none of the referenced cohorts are used in any other study:
            1. We will treat the study as one independent cohort.
        2. If any of the referenced cohorts are used in other single cohort studies:
            1. We will prioritize the single cohort study over the multiple cohort one and treat it as previously described in section 1.
        3. If any of the referenced cohorts are used in other multiple cohort studies:
            1. We will prioritize studies with the least overlap with other included single or multiple cohort studies, so that the maximum possible amount of independent data is included.

Appendix 5- Model building using the bivariate hierarchical method

We considered two main bivariate hierarchical models: (1) the "Equal variance model", assuming equal variance, and (2) the "Unequal variance model", allowing different variances for sensitivity and specificity of each score. In case of model divergence, we constrained indices with low variance for random effect (<0·1) to fixed effects. If there was a high correlation (>0·95) between the variances of random effects for sensitivity and specificity of a test, only the index with the higher variance was kept as a random effect and the other was constrained to fixed effect. The likelihood ratio test was used to choose the final bivariate hierarchical model.

In the unequal variance model, due to the high correlation between random effects of sensitivity and specificity for the revised Geneva (3 tier) score and PERC (when applied to the low-risk sub-population), we constrained the specificity index for revised Geneva and the sensitivity index for PERC (low-risk) to fixed effects. These constraints allowed the convergence of the model which outperformed the equal variance model (likelihood ratio test, Chi^2=39·7, p<0·001) and was subsequently chosen as the representative model for the bivariate hierarchical model.

Appendix 6- PRISMA flow diagram

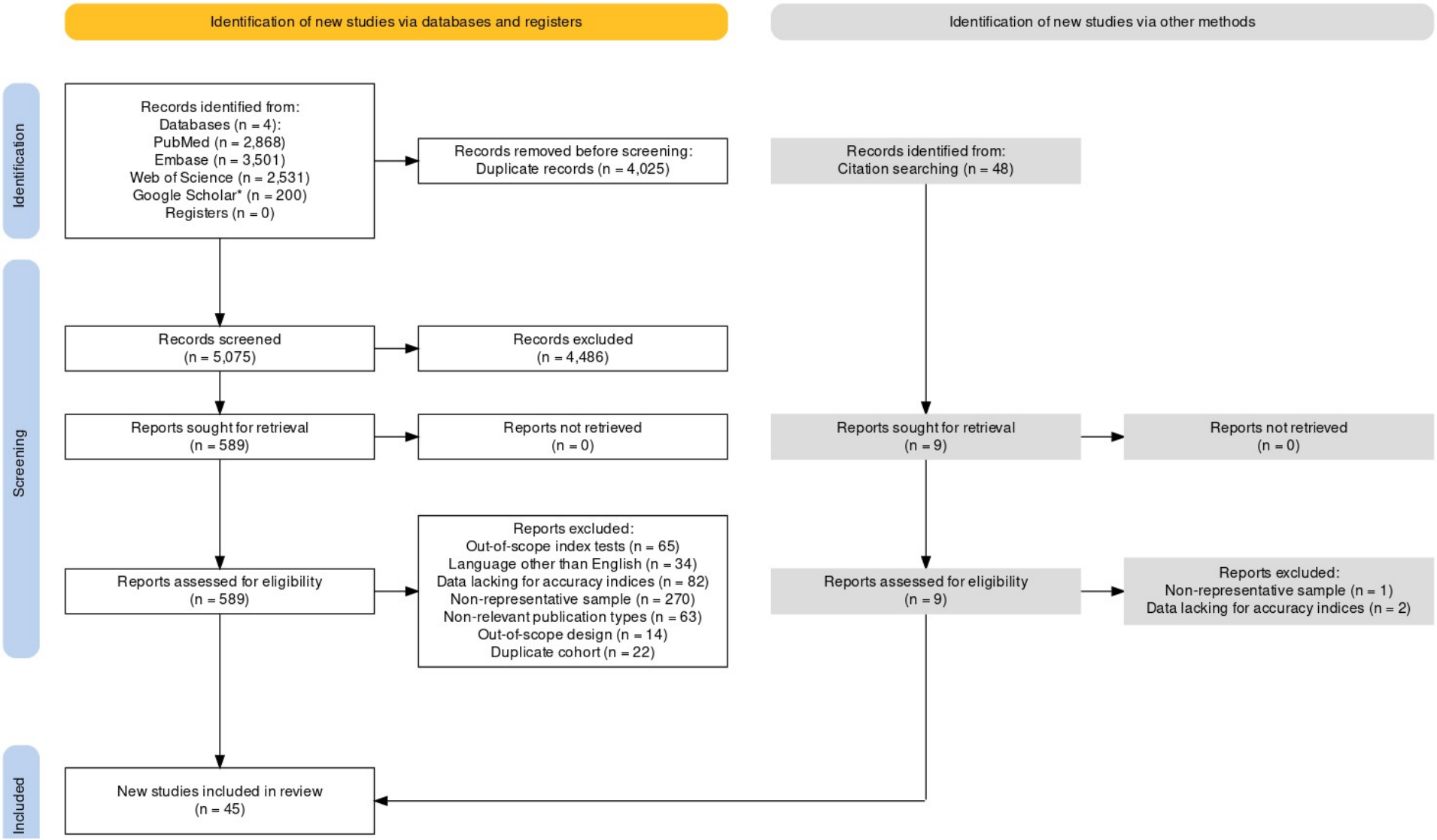

Figure a- PRISMA flow diagram for included studies

Appendix 7- Baseline Characteristics of the included primary studies

n	Author (Year)	Design	Country	Setting	Age[1]	Sex: F(%)	Hx VTE (%)	Patient ID	Assessor	Reference Standard	Tests
1	Kearon (2006)	Prospective	Canada	Outpatient; Inpatient	57	65	12	Clinical Suspicion	Physician; Nurse	Algorithm; Followup	W3
2	Esiene(2019)	Cross-sectional	Cameroon	ED	53·7	64	6·7	Clinical Suspicion	Unclear	CTPA	W3; RG3; SRG2
3	Steeghs (2005)	Prospective	Netherlands	ED	51	62	9·1	Clinical Suspicion	Physician	Algorithm; Followup	W2
4	Kurt (2010)	Prospective	Turkey	ED; Inpatients	NR[2]	40	12	Clinical Suspicion	Unclear	CTPA	W3
5	Penaloza (2017)	Prospective	France; Belgium	ED	53	58	NR	Clinical Suspicion	Physician	Algorithm; Followup	PRC
6	Kline (2008)	Prospective	USA; New Zeeland	ED; Outpatient	49·1	67	11	Availability of tests	Physician	Algorithm; Followup	PRC
7	Gulsen(2015)	Prospective	Turkey	ED; Outpatient	56·5	48	NR	Clinical Suspicion	Unclear	CTPA	W3
8	Obradovic(2015)	Prospective	Serbia	Independent Institute	51·2	45	NR	Clinical Suspicion	Unclear	CTPA	W3; W2
9	Van Belle (2006)	Prospective	Netherlands	Outpatient; Inpatient	53	58	NR	Clinical Suspicion	Attending physician	Algorithm; Followup	W2
10	Galipienzo (2012)	Prospective	Spain	ED	65·2	51	12·9	Clinical Suspicion	Attending physician	Algorithm; Followup	W2
11	Castelli (2009)	Retrospective	Italy	ED	NR	NR	NR	Clinical Suspicion	Unclear	Algorithm; Followup	W3
12	van der Hulle (2017)	Prospective	Netherlands	Outpatient; Inpatient	53	62	10	Clinical Suspicion	Attending physician	Algorithm; Followup	YRS

n	Author (Year)	Design	Country	Setting	Age[1]	Sex: F(%)	Hx VTE (%)	Patient ID	Assessor	Reference Standard	Tests
13	Aujesky (2003)	Prospective	Switzerland	ED	61	58	NR	Clinical Suspicion	Physician	Algorithm; Followup	G3
14	Geersing (2012)	Prospective	Netherlands	Primary care	48	71	14	Clinical Suspicion	Physician	Algorithm; Followup	W2
15	Douma (2011)	Prospective	Netherlands	Outpatient; Inpatient	53	61	4·8	Clinical Suspicion	Physician	Algorithm; Followup	W2; RG2; SW; SRG2
16	Goekoop(2007)	Prospective	Netherlands	ED; Outpatient	51	63	9·4	Clinical Suspicion	Physician	Algorithm; Followup	W2
17	Golshani (2020)	Cross-sectional	Iran	ED	56·29	39	25	Clinical Suspicion	Unclear	CTPA	W2
18	Kearon (2020)	Prospective	Canada	Outpatient	52	67	8·13	Clinical Suspicion	Physician	Algorithm; Followup	W3
19	Miniati (2005)	Prospective	Italy	Independent Institute	70	54	NR	Clinical Suspicion	Physician	Pulmonary Angiogram	W3; G3
20	Sanjuan(2014)	Retrospective	Spain	ED	66·9	56	NR	Availability of tests (DD and CTPA)	Unclear	MSCT; V/Q; CXR; compression US; FU	RG3
21	Penaloza (2013)	Retrospective	France; Belgium	ED	64	62	22·8	Clinical Suspicion	Physician	Algorithm; Followup	W3; RG3
22	Aydogdu(2014)	Retrospective	Turkey	ED	64	48	NR	Availability of tests	Physician	CTPA; Lower extremity Doppler USG(only in 4 patients)	W3; PRC
23	Wells(2000)	Retrospective	Canada	Outpatient; Inpatient	NR	NR	NR	Clinical Suspicion	Unclear	Algorithm; Followup	W3; W2
24	Wit (2022)	Prospective	Canada	ED	62	58	NR	Clinical Suspicion	Physician	Algorithm; Followup	YRS

