The diagnostic management of suspected pulmonary embolism in special patient populations

Stals, M.A.M.

3

Safety and efficiency of diagnostic strategies for ruling out pulmonary embolism in clinically relevant patient subgroups: a systematic review and individual-patient data meta-analysis

Milou A.M. Stals, Toshihiko Takada, Noémie Kraaijpoel, Nick van Es, Harry R. Büller, D. Mark Courtney, Yonathan Freund, Javier Galipienzo, Grégoire Le Gal, Waleed Ghanima, Menno V. Huisman, Jeffrey A. Kline, Karel G.M. Moons, Sameer Parpia, Arnaud Perrier, Marc Righini, Helia Robert-Ebadi, Pierre-Marie Roy, Maarten van Smeden, Phil S. Wells, Kerstin de Wit, Geert-Jan Geersing, Frederikus A. Klok

Ann Intern Med. 2022 Feb;175(2):244-255

ABSTRACT

Background: How diagnostic strategies for suspected pulmonary embolism (PE) perform in relevant patient subgroups defined by sex, age, cancer, and previous venous thromboembolism (VTE) is unknown.

Purpose: To evaluate the safety and efficiency of the Wells and revised Geneva scores combined with fixed and adapted D-dimer thresholds, as well as the YEARS algorithm, for ruling out acute PE in these subgroups.

Data Sources: MEDLINE from 1 January 1995 until 1 January 2021.

Study Selection: 16 studies assessing at least 1 diagnostic strategy.

Data Extraction: Individual-patient data from 20.553 patients.

Data Synthesis: Safety was defined as the diagnostic failure rate (the predicted 3-month VTE incidence after exclusion of PE without imaging at baseline). Efficiency was defined as the proportion of individuals classified by the strategy as "PE considered excluded" without imaging tests. Across all strategies, efficiency was highest in patients younger than 40 years (47% to 68%) and lowest in patients aged 80 years or older (6.0% to 23%) or patients with cancer (9.6% to 26%). However, efficiency improved considerably in these subgroups when pretest probability–dependent D-dimer thresholds were applied. Predicted failure rates were highest for strategies with adapted D-dimer thresholds, with failure rates varying between 2% and 4% in the predefined patient subgroups.

Limitations: Between-study differences in scoring predictor items and D-dimer assays, as well as the presence of differential verification bias, in particular for classifying fatal events and subsegmental PE cases, all of which may have led to an overestimation of the predicted failure rates of adapted D-dimer thresholds.

Conclusion: Overall, all strategies showed acceptable safety, with pretest probability–dependent D-dimer thresholds having not only the highest efficiency but also the highest predicted failure rate. From an efficiency perspective, this individual patient data meta-analysis supports application of adapted D-dimer thresholds.

1. INTRODUCTION

Currently recommended diagnostic strategies for suspected acute pulmonary embolism (PE) consist of a standardized assessment of the clinical pretest probability using a validated clinical decision rule (CDR) and D-dimer testing.[1] The combination of a nonhigh clinical probability and a normal D-dimer test result safely rules out acute PE, allowing clinicians to refrain from performing imaging tests.[2,3] This is important to minimize exposure to potentially harmful ionizing radiation and contrast material, as well as to reduce health care costs and turnaround time in busy clinics.[1,4–7] With the recent introduction and validation of D-dimer thresholds dependent on age or clinical pretest probability, the proportion of patients requiring an imaging test has decreased from about 70% (when using the fixed D-dimer threshold of 500 µg/L) to 40% to 50%.[2,8–11]

Nevertheless, although the overall safety and efficiency of these strategies have been demonstrated in large management studies[8,10,12–15], it is also recognized that CDRs and D-dimer tests in general may be less safe and less efficient in specific patient subgroups, such as patients with renal insufficiency, patients with cancer, and elderly patients or inpatients.[2,16–18] Thus, the preferred diagnostic strategy may be different for certain subgroups. Yet, how different CDR/D-dimer test combinations perform in relevant patient subgroups is unknown, as individual studies were often too small to perform reliable subgroup analyses.

We set out to evaluate the safety and efficiency of the most widely used and recommended CDRs (the Wells rule and revised Geneva score in combination with available strategies for interpretation of the D-dimer test [fixed, age-adjusted, and pretest probability dependent]), as well as the YEARS algorithm, a strategy with D-dimer dependent on pretest probability, for frequently encountered and clinically relevant patient subgroups. To validate these 3 diagnostic scores in clinically relevant patient subgroups, we performed an international systematic review followed by a meta-analysis of individual patient data (IPDMA) from more than 20.000 patients with suspected PE.[19]

2. METHODS

This IPDMA followed the guidance of both the PRISMA-IPD (PRISMA for Individual Patient Data systematic reviews) and PRISMA-DTA (PRISMA for Diagnostic Test Accuracy) statements on systematic reviews including individual-patient data, and followed guidance from TRIPOD (Transparent reporting of a multivariable prediction model for individual

prognosis or diagnosis).[20–24] This IPDMA was preregistered at the PROSPERO database for systematic reviews (CRD42018089366), and a protocol was published.[19]

Data Sources and Searches

MEDLINE was searched from 1 January 1995 until 1 January 2021 to retrieve studies that had evaluated diagnostic strategies for PE (Appendix, available at Annals.org). Full-text articles were independently assessed for eligibility in duplicate by 2 pairs of authors (N.K. and G.J.G., and N.v.E. and F.A.K.). Discrepancies were resolved by discussion.

Study Selection

The process of study selection was described in detail in the published protocol.[19] In short, eligible studies were those that had a prospective follow-up or cross-sectional study design, included patients with clinically suspected PE, and assessed variables to calculate at least 1 of the predefined CDRs of interest. Furthermore, the reference standard had to be imaging or clinical follow-up in those in whom PE was ruled out without imaging and who thus did not receive anticoagulant treatment. In addition, we excluded studies with qualitative D-dimer measurements only and studies including only patients with low clinical pretest probability.

Data Extraction and Quality Assessment

Principal investigators from the eligible studies were asked to provide deidentified individual-patient data (IPD) (**Appendix Figure 1**, available at Annals.org). Patient-level data collected at baseline included information on demographic characteristics, risk factors for venous thromboembolism (VTE), comorbidity, items of the diagnostic strategies of interest, D-dimer levels, and results of imaging tests. Information collected during follow-up included information on occurrence of VTE, anticoagulant therapy for reasons other than VTE, mortality, and loss to follow-up (see the Supplement, available at Annals.org). Each diagnostic study from which we retrieved IPD was assessed for potential sources of bias using the QUADAS2 (Quality Assessment of Diagnostic Accuracy Studies 2) tool (see **Appendix Figure 2**, available at Annals.org).[25] This assessment was performed independently by 3 pairs of authors (G.J.G. and T.T., N.v.E. and N.K., and F.A.K. and M.A.M.S.) who were not involved in the original included studies. Disagreements were resolved by discussion within each pair and between pairs.

Data Synthesis and Analysis

The main analysis focused on the predicted diagnostic performance of various diagnostic strategies for ruling out PE across different patient subgroups. Diagnostic strategies under evaluation were the Wells rule and revised Geneva score (both combined with D-dimer testing), and the YEARS algorithm (**Appendix Figures 3 and 4**, available at An-

nals.org). The YEARS algorithm is a strategy with a D-dimer threshold that is dependent on clinical pretest probability (CPTP) assessment (that is, it applies a higher D-dimer threshold in patients with a low CPTP).[10] The Wells rule and the revised Geneva score incorporate a fixed D-dimer threshold of 500 µg/L, an age-adjusted D-dimer threshold (age x 10 µg/L in patients aged >50 years), or a D-dimer threshold dependent on CPTP. The CPTP-dependent D-dimer threshold has not been prospectively validated for the revised Geneva score before, but we applied the same D-dimer thresholds as used in the PEGeD (The Pulmonary Embolism Graduated D-dimer) study (Wells rule with D-dimer threshold dependent on CPTP).[11]

The main outcome measures were the predicted safety and efficiency of each diagnostic strategy. Safety was defined as the failure rate, which is the proportion of patients with confirmed VTE at baseline or during follow-up divided by the total number of patients in whom PE was considered excluded at baseline based on CDR and D-dimer testing alone (as a measure of missed VTE events at baseline). Traditionally, the generally accepted safety threshold ranges between 2% and 3%, with recent data suggesting that a safety threshold dependent on PE prevalence at baseline should be used.[26] The efficiency of the diagnostic strategy was defined as the number of patients in whom PE was considered ruled out based on CDR and D-dimer alone among all included patients.

Probabilities of safety and efficiency were calculated overall and in clinically relevant patient subgroups: male versus female patients, age as a continuous variable, active cancer (as defined in the original studies), and history of VTE. Analyses in the predefined subgroups by delayed presentation, obesity, known heart failure, and known chronic obstructive pulmonary disease could not be performed because numbers for these subgroups were too small, definitions used in the original studies were too heterogeneous, or information could not be retrieved in most of the included studies.

Statistical Analysis

Multilevel logistic regression models were used to account for the clustering of patients within studies by including a random intercept. For the analyses of safety, a univariable logistic regression model was constructed with the presence or absence of VTE as the outcome and classification by each rule as a categorical covariate. By using this model, the failure rate of each model corresponds to the predicted probability of VTE in patients categorized by the model as “PE considered excluded.’’

This safety measure is frequently applied in the field of diagnostic studies in suspected PE and ideally should have a point estimate dependent on PE prevalence at baseline.[26] For the analyses of efficiency, a model with the classification by each rule as the out-

come and an intercept as a sole covariate was used. As such, efficiency was quantified as a predicted probability of being categorized as "PE considered excluded" by each rule.

Next, the predicted diagnostic performance of each decision rule in each subgroup was evaluated. To this end, each subgroup variable was added as a covariate in the multilevel logistic regression models described earlier. For the models of safety, an interaction term between each subgroup variable and the judgment based on each prediction rule was added.