n	Author (Year)	Design	Country	Setting	Age[1]	Sex: F(%)	Hx VTE (%)	Patient ID	Assessor	Reference Standard	Tests
25	Calisir (2009)	Cross-sectional	Turkey	ED; Inpatients	62	48	NR	Clinical Suspicion	Physician	CTA and CTV	W3; RG3
26	Sanson (2000)	Prospective	Netherlands	Outpatient; Inpatient	51	58	14	Clinical Suspicion	Attending physician	lung scintigraphy; Ventilation; Angiogram	W3
27	Martinez(2013)	Cross-sectional	Argentina	Inpatients	68	42	7·8	Clinical Suspicion	Medical student	Algorithm; Followup	W2
28	Gupta (2009)	Prospective	US	ED	46·9	66	NR	Clinical Suspicion	Unclear	CTPA	RG3
29	Kucher (2003)	Prospective	Switzerland	ED	62·09	41	15·18	Clinical Suspicion	Unclear	Algorithm; Followup	W3
30	Wells (2001)	Prospective	Canada	ED	50·5	63	NR	Clinical Suspicion	Physician	Algorithm; Followup	W3
31	Kline (2004)	Prospective*	US	ED	47	60	12	Clinical Suspicion	Nurse; Physician	Algorithm; Followup	PRC
32	Hasanoglu(2019)	Prospective	Turkey	ED	69	54	NR	Clinical Suspicion	Physician	CTPA	W2
33	Theunissen (2016)	Retrospective	Netherlands	ED	58	47	6·4	Clinical Suspicion; Availability of Tests (DD)	Attending physician	Algorithm; Followup	PRC
34	Hugli(2011)	Prospective	Switzerland; France; Belgium	ED	61	55	17·9	Clinical Suspicion	Unclear	Algorithm based on MDCT, PA, V/Q scan or US; Followup	PRC
35	Hogg (2011)	Prospective	England	ED	38	51	12	Clinical Suspicion	Physician	Algorithm; Followup	W3

n	Author (Year)	Design	Country	Setting	Age[1]	Sex: F(%)	Hx VTE (%)	Patient ID	Assessor	Reference Standard	Tests
36	Penaloza (2007)	Prospective	Belgium	ED	56	59	15	Clinical Suspicion	Attending physician	Algorithm; Followup	W3; W2
37	Ebadi(2017)	Prospective	Belgium; France; Netherlands; Switzerland	ED	63	56	14·1	Clinical Suspicion	Attending physician	Algorithm; Followup	RG3; SRG3
38	Le Gal [D·] (2006)[3]	Prospective	France; Switzerland	ED	60·6	59	17·2	Clinical Suspicion	Physician	Algorithm; Followup	RG3
39	Le Gal [V·] (2006)[3]	Prospective	France; Switzerland	ED	60	60	19	Clinical Suspicion	Physician	Algorithm; Followup	RG3
40	Freund (2018)	Prospective Trial	France	ED	44	98	3	Clinical Suspicion	Physician	Algorithm	PRC
41	Leclerq(2003)	Prospective	Netherlands	Outpatient; Inpatient	52	61	14	Clinical Suspicion	Attending physician	Algorithm	W3
42	Wicki (2001)	Retrospective	Switzerland	ED	62	55	19	Clinical Suspicion	Physician	Lung scan, US	G3
43	Kline_1 (2006)[4]	Prospective	US	ED	48·4	NR	NR	Clinical Suspicion	Physician	CTPA; Scan; Follwoup	W3
44	Kabrhel (2005)	Prospective	US	ED	47·9	74	20	Availability of tests	Physician	CTPA; Angiogram; Followup	W3; W2
45	Anderson (2005)	Prospective	Canada	ED	50·2	65	9	Clinical Suspicion	Physician	CTPA	W3
46	Kline_2 (2006)[4]	Prospective	US	ED	44·7	69	7	Clinical Suspicion	Physician	CTPA	W3

[1]Mean/ Median depending on the primary study

[2]NR: Not reported

[3]Le Gal et al· (2006) used two separate datasets as derivation and validation samples in the same study.

[4]Kline et al. published two studies in 2006.

Hx: History of; F: Female; ID: Identification; W3: Wells (3-tier); W2: Wells (2-tier); RG3: Revised Geneva (3-tier); PRC: PERC

Appendix 8- Results of Individual Primary Studies with PE Outcomes in Clinical Probability Categories

n	Author Year	Clinical Probability Score	Cut-Off	Analysis Sample Size	PE Prevalence (%)	Clinical Probability	PE Positive	PE Negative
1	Anderson (2005)			858	9·6	Low	10	469
						Moderate	46	278
						High	26	29
2	Aydogdu(2014)			108	49·1	Low	4	9
						Moderate	26	34
						High	23	12
3	Calisir (2009)			148	32·4	Low	4	47
						Moderate	18	50
						High	26	3
4	Castelli (2009)	Wells (3 Tier)	2,6	585	39·8	Low	24	98
						Moderate	158	235
						High	51	19
5	Esiene(2019)			30	53·3	Low	2	4
						Moderate	11	9
						High	3	1
6	Gulsen(2015)			62	41·9	Low	11	32
						Moderate	8	4
						High	7	0
7	Hogg (2011)			425	5·2	Low	12	361
						Moderate	7	31
						High	3	11

n	Author Year	Clinical Probability Score	Cut-Off	Analysis Sample Size	PE Prevalence (%)	Clinical Probability	PE Positive	PE Negative
8	Kabrhel (2005)			607	10·0	Low	13	312
						Moderate	32	202
						High	16	32
9	Kearon (2006)			1,113	14·4	Low	27	634
						Moderate	94	287
						High	39	32
10	Kearon (2020)		4·5, 6·5	2,017	7·4	Low	87	1,665
						Moderate	43	175
						High	19	28
11	Kline_1 (2006)			178	13·5	Low	3	107
						Moderate	13	42
						High	8	5
12	Kline_2 (2006)			2,302	4·7	Low	50	1,654
						Moderate	48	511
						High	10	29
13	Kucher (2003)		2,6	191	44·5	Low	11	88
						Moderate	19	15
						High	55	3
14	Kurt (2010)			58	69·0	Low	1	6
						Moderate	18	12
						High	21	0
15	Leclerq(2003)			202	29·2	Low	25	77

n	Author Year	Clinical Probability Score	Cut-Off	Analysis Sample Size	PE Prevalence (%)	Clinical Probability	PE Positive	PE Negative
						Moderate	18	50
						High	16	16
16	Miniati (2005)			215	43·3	Low	8	56
						Moderate	64	54
						High	21	12
17	Obradovic(2015)			60	83·3	Low	1	0
						Moderate	33	10
						High	16	0
18	Penaloza (2007)			185	18·4	Low	2	99
						Moderate	25	52
						High	7	0
19	Penaloza (2013)			1,038	31·3	Low	61	425
						Moderate	221	257
						High	43	31
20	Sanson (2000)			414	29·5	Low	41	106
						Moderate	78	181
						High	3	5
21	Wells (2001)			930	8·7	Low	6	521
						Moderate	52	287
						High	23	41
22	Wells(2000)			1,219	16·5	Low	16	475
						Moderate	129	510

n	Author Year	Clinical Probability Score	Cut-Off	Analysis Sample Size	PE Prevalence (%)	Clinical Probability	PE Positive	PE Negative
						High	56	33
23	Douma (2011)			807	22·9	Low	90	494
						High	95	128
24	Galipienzo (2012)			231	24·7	Low	33	145
						High	24	29
25	Geersing (2012)			598	12·0	Low	21	401
						High	51	125
26	Goekoop(2007)			876	12·3	Low	80	700
						High	28	68
27	Golshani (2020)			104	37·5	Low	21	26
						High	18	39
28	Hasanoglu(2019)	Wells (2 Tier)	4	250	66·0	Low	63	61
						High	102	24
29	Kabrhel (2005)			607	10·0	Low	25	424
						High	36	122
30	Martinez(2013)			613	36·4	Low	77	317
						High	146	73
31	Obradovic(2015)			60	83·3	Low	5	2
						High	45	8
32	Penaloza (2007)			185	18·4	Low	10	134
						High	24	17
33	Steeghs (2005)			331	13·9	Low	30	249
						High	16	36