For all outcome measures, 95% CIs and 95% prediction intervals were estimated by using the Gauss–Hermite quadrature approximation with 10 quadrature points. The prediction intervals illustrate the performance that can be expected when the diagnostic strategy is applied in a new population, taking between-study heterogeneity into account. As another measure for between-study heterogeneity, a random effect for an intercept in each multilevel logistic regression model (tau [t]) was used. Furthermore, the range of failure rates and efficiency of each diagnostic strategy in each subgroup across included studies was visualized using forest plots with the I^2statistic.[27]

Missing Data

In the data set, variables were either partially missing (that is, missing in a certain proportion of patients within a study) or systematically missing (that is, completely missing in certain studies). In accordance with statistical recommendations[28,29], those missing values were imputed using 1-stage, multilevel chained equations with all items included in the diagnostic strategies and the outcome. Ten imputation data sets were created, and the results of the analyses done separately in each set were combined using the Rubin rule.[30]

Sensitivity Analysis

Most studies included in this IPDMA used both imaging and clinical follow-up as the reference standard. However, VTE detected during follow-up could be a new event (that is, absent at baseline and thus unrelated to the index presentation), which is especially likely in high-risk patients. Such differential verification may lead to a false increase in failure rate. Thus, to evaluate the impact of this differential verification on safety, we performed a sensitivity analysis in which only VTE events diagnosed at baseline (based on imaging) were used as the outcome.

All analyses were performed using R, version 3.6.3 (R Foundation for Statistical Computing;www.R-project.org), particularly the lme4 package.

Role of the Funding Source

This study was funded by the Dutch Research Council. The steering committee, consisting of the authors, had final responsibility for the study design, oversight, and data verification and analyses. The sponsor was not involved in the study. All members of the steering committee contributed to the interpretation of the results, approved the final version of the manuscript, and vouch for the accuracy and completeness of the data reported. The final decision to submit the manuscript was made by the corresponding author on behalf of all coauthors.

3. RESULTS

Study Selection and Included Patients

The literature search retrieved 3.733 studies, of which 328 full texts were assessed for eligibility. Forty studies fulfilled the predefined eligibility criteria, and corresponding authors from these publications were invited to provide original IPD. Seventeen studies were eventually excluded after the original data files were scrutinized. In the end, 23 studies were included, with a total of 35.248 unique patients (**Appendix Figure 1**).

After exclusion of studies with qualitative D-dimer measurements only[31–36] and studies including only patients with low CPTP[35–37], 16 studies were included in the current analysis, involving a total of 20.553 patients. Of note, the inpatients from the study by Kline and colleagues[38] and the study patients with nonlow CPTP from the PERCEPIC (Pulmonary Embolism Rule-Out Criteria Rule in Patients With Low Implicit Clinical Probability) study[39] were also included in the final analysis. Characteristics and outcomes of the 16 included studies are summarized in **Table 1**.[8,10–14,38–47] Characteristics of the complete study group are provided in **Table 2**, and proportions of missing values in each study are shown in **Appendix Table 1** (available at Annals.org).

Heterogeneity in subgroup definitions between studies was low in general. Although the definitions of immobilization or surgery, clinical signs of deep venous thrombosis, and PE as the most likely diagnosis followed those as per the Wells rule in most studies, the original Geneva studies collected these variables based on Geneva scoring items. Previous VTE and active cancer followed the per-study definitions. In all studies except 2[38,42], imaging and anticoagulant therapy were withheld in patients with a low clinical probability and a negative D-dimer test result as per the decision rule and D-dimer threshold in that specific study. These patients were followed prospectively for 3 months by telephone contact or a scheduled outpatient visit, except for patients from the study by Kline and colleagues[38], who were followed for only 30 days.

Table 1. Characteristics of included studies

Study; Year (ref)	*Primary study goal*	*Study period*	*D-dimer assay*	*Diagnostic imaging test*	*Outcome adjudication*	*Number of patients provided by the authors; N*	*Patients with VTE at baseline or during FU; N (%)*
Kline et al, 2012 [38] -	Evaluate the Wells score and Revised Geneva score in conjunction with doubling the D-dimer threshold for PE unlikely patients (threshold: 1000 µg/L) in outpatients and inpatients.	January 2007 – April 2008	VIDAS (bioMérieux)	CTPA	No	333	77 (23)
Douma et al, 2011 [13] ~	To evaluate 4 CDRs* combined with D-dimer testing (threshold: 500 µg/L) in outpatients and inpatients.	July 2008 – November 2009	VIDAS / Tinaquant / STA Liatest / Innovance	CTPA	Yes	807	192 (24)^
Goekoop et al, 2007 [40]	Evaluate the dichotomized Wells score combined with D-dimer testing (threshold: 500 µg/L) in outpatients.	March 2002 – March 2004	VIDAS (bioMérieux)	CTPA or V/Q	No	876	110 (13)^
Righini et al, 2014 [8]	Evaluate the Wells or Revised Geneva score combined with D-dimer testing (threshold: 500 µg/L in patients aged <50 y and age-adjusted in patients aged ≥ 50 y) in outpatients.	January 2010 – February 2013	VIDAS / Tinaquant / Cobas h 232 / SA Liatest / HS-500 / Innovance	CTPA	Yes	3,324	639 (19)
Schouten et al, 2014 [41]	Evaluate the Wells score combined with D-dimer testing (qualitative assay) in older (≥60 y old) outpatients.	July 2007-April 2013	Simplify	CTPA or V/Q	Partly (only deaths were adjudicated)	294	83 (28)
Christopher Study Investigators, 2006 [12]	Evaluate the dichotomized Wells score combined with D-dimer testing (threshold: 500 µg/L) in outpatients and inpatients.	November 2002 – September 2004	VIDAS (bioMérieux) / Tinaquant (Roche)	CTPA	Yes	3,296	699 (21)
Van der Hulle et al, 2017 [10]	Evaluate the YEARS score combined with D-dimer testing (threshold 1000 µg/L if 0 YEARS items and 500 µg/L if 1-3 YEARS items) in outpatients and inpatients.	October 2013 – July 2015	VIDAS / Tinaquant / STA Liatest / Innovance	CTPA	Yes	3,448	473 (14)
Mos et al, 2014 [14]	Evaluate the dichotomized Wells score combined with D-dimer testing (threshold: 500 µg/L) in outpatients and inpatients with suspected recurrent PE.	November 2002 – November 2009	Tinaqaunt / VIDAS / STA Liatest / Innovance	CTPA	Yes	279	114 (41)
Wicki et al, 2001 [42]	To develop a clinical score for assessing clinical pre-test probability in outpatients. (= Geneva derivation study)	October 1992 – October 1997	Asserachrom D-Di enzyme / VIDAS (bioMérieux)	Lungscan	No	1,089	296 (27)

Table 1. Characteristics of included studies (*continued*)

Study; Year (ref)	*Primary study goal*	*Study period*	*D-dimer assay*	*Diagnostic imaging test*	*Outcome adjudication*	*Number of patients provided by the authors; N*	*Patients with VTE at baseline or during FU; N (%)*
Perrier et al, 2004 [43]	Evaluate the Geneva score combined with D-dimer testing (threshold: 500 µg/L) and CUS in outpatients.	October 2000 – June 2002	VIDAS (bioMérieux)	CTPA	Yes	965	229 (24)
Perrier et al, 2005 [44]	Evaluate the Geneva score combined with D-dimer testing (threshold: 500 µg/L) in outpatients.	August 2002 – November 2003	VIDAS (bioMérieux)	Multidetector-row CT and CUS	Partly (only deaths were adjudicated)	1,692	361 (21)
Righini et al, 2008 [45]	Evaluate the Revised Geneva score combined with 1. D-dimer testing (threshold: 500 µg/L) alone or 2. D-dimer testing followed by CUS if D-dimer above threshold or high clinical probability, in outpatients.	January 2005 – August 2006	VIDAS (bioMérieux)	CTPA	Yes	755	197 (26)
Kearon et al, 2019 [11]	Evaluate the Wells score combined with D-dimer testing (threshold: 1000 µg/L if low clinical probability and 500 µg/L if moderate clinical probability) in outpatients.**	December 2015 – May 2018	STA Liatest / HS 500 / Innovance / Triage / Other	CTPA	Yes	2,017	150 (7.4)
Galipienzo et al, 2012 [46]	Evaluate the dichotomized Wells score combined with D-dimer testing (threshold: 500 µg/L) in outpatients.	May 2007 – December 2008	VIDAS (bioMérieux)	CTPA	No	240	63 (26)
Ghanima et al, 2005 [47]	Evaluate an algorithm based on the Hyers criteria combined with D-dimer testing (threshold: 400 µg/L) in outpatients.	February 2002 – December 2003	STA Liatest	Multi-slice spiral CT (MSCT)	Partly (only deaths were adjudicated)	432	95 (22)
Penaloza et al, 2017 [39] (only patients with non-low clinical probability were included)	Evaluate the PERC rule in patients with low implicit clinical probability.~~	May 2015 – April 2016	NA	CTPA or V/Q	Yes	705	153 (22)

PE: pulmonary embolism; CDR: clinical decision rule; y: years; CUS: compression ultrasonography; PERC: pulmonary embolism rule-out criteria; NA: not available; CTPA: computed tomography pulmonary angiography; V/Q: ventilation-perfusion scan; N: number

- This was not a management study. Patients were selected based on performed CTPA, and retrospectively data on CDR and D-dimer was assessed. Therefore, no patients were managed without CTPA.

~ Patients were only ruled out from having PE if all CDRs implicated PE unlikely

* Wells rule and Revised Geneva score (both original and simplified version)

^ only patients in whom PE was excluded at baseline based on a low clinical probability and low D-dimer (so without imaging) were followed over time

** 1 inpatient was enrolled

~~ Patients were excluded if they had already been hospitalized for more than 2 days

Table 2. Characteristics of complete study group

	N	*Age (median [IQR])*	*Female sex (%)*	*Inpatients (%)*	*Previous VTE (%)*	*Tachycardia (%)*	*ImmoSurg (%)*	*Active cancer (%)*	*Hemoptysis (%)*	*Clinical signs of deep-vein thrombosis (%)*	*PE most likely diagnosis (%)*	*Ddimer Quant (median [IQR])*	*VTE diagnosis (%)*
Overall	20,553	58 [43, 71]	12,161 (59)	1606 (7.8)	2941 (14)	4383 (21)	3143 (15)	2218 (11)	942 (4.6)	1551 (7.5)	9503 (46)	800 [360, 1780]	3932 (19)
Kline et al, 2012 [38]	334	57 [47, 66]	200 (60)	334 (100)	50 (15)	60 (18)	124 (37)	73 (22)	7 (2.1)	31 (9.3)	132 (40)	1315 [610, 2758]	77 (23)
Douma et al, 2011 [13]	807	54 [40, 67]	487 (60)	163 (20)	39 (4.8)	184 (23)	176 (22)	121 (15)	40 (5.0)	47 (5.8)	456 (57)	1055 [500, 2415]	192 (24)
Goekoop et al, 2007 [40]	876	50 [38, 65]	549 (63)	0 (0)	84 (9.6)	166 (19)	50 (5.7)	17 (2.0)	31 (3.5)	11 (1.3)	415 (47)	430 [244, 1000]	110 (13)
Righini et al, 2014 [8]	3,324	63 [53, 73]	1887 (57)	0 (0)	466 (14)	722 (22)	392 (12)	429 (13)	134 (4.0)	239 (7.2)	1557 (47)	892 [473, 2101]	639 (19)
Schouten et al, 2014 [41]	294	76 [67, 84]	195 (66)	0 (0)	52 (18)	64 (22)	88 (30)	28 (9.6)	8 (2.7)	35 (12)	162 (55)	1731 [820, 3840]	83 (28)
Christopher Study Investigators, 2006 [12]	3,296	52 [39, 68]	1897 (58)	605 (18)	427 (13)	690 (21)	638 (19)	375 (11)	175 (5.3)	189 (5.7)	2027 (62)	801 [340, 2075]	699 (21)
Van der Hulle et al, 2017 [10]	3,448	54 [40, 67]	2142 (62)	468 (14)	360 (10)	693 (20)	408 (12)	336 (9.8)	137 (4.0)	112 (3.2)	1619 (47)	670 [337, 1508]	473 (14)
Mos et al, 2014 [14]	279	54 [42, 68]	164 (59)	36 (13)	279 (100)	59 (21)	41 (15)	52 (19)	18 (6.4)	32 (11)	169 (61)	982 [444, 2540]	114 (41)
Wicki et al, 2001 [42]	1,089	62 [46, 76]	597 (55)	0 (0)	202 (19)	236 (22)	316 (29)	138 (13)	63 (5.8)	147 (14)	447 (41)	860 [387, 2593]	296 (27)
Perrier et al, 2004 [43]	965	63 [45, 77]	562 (58)	0 (0)	167 (17)	186 (19)	202 (21)	90 (9.3)	43 (4.5)	174 (18)	288 (30)	950 [411, 1000]	229 (24)
Perrier et al, 2005 [44]	1,692	61 [45, 75]	923 (55)	0 (0)	300 (18)	369 (22)	258 (15)	127 (7.5)	83 (4.9)	153 (9.0)	770 (46)	890 [370, 1769]	361 (21)
Righini et al, 2008 [45]	755	63 [45, 76]	453 (60)	0 (0)	142 (19)	176 (23)	124 (16)	75 (9.9)	37 (4.9)	72 (9.6)	464 (62)	1001 [379, 1001]	197 (26)

Table 2. Characteristics of complete study group (*continued*)

	N	*Age (median [IQR])*	*Female sex (%)*	*Inpatients (%)*	*Previous VTE (%)*	*Tachycardia (%)*	*ImmoSurg (%)*	*Active cancer (%)*	*Hemoptysis (%)*	*Clinical signs of deep-vein thrombosis (%)*	*PE most likely diagnosis (%)*	*Ddimer Quant (median [IQR])*	*VTE diagnosis (%)*
Kearon et al, 2019 [11]	2,017	53 [38, 66]	1335 (66)	1 (0)	164 (8.1)	398 (20)	149 (7.4)	187 (9.3)	93 (4.6)	138 (6.8)	423 (21)	490 [270, 1160]	150 (7.4)
Galipienzo et al, 2012 [46]	240	67 [54, 78]	122 (51)	0 (0)	38 (16)	56 (23)	51 (21)	36 (15)	27 (11)	24 (10)	53 (22)	1071 [479, 2429]	63 (26)
Ghanima et al, 2005 [47]	432	58 [43, 73]	231 (54)	0 (0)	44 (10)	79 (18)	38 (8.8)	32 (7.3)	15 (3.4)	56 (13)	198 (46)	1000 [400, 2725]	95 (22)
Penaloza et al, 2017 [39]	705	61 [46, 76]	419 (59)	0 (0)	127 (18)	246 (35)	88 (13)	102 (15)	31 (4.4)	91 (13)	323 (46)	1001 [413, 2170]	153 (22)

n= number; IQR: interquartile range; VTE: venous thromboembolism; Immosurg: immobility or surgery in the last 4 weeks (Wells item); PE: pulmonary embolism; Ddimer Quant: quantitative measurement of D-dimer

Main Outcomes

Table 3 shows the predicted overall safety of the different strategies (defined as the failure rate). The predicted failure rate among patients in whom imaging was withheld was 0.36% (95% CI, 0.20% to 0.63%) for the Wells rule and 0.58% (CI, 0.37% to 0.90%) for the revised Geneva score when using the fixed D-dimer threshold of 500 µg/L. When the age-adjusted D-dimer threshold was used, the predicted failure rate was 0.76% (CI, 0.52% to 1.1%) for the Wells rule and 1.1% (CI, 0.80% to 1.5%) for the revised Geneva score. For strategies applying the D-dimer threshold dependent on pretest probability, the predicted failure rate was 1.8% (CI, 1.4% to 2.4%) for the YEARS algorithm, 2.8% (CI, 2.3% to 3.5%) for the Wells rule (PEGeD Wells), and 2.8% (CI, 2.3% to 3.5%) for the revised Geneva score (PEGeD Geneva). **Table 4** shows the predicted overall efficiency of the different diagnostic strategies. The predicted overall efficiency was highest for strategies applying the D-dimer threshold dependent on pretest probability: The Wells rule (PEGeD Wells, 47% [CI, 42% to 52%]) had the highest efficiency, followed by the revised Geneva score (PEGeD Geneva, 44% [CI, 39% to 50%]) and the YEARS algorithm (41% [CI, 36% to 47%]). The least efficient strategies were the Wells rule or the revised Geneva score combined with a fixed D-dimer threshold of 500 µg/L (26% [CI, 22% to 31%] and 30% [CI, 26% to 36%], respectively). The predicted efficiencies of the Wells rule and the revised Geneva score when using the age-adjusted D-dimer threshold were 32% (CI, 27% to 37%) and 37% (CI, 32% to 41%), respectively.

The predicted failure rate of the different diagnostic strategies across different subgroups of patients is presented in **Table 3**, **Figure 1**, and **Appendix Figure 5** (available at Annals.org). The predicted failure rate, as well as the uncertainty around this point estimate, was highest for the diagnostic strategies that used a D-dimer threshold dependent on pretest probability, especially in patients aged 80 years or older, patients with active cancer, and patients with a history of VTE. Predicted failure rates varied between 3% and 4% in these subgroups. The predicted efficiency of all strategies, presented in **Table 4**, **Figure 2**, and **Appendix Figure 5**, was highest in patients younger than 40 years, ranging from 47% to 54% when the fixed or age-adjusted D-dimer threshold was used and from 64% to 68% when the D-dimer threshold dependent on pretest probability was used.

Across all strategies, predicted efficiency was lowest in patients aged 80 years or older and in those with active cancer, although efficiency increased considerably when the age-adjusted D-dimer threshold or the D-dimer threshold dependent on pretest probability was applied, from 6% to 7% (with the fixed D-dimer threshold) to 17% to 23% in patients aged 80 years or older and from 10% to 12% (with the fixed D-dimer threshold) to 15% to 26% in patients with active cancer.

Table 3. Failure rate* of the CDRs and D-dimer testing in excluding PE, overall and in clinically relevant patient subgroups

Variable	*Overall*	*Sex*		*Active cancer*		*History VTE*	
		Male	*Female*	*No*	*Yes*	*No*	*Yes*
Wells score (fixed D-dimer), %	0.36	0.34	0.36	0.36	NE~	0.33	0.48
95% CI	0.20-0.63	0.13-0.90	0.19-0.69	0.21-0.63	NE	0.19-0.58	0.09-2.6
95% PI	0.14-0.94	0.09-1.2	0.13-1.0	0.14-0.94	NE	0.14-0.77	0.07-3.4
N	20,553	8,391	12,162	18,334	2,219	17,611	2,942
Tau (τ)	0.35	0.34		0.35		0.28	
Wells score (age-adjusted D-dimer), %	0.76	0.57	0.89	0.74	1.1	0.70	1.0
95% CI	0.52-1.1	0.28-1.1	0.59-1.3	0.50-1.1	0.33-3.6	0.49-1.0	0.30-3.5
95% PI	0.33-1.7	0.20-1.6	0.39-2.0	0.32-1.7	0.25-4.7	0.35-1.4	0.23-4.3
N	20,553	8,391	12,162	18,334	2,219	17,611	2,942
Tau (τ)	0.33	0.33		0.33		0.27	
Revised Geneva score (fixed D-dimer), %	0.58	0.60	0.56	0.55	1.3	0.48	1.2
95% CI	0.37-0.90	0.33-1.1	0.33-0.96	0.35-0.85	0.37-4.8	0.30-0.74	0.51-2.6
95% PI	0.22-1.5	0.21-1.7	0.21-1.5	0.21-1.4	0.26-6.6	0.21-1.1	0.39-3.3
N	20,553	8,391	12,162	18,334	2,219	17,611	2,942
Tau (τ)	0.39	0.39		0.39		0.29	
Revised Geneva score (age-adjusted D-dimer), %	1.1	1.0	1.2	1.0	2.5	0.87	2.5
95% CI	0.80-1.5	0.65-1.6	0.81-1.7	0.74-1.4	1.1-5.6	0.63-1.2	1.5-4.3
95% PI	0.46-2.6	0.40-2.6	0.48-2.8	0.43-2.5	0.73-8.0	0.43-1.7	1.1-5.7
N	20,553	8,391	12,162	18,334	2,219	17,611	2,942
Tau (τ)	0.38	0.37		0.38		0.28	
YEARS algorithm, %	1.8	2.2	1.6	1.7	3.4	1.5	3.5
95% CI	1.4-2.4	1.5-3.0	1.2-2.2	1.3-2.3	1.9-6.0	1.2-1.9	2.3-5.2
95% PI	0.78-4.2	0.90-5.1	0.67-3.8	0.74-4.0	1.2-9.0	0.77-2.9	1.7-7.2
N	20,553	8,391	12,162	18,334	2,219	17,611	2,942
Tau (τ)	0.38	0.38		0.38		0.28	
PEGeD algorithm (Wells), %	2.8	3.4	2.4	2.7	3.9	2.6	3.4
95% CI	2.3-3.5	2.7-4.4	1.9-3.0	2.2-3.4	2.4-6.4	2.2-3.2	2.3-5.2
95% PI	1.3-5.8	1.6-7.2	1.1-5.0	1.3-5.7	1.6-9.2	1.4-4.8	1.7-7.0
N	20,553	8,391	12,162	18,334	2,219	17,611	2,942
Tau (τ)	0.34	0.34		0.34		0.27	
PEGeD algorithm (Geneva), %	2.8	3.4	2.4	2.8	3.5	2.7	2.8
95% CI	2.3-3.5	2.7-4.4	1.9-3.1	2.3-3.4	1.8-6.6	2.2-3.3	1.7-4.6
95% PI	1.3-5.8	1.6-7.1	1.1-5.0	1.3-5.8	1.3-9.1	1.5-4.9	1.3-6.1
N	20,553	8,391	12,162	18,334	2,219	17,611	2,942
Tau (τ)	0.34	0.33		0.34		0.27	

CDR: clinical decision rule; PE: pulmonary embolism; CI: confidence interval; PI = prediction interval; N: number of patients included in the analysis; VTE = venous thromboembolism; NE: not estimable* Defined as the predicted 3-month probability of VTE in patients with a low score on the CDR combined with a negative D-dimer test result. ~ No failures were observed in patients with a low score on the CDR combined with a negative D-dimer test result.