n	Author Year	Clinical Probability Score	Cut-Off	Analysis Sample Size	PE Prevalence (%)	Clinical Probability	PE Positive	PE Negative
34	Van Belle (2006)			3,306	20·4	Low	266	1,940
						High	408	692
35	Wells(2000)			1,211	16·7	Low	63	804
						High	139	205
36	Aujesky (2003)			259	29·7	Low	10	106
						Moderate	41	58
						High	26	18
37	Miniati (2005)	Geneva Score (3 Tier)	4, 9	215	43·3	Low	13	13
						Moderate	50	78
						High	30	31
38	Wicki (2001)			986	27·1	Low	50	436
						Moderate	166	271
						High	51	12
39	Calisir (2009)			148	32·4	Low	0	15
						Moderate	28	81
						High	20	4
40	Ebadi(2017)	Revised Geneva (3 Tier)	4, 11	1,621	18·1	Low	62	563
						Moderate	202	732
						High	30	32
41	Esiene(2019)			30	53·3	Low	5	5
						Moderate	10	9
						High	1	0
42	Gupta (2009)			627	4·5	Low	6	275

n	Author Year	Clinical Probability Score	Cut-Off	Analysis Sample Size	PE Prevalence (%)	Clinical Probability	PE Positive	PE Negative
						Moderate	17	313
						High	5	11
43	Le Gal [D·] (2006)			956	23·1	Low	32	322
						Moderate	151	398
						High	38	15
44	Le Gal [V·] (2006)			749	25·6	Low	18	211
						Moderate	132	331
						High	42	15
45	Penaloza (2013)			1,038	31·3	Low	35	235
						Moderate	222	447
						High	68	31
46	Sanjuan(2014)			3,242	10·9	Low	61	1,220
						Moderate	241	1,596
						High	50	74
47	Ebadi(2017)	Simplified Revised Geneva (3 Tier)	2, 5	1,621	18·1	Low	59	549
						Moderate	220	760
						High	15	18
48	Esiene(2019)			30	53·3	Low	6	6
						Moderate	10	6
						High	0	2
49	Aydogdu(2014)	PERC rule (All)	1	108	49·1	Low	1	4
						High	52	51
50	Hugli(2011)			1,675	21·3	Low	12	209

n	Author Year	Clinical Probability Score	Cut-Off	Analysis Sample Size	PE Prevalence (%)	Clinical Probability	PE Positive	PE Negative
						High	345	1,109
51	Kline (2008)			8,138	4·6	Low	19	1,933
						High	352	5,834
52	Freund (2018)			958	1·5	Low	1	458
						High	13	486
53	Hugli(2011)			587	9·7	Low	12	176
						High	45	354
54	Kline (2004)	PERC rule (Low-risk)		1,427	8·0	Low	5	357
						High	109	956
55	Penaloza (2017)			1,050	4·5	Low	4	331
						High	43	672
56	Theunissen (2016)			377		Low	1	77
						High	16	283
57	Wit (2022)			1,703	8·0	Low	67	1,047
		YEARS rule				High	69	520
58	van der hulle (2017)			3,465	13·2	Low	55	1,688
						High	404	1,318
59		Simplified Wells				Low	65	434
						High	120	188
60	Douma (2011)	Revised Geneva (2 tiers)	5	807	22·9	Low	88	465
						High	97	157
61		Simplified Revised Geneva (2 tiers)	2			Low	95	481
						High	90	141

Appendix 9-The included pretest clinical probability scores, their items and scoring

1. Wells Score	
Items	Points
• Clinical signs and symptoms of DVT	+3
• PE is #1 diagnosis OR equally likely	+3
• Heart rate > 100	+1.5
• Immobilization at least 3 days OR surgery in the previous 4 weeks	+1.5
• Previous, objectively diagnosed PE or DVT	+1.5
• Hemoptysis	+1
• Malignancy w/ treatment within 6 months or palliative	+1
Clinical Probability	
Three-tier Wells	Score
• Low risk	<2
• Moderate risk	2-6
• High risk	>6
Two-tier Wells	Score
• PE unlikely	<4
• PE likely	≥4
2. Revised Geneva Score	
Items	Points
• Age >65	+1
• Previous DVT or PE	+3
• Surgery (under general anesthesia) or lower limb fracture in past month	+2
• Active malignant condition (solid or hematologic; currently active or considered cured less than 1 year)	+2
• Unilateral lower limb pain	+3
• Hemoptysis	+2
Heart rate:	
• <75	0
• 75-94	+3
• ≥95	+5
• Pain on deep venous palpation of lower limb and unilateral edema	+4
Clinical Probability	
Three-tier Revised Geneva	Score
• Low risk	<4
• Moderate risk	4-10
• High risk	≥11
3. PERC Score	
Items	Points
• Age ≥50	+1
• HR ≥100	+1
• O_2 sat on room air <95%	+1
• Unilateral leg swelling	+1
• Hemoptysis	+1
• Recent surgery (under general anesthesia) or trauma within the previous four weeks	+1
• Prior PE or DVT	+1
• Hormone use (Oral contraception, hormone replacement therapy or estrogen)	+1
Clinical Probability	
PERC	Score
• PE unlikely	0
• PE likely	≥1

Appendix 10- Forest plots for the sensitivity and specificity of the included primary studies

Study	TP	FP	FN	TN	Sens (95% CI)	Spec (95% CI)
Kline(2008) PRC_A	352	5834	19	1933	0.95 (0.92-0.97)	0.25 (0.24-0.26)
Hugli(2011) PRC_A	345	1109	12	209	0.97 (0.95-0.98)	0.16 (0.14-0.18)
Aydogdu(2014) PRC_A	52	51	1	4	0.98 (0.90-1.00)	0.07 (0.03-0.17)
Theunissen(2016) PRC_L	16	283	1	77	0.94 (0.73-0.99)	0.21 (0.17-0.26)
Freund(2018) PRC_L	13	486	1	458	0.93 (0.69-0.99)	0.49 (0.46-0.52)
Kline(2004) PRC_L	109	956	5	357	0.96 (0.91-0.98)	0.27 (0.25-0.29)
Hugli(2011) PRC_L	45	354	12	176	0.79 (0.67-0.88)	0.33 (0.29-0.37)
Penaloza(2017) PRC_L	43	672	4	331	0.91 (0.79-0.96)	0.33 (0.30-0.36)
Penaloza(2013) RG3	290	478	35	235	0.89 (0.85-0.92)	0.33 (0.30-0.37)
Gupta(2009) RG3	22	324	6	275	0.79 (0.61-0.90)	0.46 (0.42-0.50)
Sanjuan(2014) RG3	291	1670	61	1220	0.83 (0.79-0.87)	0.42 (0.40-0.44)
Ebadi(2017) RG3	232	764	62	563	0.79 (0.74-0.83)	0.42 (0.39-0.45)
LeGal [D.] (2006) RG3	189	413	32	322	0.86 (0.81-0.90)	0.44 (0.40-0.48)
LeGal [V.] (2006) RG3	174	346	18	211	0.91 (0.86-0.94)	0.38 (0.34-0.42)
Calisir(2009) RG3	48	85	0	15	1.00 (0.93-1.00)	0.15 (0.09-0.23)
Esiene(2019) RG3	11	9	5	5	0.69 (0.45-0.86)	0.36 (0.17-0.61)
Geersing(2012) W2	51	125	21	401	0.71 (0.60-0.80)	0.76 (0.72-0.79)
Goekoop(2007) W2	28	68	80	700	0.26 (0.19-0.35)	0.91 (0.89-0.93)
Steeghs(2005) W2	16	36	30	249	0.35 (0.23-0.49)	0.87 (0.83-0.90)
Wells(2000) W2	139	205	63	804	0.69 (0.62-0.75)	0.80 (0.77-0.82)
Penaloza(2007) W2	24	17	10	134	0.71 (0.54-0.83)	0.89 (0.83-0.93)
VanBelle(2006) W2	408	692	266	1940	0.61 (0.57-0.65)	0.74 (0.72-0.76)
Douma(2011) W2	95	128	90	494	0.51 (0.44-0.58)	0.79 (0.76-0.82)
Galipienzo(2012) W2	24	29	33	145	0.42 (0.30-0.55)	0.83 (0.77-0.88)
Martinez(2013) W2	146	73	77	317	0.65 (0.59-0.71)	0.81 (0.77-0.85)
Golshani(2020) W2	18	39	21	26	0.46 (0.31-0.61)	0.40 (0.29-0.52)
Kabrhel(2005) W2	36	122	25	424	0.59 (0.46-0.70)	0.78 (0.74-0.81)
Hasanoglu(2019) W2	102	24	63	61	0.62 (0.54-0.69)	0.72 (0.62-0.80)
Obradovic(2015) W2	45	8	5	2	0.90 (0.79-0.96)	0.20 (0.06-0.51)
Penaloza(2013) W3	264	288	61	425	0.81 (0.76-0.85)	0.60 (0.56-0.64)
Kline_2(2006) W3	58	540	50	1654	0.54 (0.45-0.63)	0.75 (0.73-0.77)
Hogg(2011) W3	10	42	12	361	0.45 (0.27-0.65)	0.90 (0.87-0.93)
Wells(2001) W3	75	328	6	521	0.93 (0.85-0.97)	0.61 (0.58-0.64)
Anderson(2005) W3	72	307	10	469	0.88 (0.79-0.93)	0.60 (0.57-0.63)
Kline_1(2006) W3	21	47	3	107	0.88 (0.70-0.96)	0.69 (0.61-0.76)
Kearon(2006) W3	133	319	27	634	0.83 (0.76-0.88)	0.67 (0.64-0.70)
Wells(2000) W3	185	543	16	475	0.92 (0.87-0.95)	0.47 (0.44-0.50)
Penaloza(2007) W3	32	52	2	99	0.94 (0.81-0.98)	0.66 (0.58-0.73)
Leclerq(2003) W3	34	66	25	77	0.58 (0.45-0.70)	0.54 (0.46-0.62)
Sanson(2000) W3	81	186	41	106	0.66 (0.57-0.74)	0.36 (0.31-0.42)
Calisir(2009) W3	44	53	4	47	0.92 (0.81-0.97)	0.47 (0.38-0.57)
Castelli(2009) W3	209	254	24	98	0.90 (0.85-0.93)	0.28 (0.24-0.33)
Gulsen(2015) W3	15	4	11	32	0.58 (0.39-0.75)	0.89 (0.75-0.96)
Miniati(2005) W3	85	66	8	56	0.91 (0.83-0.95)	0.46 (0.37-0.55)
Kucher(2003) W3	74	18	11	88	0.87 (0.78-0.93)	0.83 (0.75-0.89)
Aydogdu(2014) W3	49	46	4	9	0.92 (0.82-0.97)	0.16 (0.09-0.28)
Esiene(2019) W3	14	10	2	4	0.88 (0.65-0.97)	0.29 (0.12-0.55)
Kabrhel(2005) W3	48	234	13	312	0.79 (0.67-0.87)	0.57 (0.53-0.61)
Kurt(2010) W3	39	12	1	6	0.98 (0.88-1.00)	0.33 (0.16-0.56)
Obradovic(2015) W3	49	10	1	0	0.98 (0.90-1.00)	0.00 (0.00-0.28)