Table 4. Efficiency* of the CDRs and D-dimer testing in excluding PE, overall and in clinically relevant patient subgroups

Variable	*Overall*	*Sex*		*Active cancer*		*History VTE*	
		Male	*Female*	*No*	*Yes*	*No*	*Yes*
Wells score (fixed D-dimer), %	26	26	27	28	9.6	30	12
95% CI	22-31	22-31	22-32	24-34	7.4-12	25-35	9.5-15
95% PI	11-51	11-51	11-52	12-53	3.4-24	13-55	4.5-28
N	20,553	8,391	12,162	18,334	2,219	17,611	2,942
Tau (τ)	0.52	0.52		0.50		0.49	
Wells score (age-adjusted D-dimer), %	32	31	32	34	15	36	15
95% CI	27-37	27-36	27-37	29-39	12-18	31-40	12-18
95% PI	15-55	15-55	15-55	17-56	6.2-31	19-57	6.7-30
N	20,553	8,391	12,162	18,334	2,219	17,611	2,942
Tau (τ)	0.46	0.46		0.44		0.41	
Revised Geneva score (fixed D-dimer), %	30	30	31	33	12	33	21
95% CI	26-36	25-35	26-36	28-38	9.3-15	28-38	17-26
95% PI	13-55	13-54	14-55	15-57	4.6-27	15-58	8.6-43
N	20,553	8,391	12,162	18,334	2,219	17,611	2,942
Tau (τ)	0.49	0.49		0.47		0.49	
Revised Geneva score (age-adjusted D-dimer), %	37	36	37	39	18	39	27
95% CI	32-41	31-41	33-42	35-44	15-21	35-44	23-32
95% PI	19-58	19-57	20-59	22-60	8.5-34	21-60	14-47
N	20,553	8,391	12,162	18,334	2,219	17,611	2,942
Tau (τ)	0.41	0.41		0.39		0.40	
YEARS algorithm, %	41	40	42	44	21	44	32
95% CI	36-47	35-45	37-47	39-49	17-25	39-49	27-37
95% PI	22-64	21-62	23-65	24-66	9.6-39	24-66	16-54
N	20,553	8,391	12,162	18,334	2,219	17,611	2,942
Tau (τ)	0.43	0.43		0.42		0.42	
PEGeD algorithm (Wells), %	47	45	48	50	26	51	31
95% CI	42-52	40-50	43-54	45-55	22-30	46-56	26-35
95% PI	26-69	24-67	27-70	29-71	12-46	30-72	15-52
N	20,553	8,391	12,162	18,334	2,219	17,611	2,942
Tau (τ)	0.44	0.44		0.42		0.42	
PEGeD algorithm (Geneva), %	44	42	46	48	17	48	26
95% CI	39-50	37-48	40-51	42-53	14-21	43-54	22-31
95% PI	23-68	22-66	24-69	26-70	7.3-35	27-71	12-48
N	20,553	8,391	12,162	18,334	2,219	17,611	2,942
Tau (τ)	0.46	0.46		0.44		0.44	

CDR: clinical decision rule; PE: pulmonary embolism; CI: confidence interval; PI = prediction interval; N: number of patients included in the analysis; VTE = venous thromboembolism *Defined as the predicted probability of ruling out PE based on CDR and D-dimer testing alone. For example, the overall predicted efficiency for the Wells score with fixed D-dimer threshold was 26%. This number means that if one uses this strategy, 26% of the patients are likely to be considered ''PE ruled out'' based on the strategy alone; thus imaging can likely be avoided in 26% of the patients.

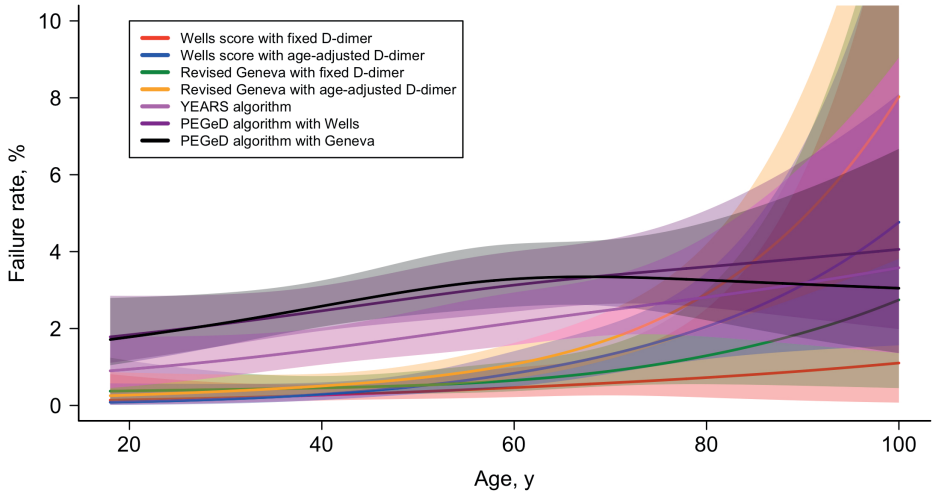


Figure 1. Failure rates with 95% CIs of the clinical decision rules and D-dimer testing in excluding pulmonary embolism versus age as a continuous variable
CI: confidence interval; y: years
Shading indicates 95% CIs.
Tau value in consecutive order from 'Wells score with fixed D-dimer' to 'PEGeD algorithm with Geneva': 0.34; 0.32; 0.39; 0.37; 0.38; 0.34; 0.34.

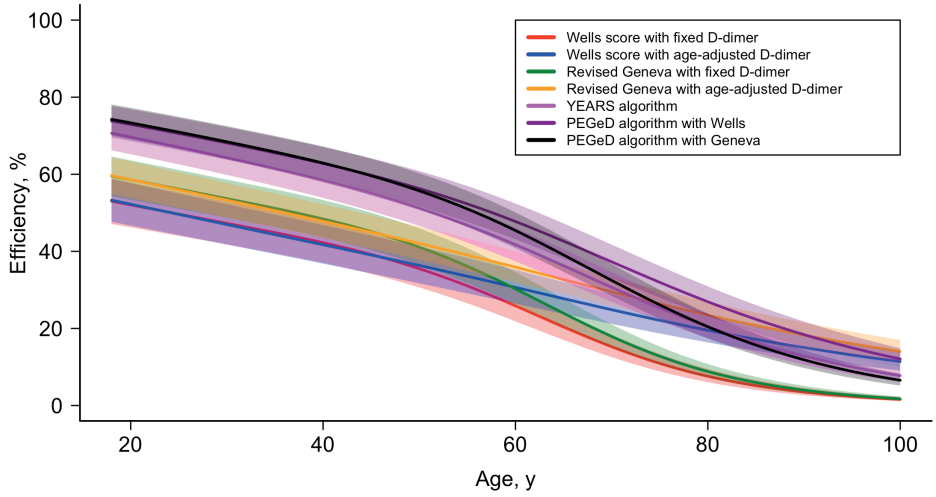


Figure 2. Efficiency with 95% CIs of the clinical decision rules and D-dimer testing in excluding pulmonary embolism versus age as a continuous variable
CI: confidence interval; y: years
Shading indicates 95% CIs.
Tau value in consecutive order from 'Wells score with fixed D-dimer' to 'PEGeD algorithm with Geneva': 0.44; 0.42; 0.39; 0.35; 0.35; 0.37; 0.37

For both failure rate and efficiency of each diagnostic strategy, in general, there was considerable between study heterogeneity as shown by wide prediction intervals, the t coefficients in each subgroup, and the forest plots (**Appendix Figures 6 and 7**, available at Annals.org).

The sensitivity analysis using only VTE events diagnosed at baseline as the outcome yielded point estimates for failure rate that were slightly lower than in the main analysis (**Appendix Table 2**, available at Annals.org).

4. DISCUSSION

The main finding of this diagnostic IPDMA is that the performance of the diagnostic strategies under study varied substantially across different patient subgroups. The predicted failure rate was generally highest for strategies incorporating adapted D-dimer thresholds. However, at the same time, predicted overall efficiency was substantially higher with these strategies versus strategies with a fixed D-dimer threshold as well. Efficiency was highest in patients younger than 40 years and gradually decreased with age, with the lowest efficiency in patients aged 80 years or older across all strategies. The considerable increase in efficiency when applying variable D-dimer thresholds in patients with cancer, elderly patients, and patients with a history of VTE was accompanied by predicted failure rates varying between 2% and 4% and, importantly, also by increasing uncertainty around these estimates as reflected by wide confidence and prediction margins (decreased precision and increased heterogeneity, respectively).