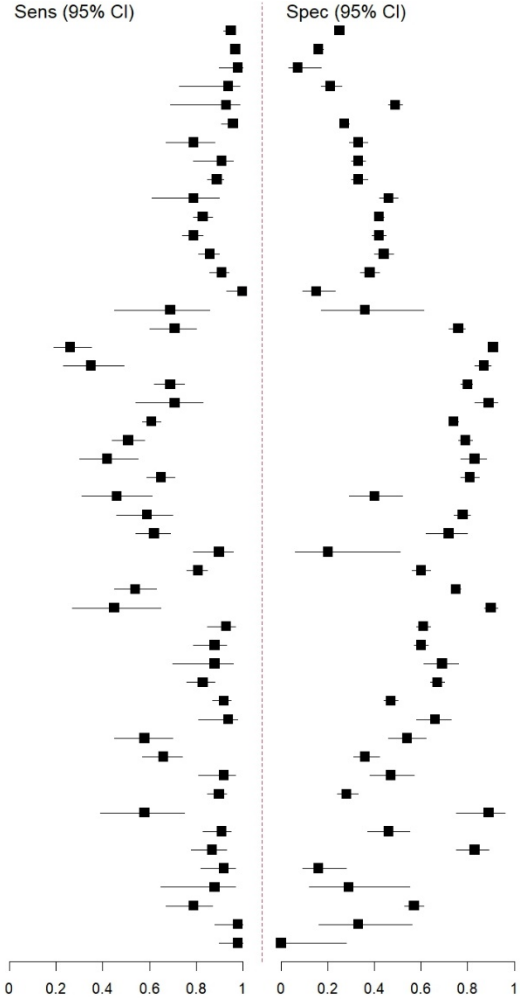

Figure b- PE rule-out model

Study	TP	FP	FN	TN	Sens (95% CI)	Spec (95% CI)
Kline(2008) PRC_A	352	5834	19	1933	0.95 (0.92-0.97)	0.25 (0.24-0.26)
Hugli(2011) PRC_A	345	1109	12	209	0.97 (0.95-0.98)	0.16 (0.14-0.18)
Aydogdu(2014) PRC_A	52	51	1	4	0.98 (0.90-1.00)	0.07 (0.03-0.17)
Theunissen(2016) PRC_L	16	283	1	77	0.94 (0.73-0.99)	0.21 (0.17-0.26)
Freund(2018) PRC_L	13	486	1	458	0.93 (0.69-0.99)	0.49 (0.46-0.52)
Kline(2004) PRC_L	109	956	5	357	0.96 (0.91-0.98)	0.27 (0.25-0.29)
Hugli(2011) PRC_L	45	354	12	176	0.79 (0.67-0.88)	0.33 (0.29-0.37)
Penaloza(2017) PRC_L	43	672	4	331	0.91 (0.79-0.96)	0.33 (0.30-0.36)
Penaloza(2013) RG3	68	31	257	682	0.21 (0.17-0.26)	0.96 (0.94-0.97)
Gupta(2009) RG3	5	11	23	588	0.18 (0.08-0.36)	0.98 (0.97-0.99)
Sanjuan(2014) RG3	50	74	302	2816	0.14 (0.11-0.18)	0.97 (0.96-0.98)
Ebadi(2017) RG3	30	32	264	1295	0.10 (0.07-0.14)	0.98 (0.97-0.99)
LeGal [D.] (2006) RG3	38	15	183	720	0.17 (0.13-0.23)	0.98 (0.97-0.99)
LeGal [V.] (2006) RG3	42	15	150	542	0.22 (0.17-0.28)	0.97 (0.95-0.98)
Calisir(2009) RG3	20	4	28	96	0.42 (0.29-0.56)	0.96 (0.90-0.98)
Esiene(2019) RG3	1	0	15	14	0.06 (0.01-0.28)	1.00 (0.78-1.00)
Geersing(2012) W2	51	125	21	401	0.71 (0.60-0.80)	0.76 (0.72-0.79)
Goekoop(2007) W2	28	68	80	700	0.26 (0.19-0.35)	0.91 (0.89-0.93)
Steeghs(2005) W2	16	36	30	249	0.35 (0.23-0.49)	0.87 (0.83-0.90)
Wells(2000) W2	139	205	63	804	0.69 (0.62-0.75)	0.80 (0.77-0.82)
Penaloza(2007) W2	24	17	10	134	0.71 (0.54-0.83)	0.89 (0.83-0.93)
VanBelle(2006) W2	408	692	266	1940	0.61 (0.57-0.65)	0.74 (0.72-0.76)
Douma(2011) W2	95	128	90	494	0.51 (0.44-0.58)	0.79 (0.76-0.82)
Galipienzo(2012) W2	24	29	33	145	0.42 (0.30-0.55)	0.83 (0.77-0.88)
Martinez(2013) W2	146	73	77	317	0.65 (0.59-0.71)	0.81 (0.77-0.85)
Golshani(2020) W2	18	39	21	26	0.46 (0.31-0.61)	0.40 (0.29-0.52)
Kabrhel(2005) W2	36	122	25	424	0.59 (0.46-0.70)	0.78 (0.74-0.81)
Hasanoglu(2019) W2	102	24	63	61	0.62 (0.54-0.69)	0.72 (0.62-0.80)
Obradovic(2015) W2	45	8	5	2	0.90 (0.79-0.96)	0.20 (0.06-0.51)
Penaloza(2013) W3	43	31	282	682	0.13 (0.10-0.17)	0.96 (0.94-0.97)
Kline_2(2006) W3	10	29	98	2165	0.09 (0.05-0.16)	0.99 (0.98-0.99)
Hogg(2011) W3	3	11	19	392	0.14 (0.05-0.34)	0.97 (0.95-0.98)
Wells(2001) W3	23	41	58	808	0.28 (0.19-0.39)	0.95 (0.93-0.96)
Anderson(2005) W3	26	29	56	747	0.32 (0.23-0.43)	0.96 (0.94-0.97)
Kline_1(2006) W3	8	5	16	149	0.33 (0.18-0.53)	0.97 (0.93-0.99)
Kearon(2006) W3	39	32	121	921	0.24 (0.18-0.31)	0.97 (0.96-0.98)
Wells(2000) W3	56	33	145	985	0.28 (0.22-0.35)	0.97 (0.96-0.98)
Penaloza(2007) W3	7	0	27	151	0.21 (0.11-0.37)	1.00 (0.98-1.00)
Leclerq(2003) W3	16	16	43	127	0.27 (0.17-0.39)	0.89 (0.83-0.93)
Sanson(2000) W3	3	5	119	287	0.02 (0.01-0.06)	0.98 (0.96-0.99)
Calisir(2009) W3	26	3	22	97	0.54 (0.40-0.67)	0.97 (0.92-0.99)
Castelli(2009) W3	51	19	182	333	0.22 (0.17-0.28)	0.95 (0.92-0.97)
Gulsen(2015) W3	7	0	19	36	0.27 (0.14-0.46)	1.00 (0.90-1.00)
Miniati(2005) W3	21	12	72	110	0.23 (0.16-0.33)	0.90 (0.83-0.94)
Kucher(2003) W3	55	3	30	103	0.65 (0.54-0.74)	0.97 (0.92-0.99)
Aydogdu(2014) W3	23	12	30	43	0.43 (0.31-0.56)	0.78 (0.65-0.87)
Esiene(2019) W3	3	1	13	13	0.19 (0.07-0.43)	0.93 (0.69-0.99)
Kabrhel(2005) W3	16	32	45	514	0.26 (0.17-0.38)	0.94 (0.92-0.96)
Kurt(2010) W3	21	0	19	18	0.52 (0.37-0.67)	1.00 (0.82-1.00)
Obradovic(2015) W3	16	0	34	10	0.32 (0.21-0.46)	1.00 (0.72-1.00)

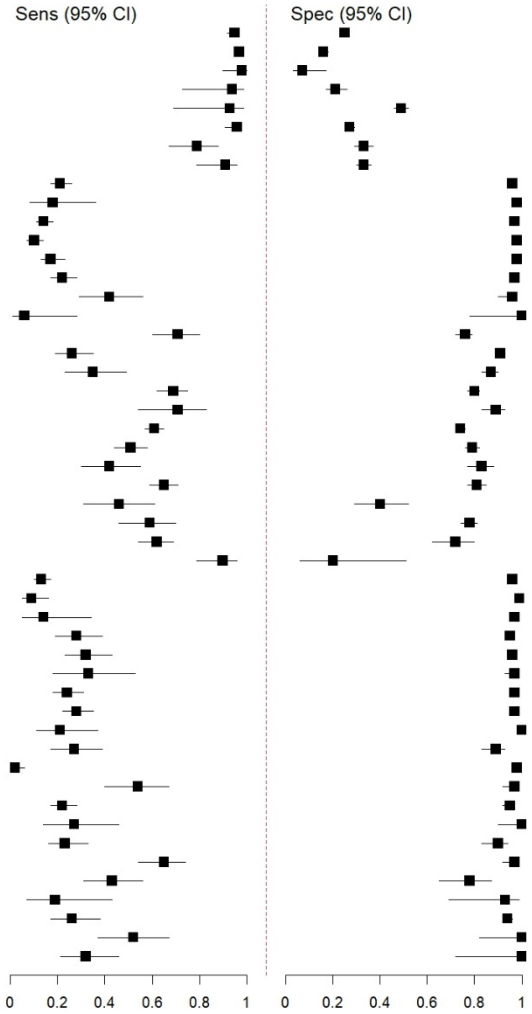

Figure c- PE diagnostic test assignment model

Appendix 11- The Summary ROC plots for the included primary studies

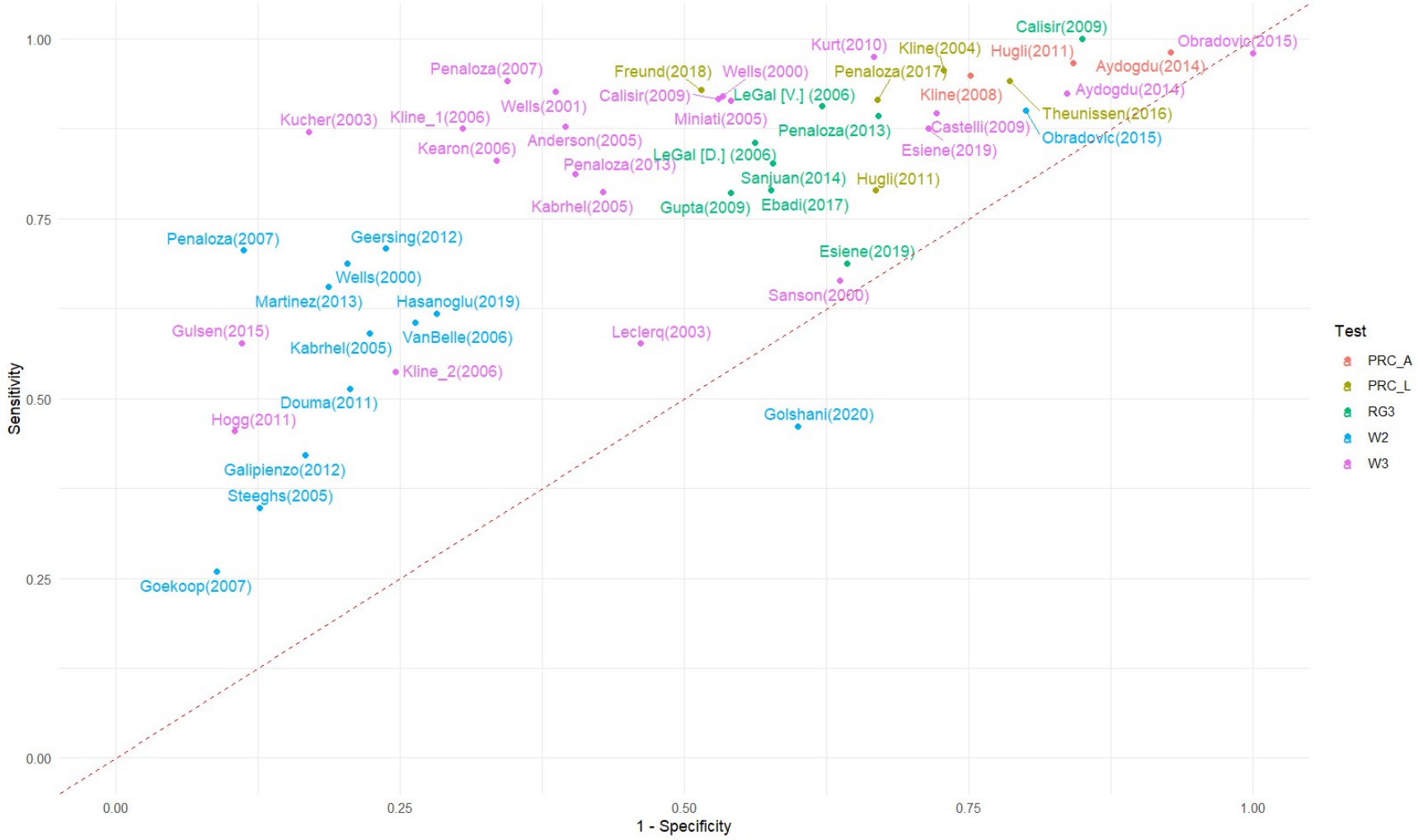

Figure d- The PE rule-out model

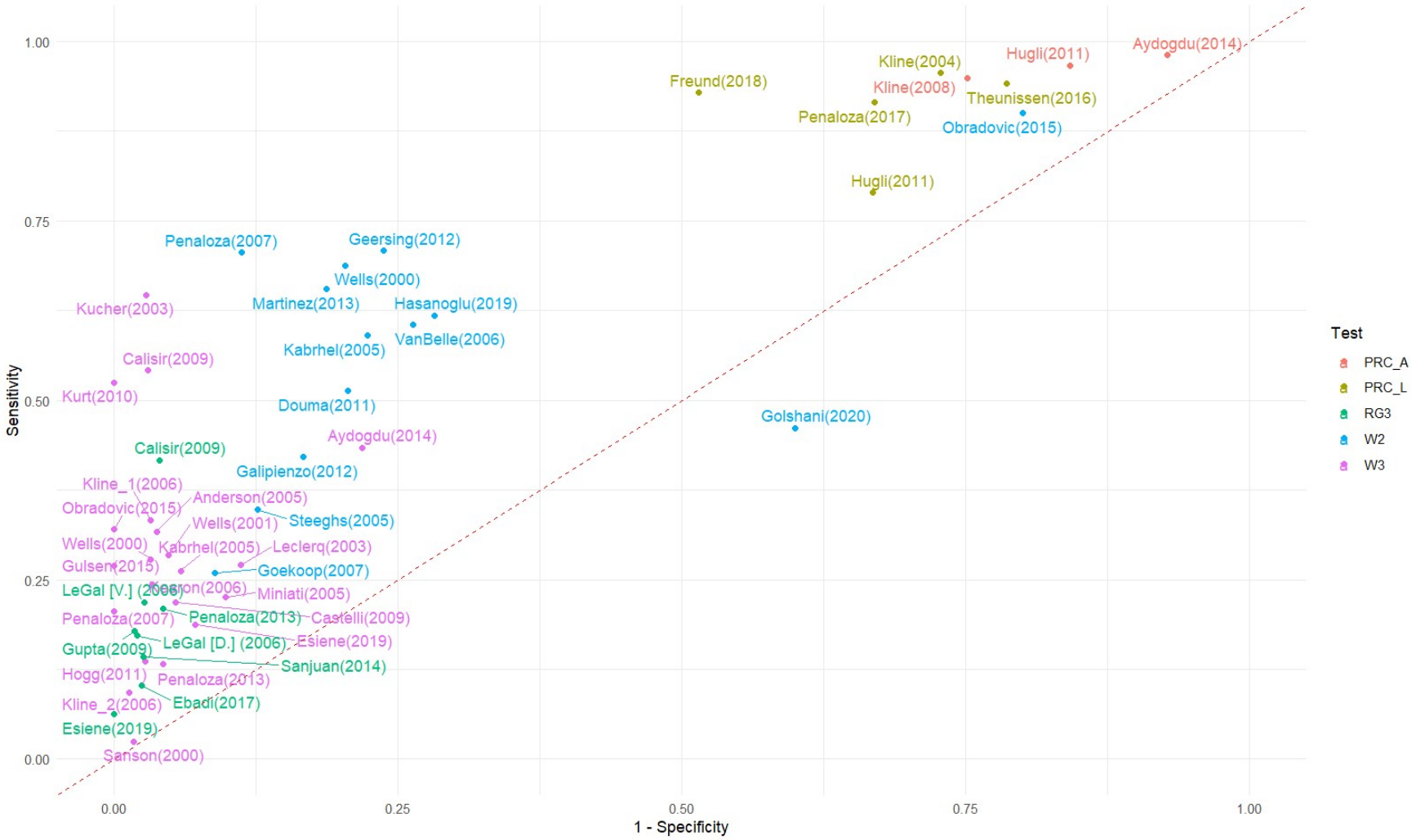

Figure e- The PE diagnostic test assignment model

Appendix 12- The network graph depicting the number of direct comparisons in the included primary studies

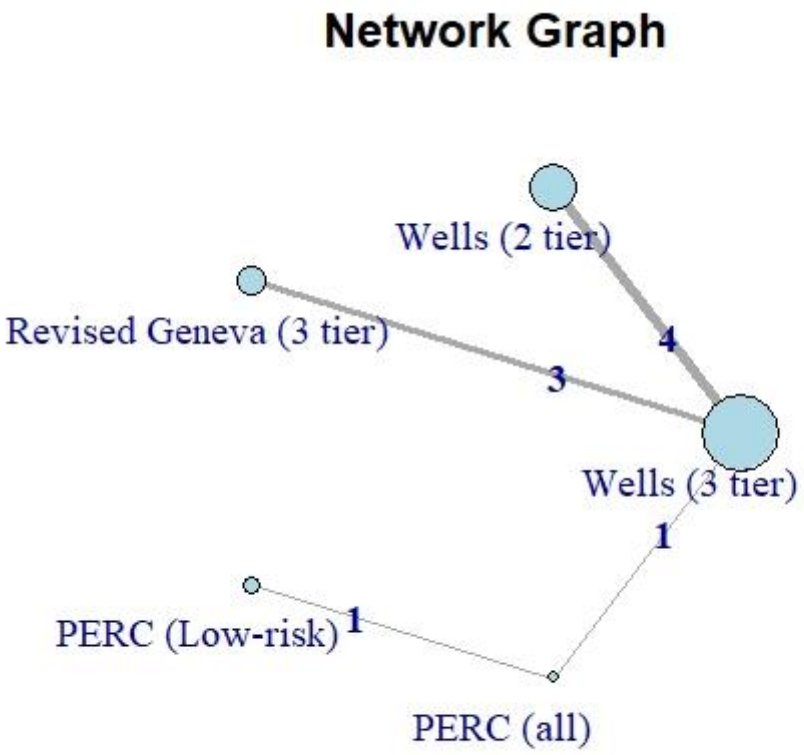

Figure g- Network graph of the direct interventions in the primary studies included in the meta-analysis; The size of the nodes and edges are proportional to the number of relevant records

Appendix 13- The traffic light and summary plots of the assessment of risk of bias and applicability using the QUADAS-2 tool