The clinical consequences of interpreting the safety of the different strategies are not straightforward for several reasons. First, as dictated by the Bayes theorem, a higher failure rate is to be expected in groups with a higher PE risk, such as patients with cancer.[26] These patients often have persistent risk factors for PE, which can lead to inflation of the predicted failure rate of the strategy. This is supported by data showing that computed tomography pulmonary angiography (CTPA) itself is not “failure rate free”, with a reported 3-month VTE incidence of 1.2% (CI, 0.48% to 2.6%).[48] However, in patients with a likely or high clinical probability (score >4 on the dichotomized Wells score and ≥6 on the 3-level Wells score), the failure rate was even higher (2.0% [CI, 1.0% to 4.1%] and 6.3% [CI, 3.0% to 13%], respectively).[48] Consequently, VTE diagnosis after initial negative testing for PE is not necessarily a “true” failure of the diagnostic strategy at baseline, because some of these are likely de novo thrombotic events that are unrelated to the index presentation. Therefore, a strategy with a failure rate exceeding the margin of 2% recommended by International Society on Thrombosis and Haemostasis standards is

not unsafe in high-risk patients per se.[26] Notably, this standard was meant to guide the design of prospective studies and not to determine clinical practice.

Second, and supporting the statements in the previous paragraph, increased use of imaging tests will lead to a substantial increase in the detection of isolated subsegmental PE, a condition that may not always require treatment.[49,50] Notably, in this IPDMA, the YEARS algorithm and the Wells rule with a D-dimer threshold dependent on pretest probability (PEGeD) appeared to be less safe than reported in their original validation studies, as was the case for use of an age-adjusted D-dimer threshold. These strategies refer fewer patients for imaging than would have been the case with the fixed D-dimer threshold of 500 µg/L. Because the strategy with a D-dimer threshold dependent on pretest probability was applied retrospectively in almost all studies in this IPDMA, more patients actually underwent imaging than in the prospective studies that originally validated this strategy. As a result, a greater number of isolated subsegmental PE cases were probably detected, contributing to differential verification bias, as confirmed by a recent study.[51] Unfortunately, data on the location of PE were not available in this IPDMA data set, precluding strong conclusions.

Lastly, the failure rate of the strategies may have been overestimated, especially for elderly patients and patients with cancer, due to misclassification; patients who died during follow-up, which occurred frequently in these 2 categories, were often considered to have had recurrent PE by the clinical event committees, even though these recurrences were not confirmed by imaging or autopsy.[52] In this IPDMA, 40% of all PE recurrences were fatal (in the studies where this information was available), representing patients who died during follow-up without sufficient information to determine the likely cause of death. Only in recent years was a more fair and practical definition of “fatal PE” for clinical trials adapted, classifying fatal PE based on autopsy, imaging tests, or most likely cause of death and not based on undetermined cause of death.[53] Importantly, patients who undergo imaging are not included in the safety analysis of our study. In strategies using adapted D-dimer thresholds, more patients managed without imaging will die during follow-up, simply because more patients are managed without imaging in general. This may further lead to an overestimation of the failure rate associated with adapted D-dimer thresholds. From that point of view, studying the safety of a diagnostic strategy in the complete population may be preferred, rather than only in those managed without imaging. Considering all of these factors, we do not believe that there are safety concerns with the available strategies in the patient subgroups included in our analyses, notwithstanding the observation that some uncertainty and heterogeneity of the failure rate remains, especially in the oldest patients. Thus, given this uncertainty, and acknowledging that patients in the subgroups studied in our analysis also remain

at high risk for new thrombotic events during follow-up, a reassessment should be initiated at a relatively low threshold if symptoms progress or persist.

The results of this IPDMA demonstrate that efficiency was highest for strategies applying adapted D-dimer thresholds (age-adjusted or pretest probability–dependent), as they increase the number of patients in whom PE can be ruled out without imaging by up to 20% overall. The relative efficiency increase with these variable D-dimer thresholds was highest in the subgroups of elderly patients, patients with cancer, and patients with a previous VTE. From our analysis, it seems that D-dimer thresholds dependent on pretest probability were more efficient than age-adjusted thresholds.

Even though we could not identify an overall preferred diagnostic strategy, the numbers presented in this study will inform physicians and policymakers as they decide on the optimal strategy in their particular patient subgroup, by balancing the risks of unnecessary CTPA with possibly untreated PE on an individual basis. Our interpretation of the findings is as follows. All studied strategies can be used in both the overall population with suspected PE and in relevant patient subgroups, including elderly patients, those with cancer, or those with suspected recurrent VTE. In our practice, we therefore apply a strategy with adapted D-dimer thresholds for obvious reasons of efficacy, with a number needed to test to avoid 1 CTPA ranging from 10 patients when a fixed D-dimer threshold is used to 3 or 4 patients when an adapted D-dimer threshold is used. We acknowledge that the clinical utility remains limited in the most elderly patients, even when an adapted D-dimer threshold is applied. Importantly, as the benefit of diagnostic strategies for suspected PE is largely dependent on their correct application, we propose incorporating 1 strategy as the standard of care in each individual hospital rather than choosing a particular strategy based on the characteristics of individual patients. After all, standardization is key to achieving optimal adherence. Whether clinicians should rely on the Wells rule, the YEARS algorithm, or the revised Geneva score becomes a matter of local preference and experience. Ultimately, in light of the increasing uncertainty of our findings in specific subgroups, randomized controlled trials directly comparing the application of different diagnostic strategies in these subgroups are necessary to understand which diagnostic strategy is superior. As an example, an ongoing international randomized controlled trial in patients with cancer and suspected PE is evaluating the safety and efficiency of the YEARS algorithm, directly compared with CTPA, in all patients (Netherlands Trial Register NL7752).

The large number of patients included in this meta-analysis is a major strength. This enabled more robust subgroup analyses on frequently encountered subgroups than was possible in the individual original studies alone. Moreover, all studies reported original

data on the method of pretest probability assessment, assessed variables to calculate at least 1 decision rule, and used a well-accepted diagnostic reference standard. Furthermore, after multilevel imputation of missing values, diagnostic strategies were directly compared in the same set of studies, limiting bias due to between-study heterogeneity.[54] Still, important limitations need to be discussed. First, although information collected during follow-up included information on anticoagulant therapy for reasons other than VTE and loss to follow-up, we were not able to exclude these patients in the failure rate analyses as this information was not available in most of the studies. This approach could have led to an underestimation of the observed failure rates. Second, there were some systematically missing values in our IPD. Rather than excluding studies that had any systematically missing values, we used 1-stage multilevel chained equations to impute them. However, as in any other imputation method, these methods require assumptions. Therefore, it remains possible that misspecification of our imputation model may have affected our results. Finally, the availability and definition of items included in the diagnostic strategies differed between included studies. This between-study heterogeneity was illustrated by the relatively wide prediction intervals around the estimates, notably for elderly patients, patients with cancer, and patients with a history of VTE. Also, various D-dimer assays were used in the different studies. Although these widely used quantitative assays have a high sensitivity for diagnosing PE, performance of these assays could have evolved over the course of 20 years. Nevertheless, we believe that this reflects current clinical practice, strengthening the external validity of our results.

In conclusion, in this IPDMA, the safety and efficiency of the studied diagnostic strategies varied across different patient subgroups. Overall, the studied strategies might all be considered safe across the predefined patient subgroups, which does not allow for favoring one over the other. Importantly, this conclusion was drawn on the basis of the arguments of the Bayes theorem as well as verification and misclassification bias, which may have led to an overestimation of the failure rate of strategies with adapted D-Dimer thresholds. From an efficiency perspective, this IPDMA supports the use of these adapted D-dimer thresholds. Pending the results of ongoing diagnostic randomized trials, physicians and guideline committees should balance the interlink between safety and efficiency of available diagnostic strategies.

REFERENCES

1. Huisman MV, Barco S, Cannegieter SC, et al. Pulmonary embolism. Nat Rev Dis Primers. 2018;4:18028.
2. van Es N, van der Hulle T, van Es J, et al. Wells rule and d-dimer testing to rule out pulmonary embolism: a systematic review and individual-patient data meta-analysis. Ann Intern Med. 2016;165:25361.
3. Pasha SM, Klok FA, Snoep JD, et al. Safety of excluding acute pulmonary embolism based on an unlikely clinical probability by the Wells rule and normal D-dimer concentration: a meta-analysis. Thromb Res. 2010;125:e123-7.
4. Hurwitz LM, Reiman RE, Yoshizumi TT, et al. Radiation dose from contemporary cardiothoracic multidetector CT protocols with an anthropomorphic female phantom: implications for cancer induction. Radiology. 2007;245:742-50.
5. Kooiman J, Klok FA, Mos IC, et al. Incidence and predictors of contrast-induced nephropathy following CT-angiography for clinically suspected acute pulmonary embolism [Letter]. J Thromb Haemost. 2010;8:409-11.
6. van der Pol LM, Dronkers CEA, van der Hulle T, et al. The YEARS algorithm for suspected pulmonary embolism: shorter visit time and reduced costs at the emergency department. J Thromb Haemost. 2018;16:725-733.
7. Huisman MV, Klok FA. How I diagnose acute pulmonary embolism. Blood. 2013;121:4443-8.
8. Righini M, Van Es J, Den Exter PL, et al. Age-adjusted D-dimer cutoff levels to rule out pulmonary embolism: the ADJUST-PE study. JAMA. 2014;311:1117-24.
9. Douma RA, le Gal G, Söhne M, et al. Potential of an age adjusted D-dimer cut-off value to improve the exclusion of pulmonary embolism in older patients: a retrospective analysis of three large cohorts. BMJ. 2010;340:c1475.
10. van der Hulle T, Cheung WY, Kooij S, et al; YEARS study group. Simplified diagnostic management of suspected pulmonary embolism (the YEARS study): a prospective, multicentre, cohort study. Lancet. 2017;390:289-297.
11. Kearon C, de Wit K, Parpia S, et al; PEGeD Study Investigators. Diagnosis of pulmonary embolism with d-dimer adjusted to clinical probability. N Engl J Med. 2019;381:2125-2134.
12. van Belle A, Büller HR, Huisman MV, et al; Christopher Study Investigators. Effectiveness of managing suspected pulmonary embolism using an algorithm combining clinical probability, Ddimer testing, and computed tomography. JAMA. 2006;295:172-9.
13. Douma RA, Mos IC, Erkens PM, et al; Prometheus Study Group. Performance of 4 clinical decision rules in the diagnostic management of acute pulmonary embolism: a prospective cohort study. Ann Intern Med. 2011;154:709-18.
14. Mos IC, Douma RA, Erkens PM, et al; Prometheus Study Group. Diagnostic outcome management study in patients with clinically suspected recurrent acute pulmonary embolism with a structured algorithm. Thromb Res. 2014;133:1039-44.
15. van der Pol LM, Tromeur C, Bistervels IM, et al; Artemis Study Investigators. Pregnancy-adapted YEARS algorithm for diagnosis of suspected pulmonary embolism. N Engl J Med. 2019;380:11391149.
16. van der Hulle T, den Exter PL, Mos IC, et al. Optimization of the diagnostic management of clinically suspected pulmonary embolism in hospitalized patients. Br J Haematol. 2014;167:681-6.
17. Karami-Djurabi R, Klok FA, Kooiman J, et al. D-dimer testing in patients with suspected pulmonary embolism and impaired renal function. Am J Med. 2009;122:1050-3.