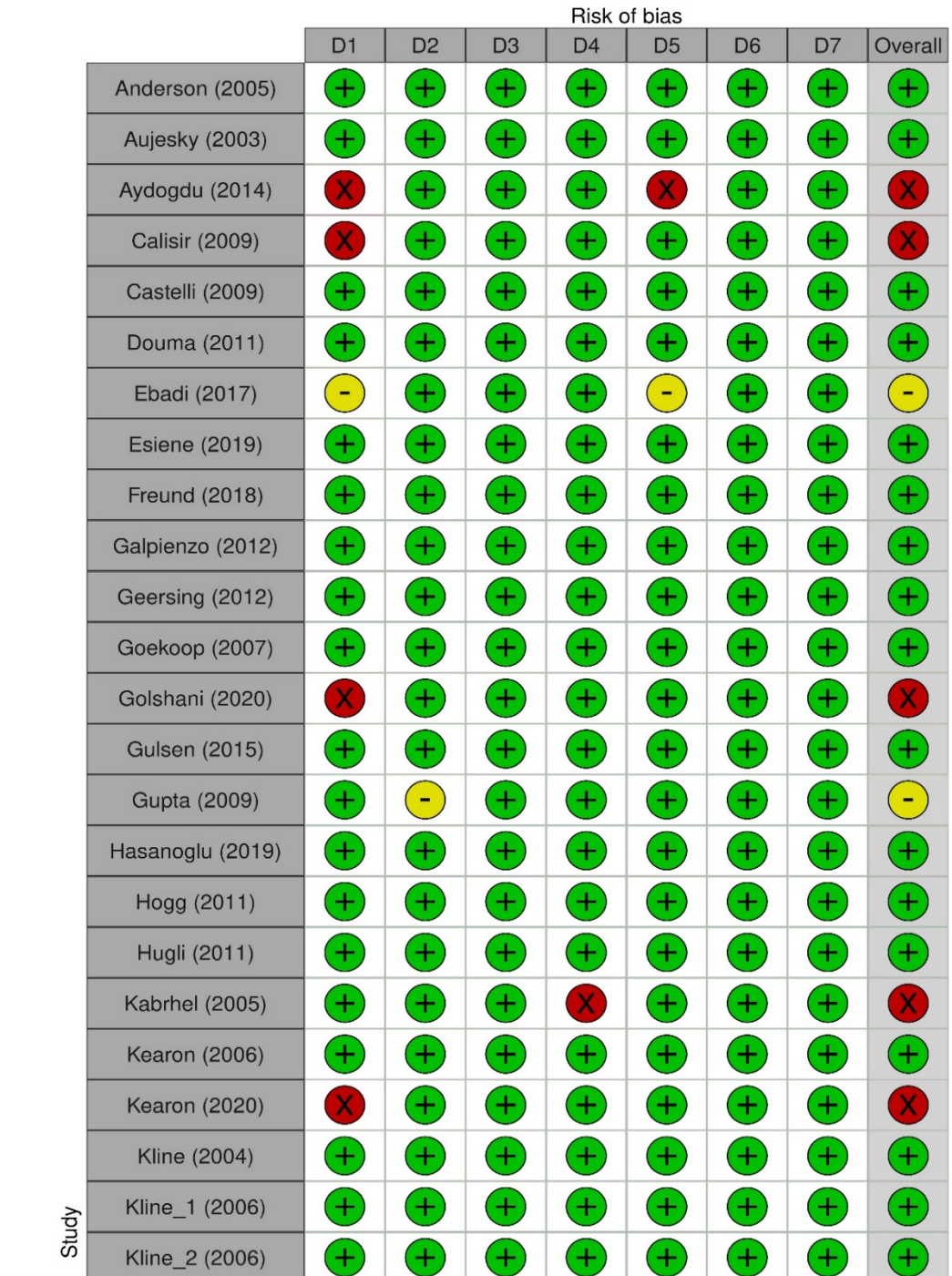

Figure h- Traffic light plot

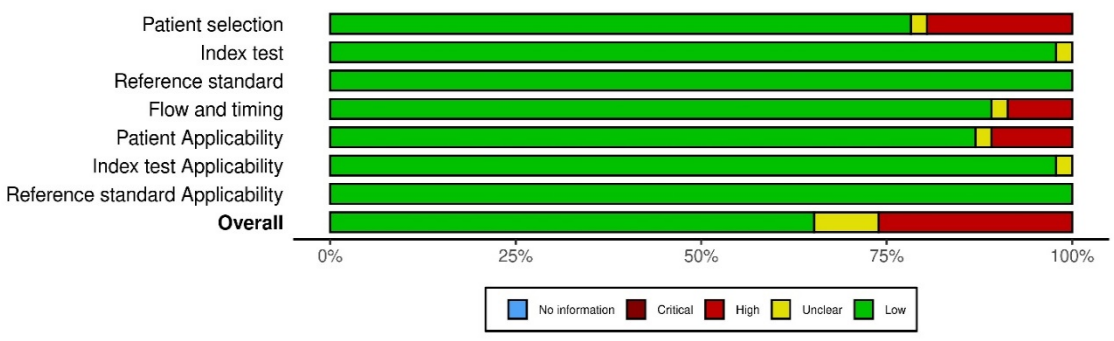

Figure i- Summary Plot

Appendix 14- The number of false negative and false positives in a sample of 100 patients in the range of reported prevalences of PE

PE-Unlikely model							
Prevalence	Test	Sensitivity	Specificity	False Negative	False Positive	Diff FN	Diff FP
0·05	1	0·88	0·34	0·60	62·70	Ref	Ref
0·05	2	0·85	0·38	0·75	58·90	0·15	-3·8
0·05	3	0·58	0·75	2·10	23·75	1·5	-38·95
0·05	4	0·82	0·55	0·90	42·75	0·3	-19·95
0·10	1	0·88	0·34	1·20	59·40	Ref	Ref
0·10	2	0·85	0·38	1·50	55·80	0·3	-3·6
0·10	3	0·58	0·75	4·20	22·50	3	-36·9
0·10	4	0·82	0·55	1·80	40·50	0·6	-18·9
0·15	1	0·88	0·34	1·80	56·10	Ref	Ref
0·15	2	0·85	0·38	2·25	52·70	0·45	-3·4
0·15	3	0·58	0·75	6·30	21·25	4·5	-34·85
0·15	4	0·82	0·55	2·70	38·25	0·9	-17·85
0·20	1	0·88	0·34	2·40	52·80	Ref	Ref
0·20	2	0·85	0·38	3·00	49·60	0·6	-3·2
0·20	3	0·58	0·75	8·40	20·00	6	-32·8
0·20	4	0·82	0·55	3·60	36·00	1·2	-16·8
0·25	1	0·88	0·34	3·00	49·50	Ref	Ref
0·25	2	0·85	0·38	3·75	46·50	0·75	-3
0·25	3	0·58	0·75	10·50	18·75	7·5	-30·75
0·25	4	0·82	0·55	4·50	33·75	1·5	-15·75
0·30	1	0·88	0·34	3·60	46·20	Ref	Ref
0·30	2	0·85	0·38	4·50	43·40	0·9	-2·8
0·30	3	0·58	0·75	12·60	17·50	9	-28·7
0·30	4	0·82	0·55	5·40	31·50	1·8	-14·7

Test 1: PERC (Low-risk), Test 2: Revised Geneva (3-tier), Test 3: Wells (2-tier), Test 4: Wells (3-tier);
Diff: The difference between each Test's FN or FP with that of Test 1 (PERC, low-risk)

PE-Likely model							
Prevalence	Test	Sensitivity	Specificity	False Negative	False Positive	Diff FN	Diff FP
0·05	1	0·88	0·33	0·6	63·65	Ref	Ref
0·05	2	0·20	0·97	4·0	2·85	3·4	-60·8
0·05	3	0·58	0·76	2·1	22·80	1·5	-40·85
0·05	4	0·28	0·95	3·6	4·75	3	-58·9
0·10	1	0·88	0·33	1·2	60·30	Ref	Ref
0·10	2	0·20	0·97	8·0	2·70	6·8	-57·6
0·10	3	0·58	0·76	4·2	21·60	3	-38·7
0·10	4	0·28	0·95	7·2	4·50	6	-55·8
0·15	1	0·88	0·33	1·8	56·95	Ref	Ref
0·15	2	0·20	0·97	12·0	2·55	10·2	-54·4
0·15	3	0·58	0·76	6·3	20·40	4·5	-36·55
0·15	4	0·28	0·95	10·8	4·25	9	-52·7
0·20	1	0·88	0·33	2·4	53·60	Ref	Ref