18. Stals MAM, Klok FA, Huisman MV. Diagnostic management of acute pulmonary embolism in special populations. Expert Rev Respir Med. 2020;14:729-736.
19. Geersing GJ, Kraaijpoel N, Büller HR, et al. Ruling out pulmonary embolism across different subgroups of patients and healthcare settings: protocol for a systematic review and individual patient data meta-analysis (IPDMA). Diagn Progn Res. 2018;2:10.
20. Stewart LA, Clarke M, Rovers M, et al; PRISMA-IPD Development Group. Preferred reporting items for systematic review and meta-analyses of individual participant data: the PRISMA-IPD statement. JAMA. 2015;313:1657-65.
21. McInnes MDF, Moher D, Thombs BD, et al; PRISMA-DTA Group. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA. 2018;319:388-396.
22. Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162:55-63.
23. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med. 2019;170:W1-W33.
24. Wolff RF, Moons KGM, Riley RD, et al; PROBAST Group. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170:51-58.
25. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529-36.
26. Dronkers CEA, van der Hulle T, Le Gal G, et al; Subcommittee on Predictive and Diagnostic Variables in Thrombotic Disease. Towards a tailored diagnostic standard for future diagnostic studies in pulmonary embolism: communication from the SSC of the ISTH. J Thromb Haemost. 2017;15:1040-1043.
27. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557-60.
28. Jolani S, Debray TP, Koffijberg H, et al. Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE. Stat Med. 2015;34:184163.
29. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393.
30. Marshall A, Altman DG, Holder RL, et al. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Med Res Methodol. 2009;9:57.
31. Sanson BJ, Lijmer JG, Mac Gillavry MR, et al. Comparison of a clinical probability estimate and two clinical models in patients with suspected pulmonary embolism. ANTELOPE-Study Group. Thromb Haemost. 2000;83:199-203.
32. Geersing GJ, Erkens PM, Lucassen WA, et al. Safe exclusion of pulmonary embolism using the Wells rule and qualitative D-dimer testing in primary care: prospective cohort study. BMJ. 2012;345: e6564.
33. Kearon C, Ginsberg JS, Douketis J, et al; Canadian Pulmonary Embolism Diagnosis Study (CANPEDS) Group. An evaluation of D-dimer in the diagnosis of pulmonary embolism: a randomized trial. Ann Intern Med. 2006;144:812-21.
34. Kline JA, Nelson RD, Jackson RE, et al. Criteria for the safe use of D-dimer testing in emergency department patients with suspected pulmonary embolism: a multicenter US study. Ann Emerg Med. 2002;39:144-52.

35. Kline JA, Runyon MS, Webb WB, et al. Prospective study of the diagnostic accuracy of the Simplify D-dimer assay for pulmonary embolism in emergency department patients. Chest. 2006;129:1417-23.
36. Kline JA, Courtney DM, Kabrhel C, et al. Prospective multicenter evaluation of the pulmonary embolism rule-out criteria. J Thromb Haemost. 2008;6:772-80.
37. Runyon MS, Beam DM, King MC, et al. Comparison of the Simplify D-dimer assay performed at the bedside with a laboratorybased quantitative D-dimer assay for the diagnosis of pulmonary embolism in a low prevalence emergency department population. Emerg Med J. 2008;25:70-5.
38. Kline JA, Hogg MM, Courtney DM, et al. D-dimer threshold increase with pretest probability unlikely for pulmonary embolism to decrease unnecessary computerized tomographic pulmonary angiography. J Thromb Haemost. 2012;10:572-81.
39. Penaloza A, Soulie C, Moumneh T, et al. Pulmonary embolism rule-out criteria (PERC) rule in European patients with low implicit clinical probability (PERCEPIC): a multicentre, prospective, observational study. Lancet Haematol. 2017;4:e615-e621.
40. Goekoop RJ, Steeghs N, Niessen RW, et al. Simple and safe exclusion of pulmonary embolism in outpatients using quantitative D-dimer and Wells' simplified decision rule. Thromb Haemost. 2007;97:146-50.
41. Schouten HJ, Geersing GJ, Oudega R, et al. Accuracy of the Wells clinical prediction rule for pulmonary embolism in older ambulatory adults. J Am Geriatr Soc. 2014;62:2136-41.
42. Wicki J, Perneger TV, Junod AF, et al. Assessing clinical probability of pulmonary embolism in the emergency ward: a simple score. Arch Intern Med. 2001;161:92-7.
43. Perrier A, Roy PM, Aujesky D, et al. Diagnosing pulmonary embolism in outpatients with clinical assessment, D-dimer measurement, venous ultrasound, and helical computed tomography: a multicenter management study. Am J Med. 2004;116:291-9.
44. Perrier A, Roy PM, Sanchez O, et al. Multidetector-row computed tomography in suspected pulmonary embolism. N Engl J Med. 2005;352:1760-8.
45. Righini M, Le Gal G, Aujesky D, et al. Diagnosis of pulmonary embolism by multidetector CT alone or combined with venous ultrasonography of the leg: a randomised non-inferiority trial. Lancet. 2008;371:1343-52.
46. Galipienzo J, Garcia de Tena J, Flores J, et al. Effectiveness of a diagnostic algorithm combining clinical probability, D-dimer testing, and computed tomography in patients with suspected pulmonary embolism in an emergency department. Rom J Intern Med. 2012 Jul-Sep;50:195-202.
47. Ghanima W, Almaas V, Aballi S, et al. Management of suspected pulmonary embolism (PE) by D-dimer and multi-slice computed tomography in outpatients: an outcome study. J Thromb Haemost. 2005;3:1926-32.
48. van der Hulle T, van Es N, den Exter PL, et al. Is a normal computed tomography pulmonary angiography safe to rule out acute pulmonary embolism in patients with a likely clinical probability? A patient-level meta-analysis. Thromb Haemost. 2017;117:1622-9.
49. Carrier M, Klok FA. Symptomatic subsegmental pulmonary embolism: to treat or not to treat. Hematology Am Soc Hematol Educ Program. 2017;2017:237-241.
50. den Exter PL, Kroft LJM, Gonsalves C, et al. Establishing diagnostic criteria and treatment of subsegmental pulmonary embolism: a Delphi analysis of experts. Res Pract Thromb Haemost. 2020;4:1251-1261.
51. van der Pol LM, Bistervels IM, van Mens TE, et al. Lower prevalence of subsegmental pulmonary embolism after application of the YEARS diagnostic algorithm. Br J Haematol. 2018;183:629635.

52. Girard P, Penaloza A, Parent F, et al. Reproducibility of clinical events adjudications in a trial of venous thromboembolism prevention. J Thromb Haemost. 2017;15:662-669.
53. Tritschler T, Kraaijpoel N, Girard P, et al; Subcommittee on Predictive and Diagnostic Variables in Thrombotic Disease. Definition of pulmonary embolism-related death and classification of the cause of death in venous thromboembolism studies: communication from the SSC of the ISTH. J Thromb Haemost. 2020;18:1495-1500.
54. Wang J, Bossuyt P, Geskus R, et al; IMPORT Study Group. Using individual patient data to adjust for indirectness did not successfully remove the bias in this case of comparative test accuracy. J Clin Epidemiol. 2015;68:290-8.

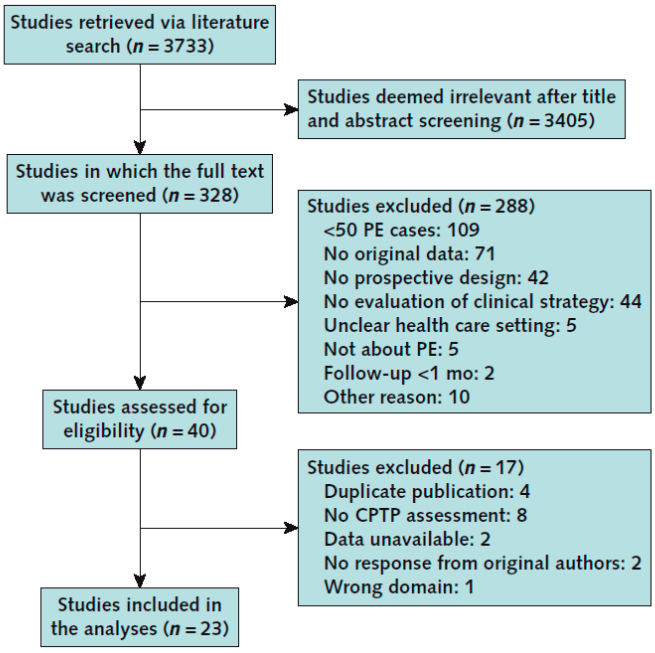
APPENDIX

Search String

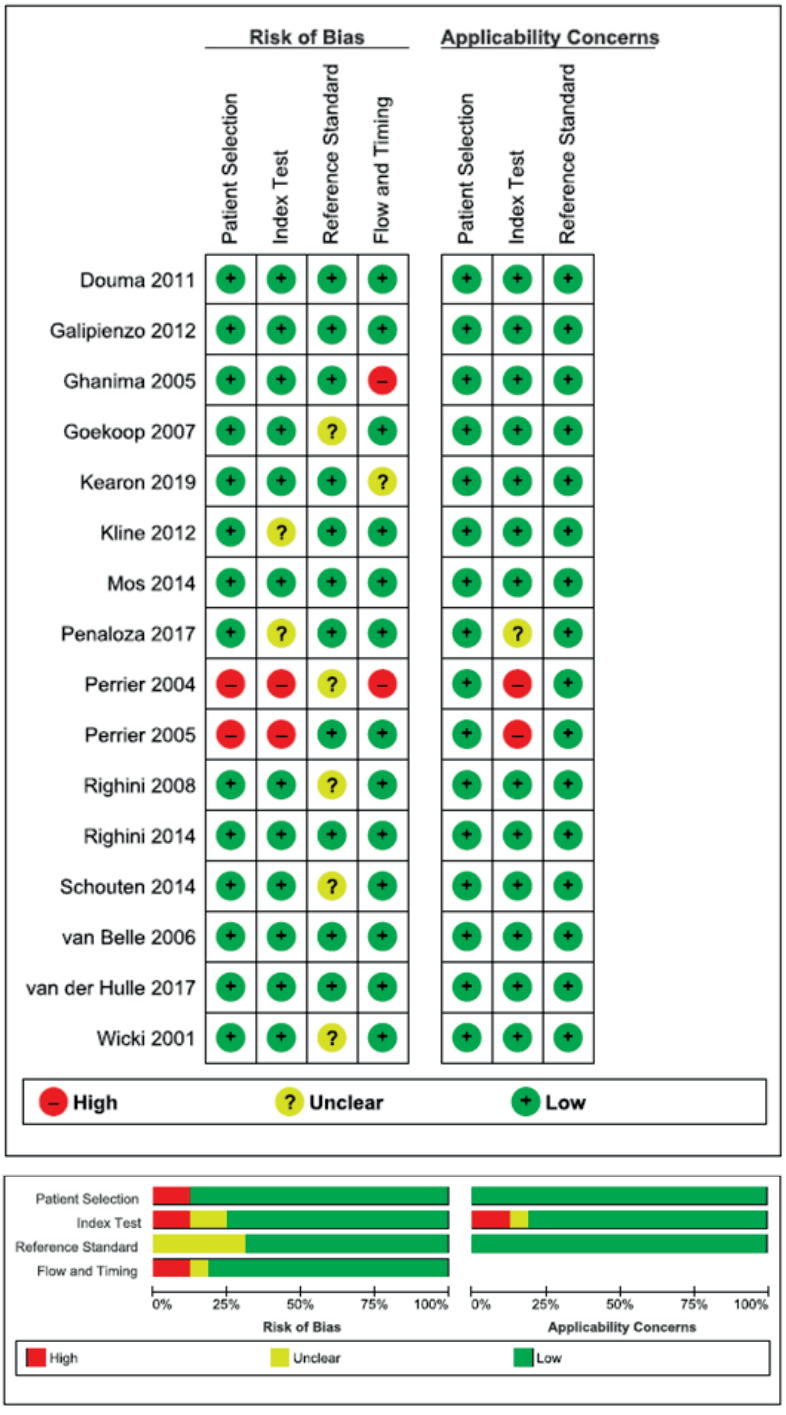
((Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$)) OR ("Stratification" OR "ROC Curve"[Mesh] OR "Discrimination" OR "Discriminate" OR "c-statistic" OR "c statistic" OR "Area under the curve" OR "AUC" OR "Calibration" OR "Indices" OR "Algorithm" OR "Multivariable"))

AND

("pulmonary embolism"[MeSH Terms] OR ("pulmonary"[All Fields] AND "embolism"[All Fields]) OR "pulmonary embolism"[All Fields])



Appendix Figure 1. Flowchart of included studies
CPTP = clinical pretest probability; PE = pulmonary embolism.



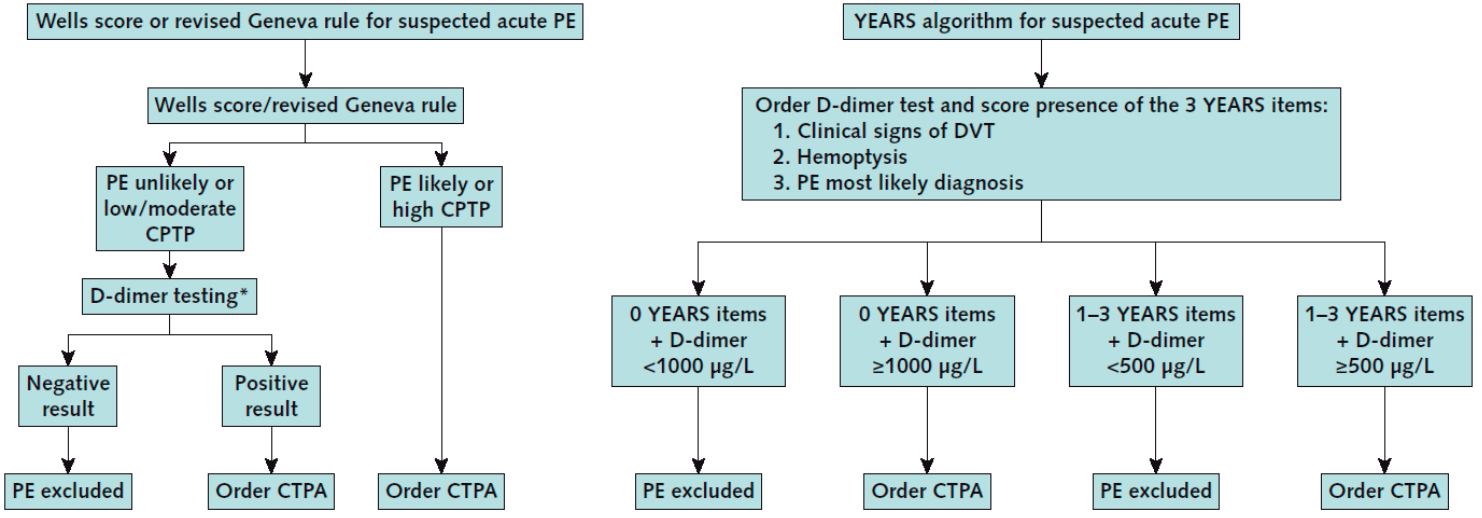
Appendix Figure 2. Risk-of-bias assessment using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies 2) tool

Diagnostic strategy	Wells rule		Revised Geneva score		YEARS algorithm	
Items and points	Previous VTE	1.5	Previous VTE	3	Clinical signs of deep-vein thrombosis	
	Heart rate >100/min	1.5	Heart rate		Hemoptysis	
	Surgery or immobilization <4 weeks	1.5	75-94/min	3	PE most likely diagnosis	
	Hemoptysis	1	≥ 95/min	5		
	Active cancer	1	Surgery or fracture <1 month	2		
	Clinical signs of deep-vein thrombosis	3	Hemoptysis	2		
	Alternative diagnosis less likely than PE	3	Active cancer	2		
			Unilateral lower limb pain	3		
			Pain on lower limb, deep venous palpation, and unilateral edema	4		
			Age >65 years	1		
Clinical pre-test probability assessment	**The original classification**		**The original classification**		Low	0 items
	Low	0-1.5	Low	0-3	High	≥ 1 items
	Intermediate	2-6	Intermediate	4-10		
	High	≥ 6.5	High	≥11		
	Unlikely	0-4	**For only D-dimer dependent on clinical pre-test probability**			
	Likely	≥ 4.5	Low	0-5		
	For only D-dimer dependent on clinical pre-test probability		Intermediate	6-10		
	Low	0-4	High	≥11		
	Moderate	4.5-6				
	High	≥ 6.5				

Diagnostic strategy	Wells rule		Revised Geneva score		YEARS algorithm	
Assessment of D-dimer testing						
Qualitative D-dimer	Normal	Negative	Normal	Negative		
	Abnormal	Positive	Abnormal	Positive		NA
Quantitative D-dimer with the traditional cut-off	Normal	<500 ng/ml	Normal	<500 ng/ml		
	Abnormal	≥ 500ng/ml	Abnormal	≥ 500ng/ml		NA
Quantitative D-dimer adjusted to age	Normal	<500 ng/ml for younger than 50 years	Normal	<500 ng/ml for younger than 50 years		NA
		< Age * 10 for 50 years or older		< Age * 10 for 50 years or older		
	Abnormal	≥ 500 ng/ml for younger than 50 years	Abnormal	≥ 500 ng/ml for younger than 50 years		
		≥ Age * 10 for 50 years or older		≥ Age * 10 for 50 years or older		
Quantitative D-dimer dependent on clinical pre-test probability	Normal	<1000 ng/ml for a low clinical probability	Normal	<1000 ng/ml for a low clinical probability	Normal	<1000 ng/ml for a low clinical probability
		<500 ng/ml for a moderate clinical probability		<500 ng/ml for an intermediate clinical probability		<500 ng/ml for a high clinical probability
	Abnormal	≥ 1000 ng/ml for a low clinical probability	Abnormal	≥ 1000 ng/ml for a low clinical probability	Abnormal	≥ 1000 ng/ml for a low clinical probability
		≥ 500 ng/ml for a moderate clinical probability		≥ 500 ng/ml for an intermediate clinical probability		≥ 500 ng/ml for a high clinical probability

Diagnostic strategy	Wells rule	Revised Geneva score	YEARS algorithm
Further testing	For traditional or age-adjusted D-dimer testing, Unlikely, Low, or Intermediate plus Abnormal D-dimer testing, or High or Likely regardless of D-dimer testing.	Low or Intermediate plus abnormal D-dimer testing. High or Likely regardless of D-dimer testing. In all other patients, PE is considered ruled-out.	Low or High plus abnormal D-dimer testing. In all other patients, PE is considered ruled out.
	For D-dimer dependent on clinical pre-test probability, Low or Moderate plus abnormal D-dimer testing, or high regardless of D-dimer testing		
	In all other patients, PE is considered ruled out.		