0·20	2	0·20	0·97	16·0	2·40	13·6	-51·2
0·20	3	0·58	0·76	8·4	19·20	6	-34·4
0·20	4	0·28	0·95	14·4	4·00	12	-49·6
0·25	1	0·88	0·33	3·0	50·25	Ref	Ref
0·25	2	0·20	0·97	20·0	2·25	17	-48
0·25	3	0·58	0·76	10·5	18·00	7·5	-32·25
0·25	4	0·28	0·95	18·0	3·75	15	-46·5
0·30	1	0·88	0·33	3·6	46·90	Ref	Ref
0·30	2	0·20	0·97	24·0	2·10	20·4	-44·8
0·30	3	0·58	0·76	12·6	16·80	9	-30·1
0·30	4	0·28	0·95	21·6	3·50	18	-43·4

Test 1: PERC (Low-risk), Test 2: Revised Geneva (3-tier), Test 3: Wells (2-tier), Test 4: Wells (3-tier);
Diff: The difference between each Test's FN or FP with that of Test 1 (PERC, low-risk)

Appendix 15- The outcomes of the baseline bivariate hierarchical model

Test[1]	Sensitivity (95% CI)	Specificity (95% CI)	LR- (95% CI)	LR+ (95% CI)	DOR (95% CI)
Model: PE-Unlikely model					
PERC (All)	0·96 (0·94-0·97)	0·17 (0·1-0·26)	0·25 (0·11-0·56)	1·15 (1·08-1·28)	4·66 (1·91-11·54)
PERC (Low-risk)	0·91 (0·83-0·96)	0·32 (0·25-0·41)	0·28 (0·11-0·68)	1·34 (1·27-1·39)	4·86 (1·86-12·84)
Revised Geneva (3 tier)	0·86 (0·81-0·9)	0·41 (0·4-0·42)	0·34 (0·23-0·49)	1·46 (1·5-1·39)	4·34 (3·09-6·13)
Wells (2 tier)	0·58 (0·48-0·68)	0·77 (0·68-0·84)	0·54 (0·38-0·76)	2·54 (2·14-3·01)	4·72 (2·83-7·87)
Wells (3 tier)	0·85 (0·79-0·9)	0·54 (0·43-0·65)	0·27 (0·15-0·49)	1·87 (1·59-2·27)	6·95 (3·22-15·09)
Model: PE-Likely model					
PERC (All)	0·96 (0·94-0·97)	0·17 (0·1-0·26)	0·25 (0·11-0·56)	1·15 (1·08-1·28)	4·66 (1·91-11·54)
PERC (Low-risk)	0·91 (0·83-0·96)	0·32 (0·25-0·41)	0·28 (0·11-0·68)	1·34 (1·27-1·39)	4·86 (1·86-12·84)
Revised Geneva (3 tier)	0·18 (0·13-0·24)	0·97 (0·97-0·98)	0·84 (0·78-0·89)	6·92 (8·03-5·74)	8·22 (8·98-7·39)
Wells (2 tier)	0·58 (0·48-0·68)	0·77 (0·68-0·84)	0·54 (0·38-0·76)	2·54 (2·14-3·01)	4·72 (2·83-7·87)
Wells (3 tier)	0·26 (0·19-0·33)	0·96 (0·95-0·97)	0·77 (0·69-0·85)	6·74 (6·17-7·11)	8·71 (7·22-10·37)

[1]As PERC and Wells (2-tier) scores are characterized by only two defined categories (low and high risk), their indices remain constant across models.

PE: Pulmonary embolism; LR: Likelihood ratio; CI: Confidence interval.

Appendix 16- Step by step approach in evaluating the sources of heterogeneity

Pretest clinical probability score	Index	Variance of Random effects				
		Step 0 (Baseline)	Step 1	Step 2	Step 3	Step 4
		Baseline model (50 records)[2]	Subgroup of studies conducted in the ED (32 records)	Subgroup of studies with >250 sample size (31 records)	Subgroup of studies with >500 sample size (27 records)	Subgroup of studies conducted in the ED with >500 sample size (18 records)
Wells (3-categories)	Sensitivity	0·95	0·67	0·7	0·45	0·49
	Specificity	1·04	0·95	0·65	0·31	0·36
Wells (2-categories)	Sensitivity	0·49	0·16	0·38	0·33	0[3]
	Specificity	0·62	0·62	0·16	0·15	0[3]
Revised Geneva (3-categories)	Sensitivity	0·24	0·09	0·08	0·08	0·08
	Specificity[1]	0	0	0	0	0
PERC (Low-risk)	Sensitivity	0·35	0·35	0·35	0·35	0·35
	Specificity	0·16	0·16	0·16	0·11	0·11
PERC (All)	Sensitivity[1]	0	0	0	0	0
	Specificity	0·19	0[3]	0·08	0·08	0[3]

1· As described in Appendix 5, the sensitivity index for these tests either had low levels of random-effect variance (<0·15) and was highly correlated with the random-effect variance of specificity (correlation effect size >0·95), so they were constrained to fixed effects.
2· As a number of primary studies evaluated more than one score, the 40 included studies in the meta-analysis corresponded to 50 records.
3· Due to the low level of random-effect variance, we constrained the index to a fixed effect in this subgroup analysis to ensure model fit.

Appendix 17- The outcomes of the bivariate hierarchical models, limited to primary studies with >500 sample size resulting in low-to-moderate levels of heterogeneity

Test[1]	Sensitivity (95% CI)	Specificity (95% CI)	LR- (95% CI)	LR+ (95% CI)	DOR (95% CI)
Model: PE-Unlikely model					
PERC (All)	0·96 (0·94-0·97)	0·2 (0·14-0·27)	0·21 (0·11-0·41)	1·2 (1·13-1·29)	5·6 (2·74-11·71)
PERC (Low-risk)	0·9 (0·8-0·96)	0·35 (0·28-0·43)	0·27 (0·1-0·7)	1·4 (1·33-1·41)	5·12 (1·9-13·86)
Revised Geneva (3 tier)	0·85 (0·81-0·88)	0·41 (0·4-0·43)	0·36 (0·27-0·47)	1·45 (1·48-1·41)	4·07 (3·18-5·24)
Wells (2 tier)	0·57 (0·46-0·68)	0·8 (0·75-0·85)	0·53 (0·38-0·71)	2·92 (2·76-3)	5·5 (3·86-7·88)
Wells (3 tier)	0·84 (0·76-0·9)	0·57 (0·48-0·66)	0·28 (0·16-0·5)	1·97 (1·71-2·27)	7·12 (3·44-14·66)
Model: PE-Likely model					
PERC (All)	0·96 (0·94-0·97)	0·2 (0·14-0·27)	0·21 (0·11-0·41)	1·2 (1·13-1·29)	5·6 (2·74-11·71)
PERC (Low-risk)	0·9 (0·8-0·96)	0·35 (0·28-0·43)	0·27 (0·1-0·7)	1·4 (1·33-1·41)	5·12 (1·9-13·86)
Revised Geneva (3 tier)	0·16 (0·13-0·2)	0·97 (0·97-0·98)	0·86 (0·81-0·9)	6·35 (5·7-6·83)	7·4 (7-7·63)
Wells (2 tier)	0·57 (0·46-0·68)	0·8 (0·75-0·85)	0·53 (0·38-0·71)	2·92 (2·76-3)	5·5 (3·86-7·88)
Wells (3 tier)	0·22 (0·16-0·28)	0·96 (0·95-0·97)	0·81 (0·74-0·88)	6 (5·71-6·04)	7·38 (6·49-8·16)

[1]As PERC and Wells (2-tier) scores are characterized by only two defined categories (low and high risk), their indices remain constant across models.

PE: Pulmonary embolism; LR: Likelihood ratio; CI: Confidence interval

Appendix 18- The outcomes of the bivariate hierarchical models, limited to primary studies in the ED setting with >500 sample size, resulting in low-to-moderate levels of heterogeneity

Test[1]	Sensitivity (95% CI)	Specificity (95% CI)	LR- (95% CI)	LR+ (95% CI)	DOR (95% CI)
Model: PE-Unlikely model					
PERC (All)	0·97 (0·94-0·98)	0·16 (0·14-0·18)	0·21 (0·11-0·41)	1·15 (1·14-1·15)	5·37 (2·75-10·81)
PERC (Low-risk)	0·9 (0·8-0·96)	0·35 (0·28-0·43)	0·27 (0·1-0·7)	1·4 (1·33-1·41)	5·12 (1·9-13·86)
Revised Geneva (3 tier)	0·85 (0·81-0·88)	0·41 (0·4-0·43)	0·36 (0·27-0·47)	1·45 (1·48-1·41)	4·07 (3·18-5·24)
Wells (2 tier)	0·59 (0·46-0·71)	0·78 (0·74-0·81)	0·53 (0·36-0·72)	2·65 (2·72-2·44)	5·01 (3·75-6·73)
Wells (3 tier)	0·82 (0·72-0·9)	0·57 (0·45-0·69)	0·31 (0·15-0·62)	1·93 (1·64-2·3)	6·33 (2·65-15·15)
Model: PE-Likely model					
PERC (All)	0·97 (0·94-0·98)	0·16 (0·14-0·18)	0·21 (0·11-0·41)	1·15 (1·14-1·15)	5·37 (2·75-10·81)
PERC (Low-risk)	0·9 (0·8-0·96)	0·35 (0·28-0·43)	0·27 (0·1-0·7)	1·4 (1·33-1·41)	5·12 (1·9-13·86)
Revised Geneva (3 tier)	0·16 (0·13-0·2)	0·97 (0·97-0·98)	0·86 (0·81-0·9)	6·35 (6·83-5·7)	7·4 (7·63-7)
Wells (2 tier)	0·59 (0·46-0·71)	0·78 (0·74-0·81)	0·53 (0·36-0·72)	2·65 (2·72-2·44)	5·01 (3·75-6·73)
Wells (3 tier)	0·2 (0·14-0·28)	0·96 (0·94-0·98)	0·83 (0·74-0·92)	5·21 (4·98-5·44)	6·25 (5·44-7·36)

[1]As PERC and Wells (2-tier) scores are characterized by only two defined categories (low and high risk), their indices remain constant across models.

Appendix 19- The outcomes of the bivariate hierarchical models with retrospective studies excluded

Test[1]	Sensitivity (95% CI)	Specificity (95% CI)	LR- (95% CI)	LR+ (95% CI)	DOR (95% CI)
Model: PE-Unlikely model					
PERC (All)	0.96 (0.94-0.97)	0.2 (0.14-0.27)	0.21 (0.11-0.41)	1.2 (1.13-1.29)	5.6 (2.74-11.71)
PERC (Low-risk)	0.9 (0.8-0.96)	0.35 (0.28-0.43)	0.27 (0.1-0.7)	1.4 (1.33-1.41)	5.12 (1.9-13.86)
Revised Geneva (3 tier)	0.87 (0.77-0.93)	0.42 (0.4-0.43)	0.32 (0.17-0.58)	1.49 (1.55-1.36)	4.62 (2.67-8.17)
Wells (2 tier)	0.58 (0.47-0.68)	0.76 (0.66-0.84)	0.56 (0.39-0.8)	2.45 (2.01-2.98)	4.4 (2.52-7.73)
Wells (3 tier)	0.84 (0.76-0.9)	0.59 (0.47-0.7)	0.27 (0.14-0.52)	2.07 (1.72-2.53)	7.79 (3.31-18.44)
Model: PE-Likely model					
PERC (All)	0.96 (0.94-0.97)	0.2 (0.14-0.27)	0.21 (0.11-0.41)	1.2 (1.13-1.29)	5.6 (2.74-11.71)
PERC (Low-risk)	0.9 (0.8-0.96)	0.35 (0.28-0.43)	0.27 (0.1-0.7)	1.4 (1.33-1.41)	5.12 (1.9-13.86)
Revised Geneva (3 tier)	0.18 (0.12-0.27)	0.98 (0.97-0.98)	0.84 (0.74-0.91)	7.91 (9.41-6.11)	9.45 (10.34-8.24)
Wells (2 tier)	0.58 (0.47-0.68)	0.76 (0.66-0.84)	0.56 (0.39-0.8)	2.45 (2.01-2.98)	4.4 (2.52-7.73)
Wells (3 tier)	0.26 (0.18-0.35)	0.97 (0.95-0.98)	0.77 (0.66-0.86)	7.82 (7.33-8.27)	10.19 (8.53-12.49)

[1]As PERC and Wells (2-tier) scores are characterized by only two defined categories (low and high risk), their indices remain constant across models.