Appendix Figure 3. Diagnostic strategies under evaluation
CPTP = clinical pretest probability; DVT = deep-vein thrombosis; NA = not applicable; VTE = venous thromboembolism; PE = pulmonary embolism

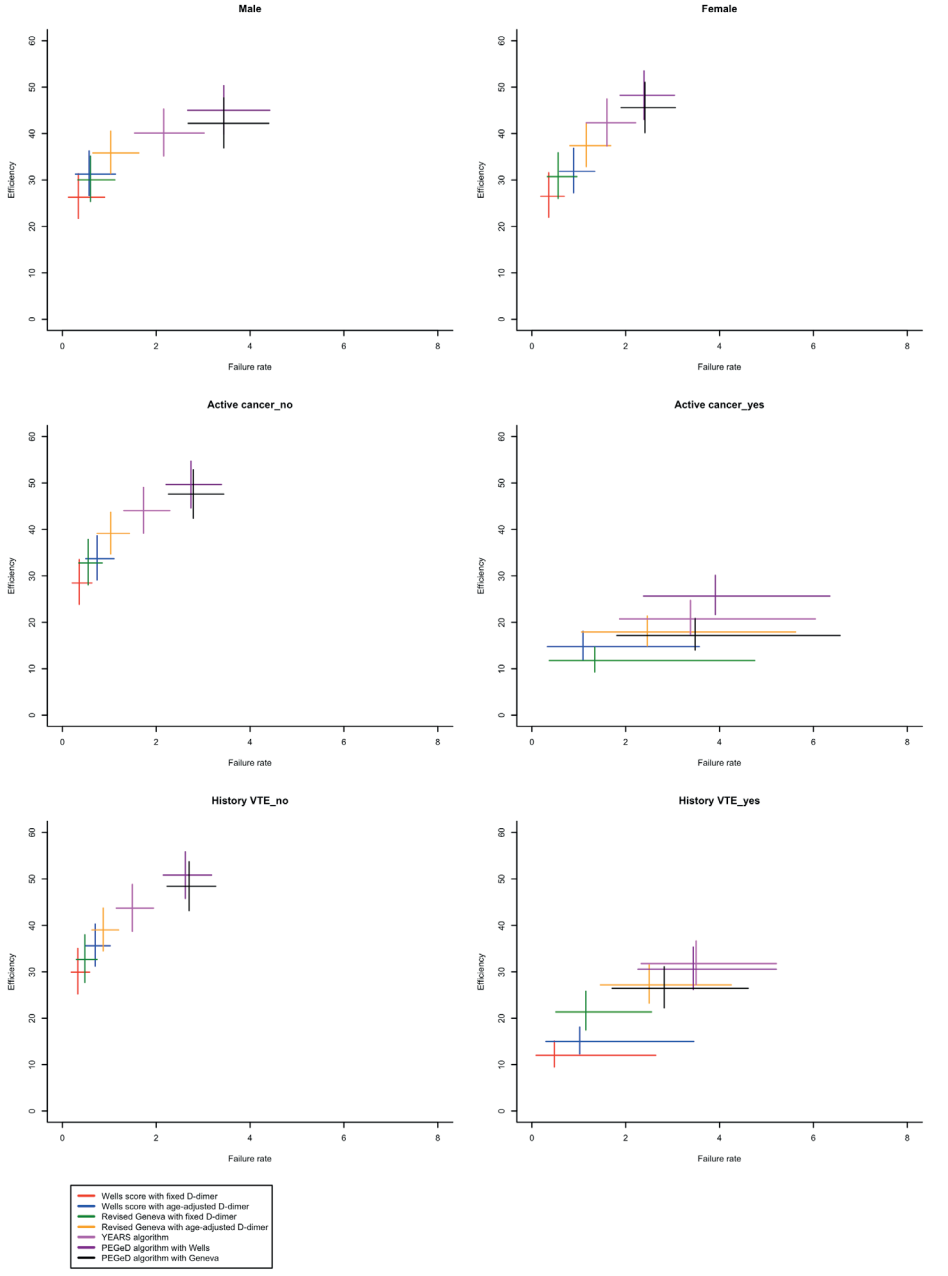


Appendix Figure 4. Diagnostic strategies for suspected pulmonary embolism in the present individual-patients data meta-analysis
PE: pulmonary embolism; CPTP: clinical pre-test probability assessment; CTPA: CT pulmonary angiography; DVT: deep-vein thrombosis
* D-dimer testing: fixed (<500 µg/L) or age-adjusted D-dimer testing (age × 10 µg/L in patients aged >50 years) in PE unlikely patients, or D-dimer testing dependent on CPTP (<1000 ng/ml for a low clinical probability and <500 ng/ml for a moderate clinical probability) in patients with a low and moderate CPTP, respectively

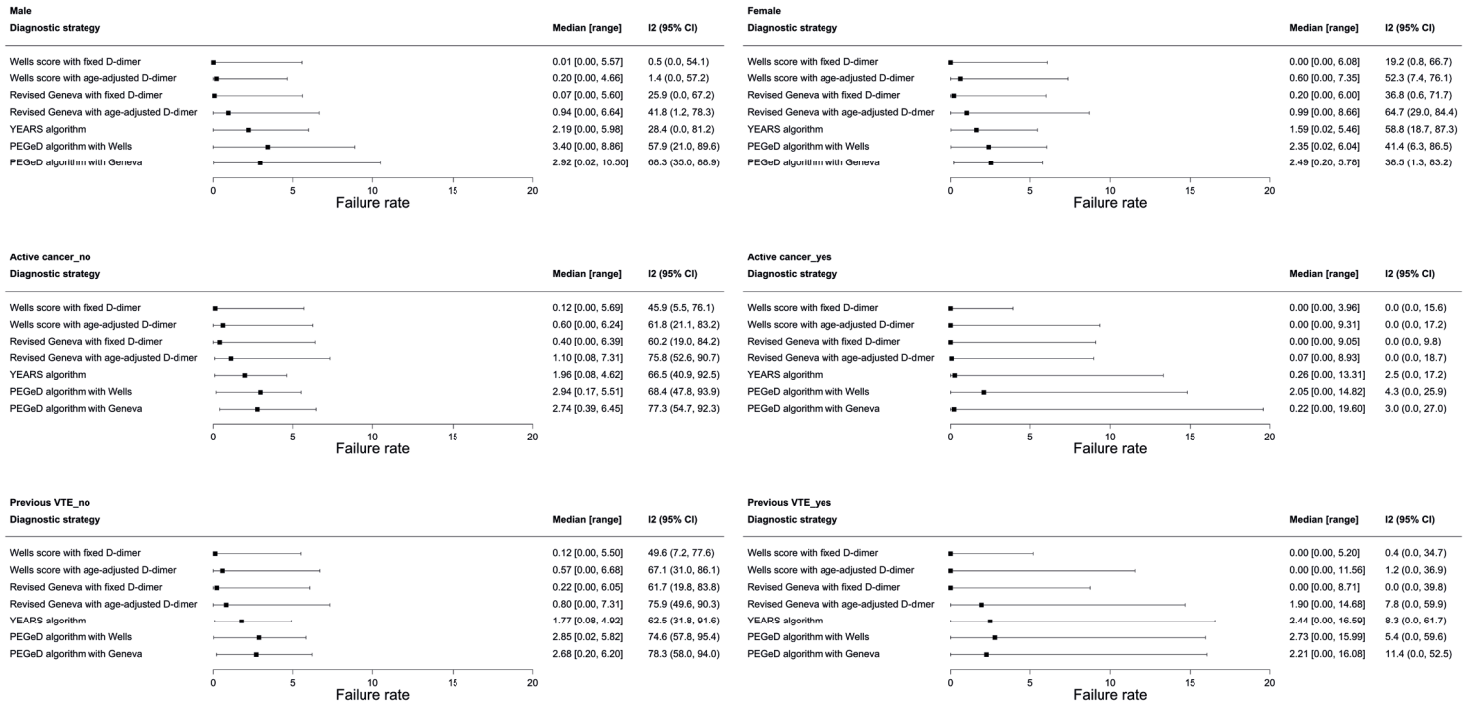
Appendix Table 1. Proportions of missing values in each study

	Kline et al, 2012	*Douma et al, 2011*	*Goekoop et al, 2007*	*Righini et al, 2014*	*Schouten et al, 2014*	*Christopher Study Investigators, 2006*	*Van der Hulle et al, 2017*	*Mos et al, 2014*	*Wicki et al, 2001*	*Perrier et al, 2004*	*Perrier et al, 2005*	*Righini et al, 2008*	*Kearon et al, 2019*	*Galipienzo et al, 2012*	*Ghanima et al, 2005*	*Penaloza et al, 2017*
n	334	807	876	3324	294	3296	3448	279	1089	965	1692	755	2017	240	432	705
Inpatient	0	0	0	0	0	0	0	2,9	0	0	0	0	0	0	0	0
Sex	0	0	0	0	0	0,3	0	0,4	0	0	0	0	0	0	0	0
Age	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BMI	2,1	9,8	100	31,7	100	100	47,1	33,7	100	100	100	100	100	100	100	100
PreviousVTE	0	0	0,2	0	1	0,2	0,1	0	0	0,2	0	0	0	0	0,5	0
HR	0	0	100	4,9	3,4	100	2	61,3	0,1	0,4	0,1	0,4	100	100	4,2	0
Tachycardia	0	0	0,1	4,9	3,4	0,2	2	61,3	0,1	0,4	0,1	0,4	0	0	4,2	0
ImmoSurg	0	0	0	0	0,3	0,2	0,1	24,4	0	0	0	0	0	0	0	0
Hemoptysis	0	0	0,2	0,1	0,7	0,2	0	26,2	0	0	0	0	0	0	28,5	0
Cancer	0	0	0,1	0	0,3	0,2	0,1	20,8	0	0,3	0	0	0	0	0,7	0
CHF	0	0,2	1,4	3,4	1	0,5	0	10	100	0	13,6	0	100	100	100	3,1
ChronicLungDisease	0	0,1	1,4	3,1	1,7	0,5	0	11,5	0	0	0	0	100	100	100	100
SympDVT	0	0	0,1	2,7	0	0,2	0	20,8	37,3	0	0,1	0,1	0	0	28,7	0
Estrogen	0	0,7	1,3	1,5	0	1,1	1	15,1	0	0	0	0	66,2	0	0,5	0
DurationSymptoms	100	2	0,5	100	2,7	2,1	0,7	1,4	100	100	100	100	0,1	100	28,7	29,2
PEmostlikely	33	0	0	0	0	0,2	0	17,9	100	3,5	2,3	1,1	0	0	100	0
SBP	0,9	100	100	6,3	11,6	100	100	100	0,3	0,6	0,3	0,5	100	100	100	100
SpO2peripheral	0,9	100	100	9,1	100	100	100	100	100	14,3	5,5	2,3	100	100	100	100
DdimerQual	100	100	100	100	6,1	100	100	100	0	0	0,5	0,1	100	100	100	100
DdimerQuant	0	7,4	8,9	7,3	50	15,6	0,2	27,2	0	0,1	0,5	1,2	0,6	2,5	0	13,2
VTEfinal	0	0	0	0	0,3	0	0	0	0	0	0	0	0	0	0	0

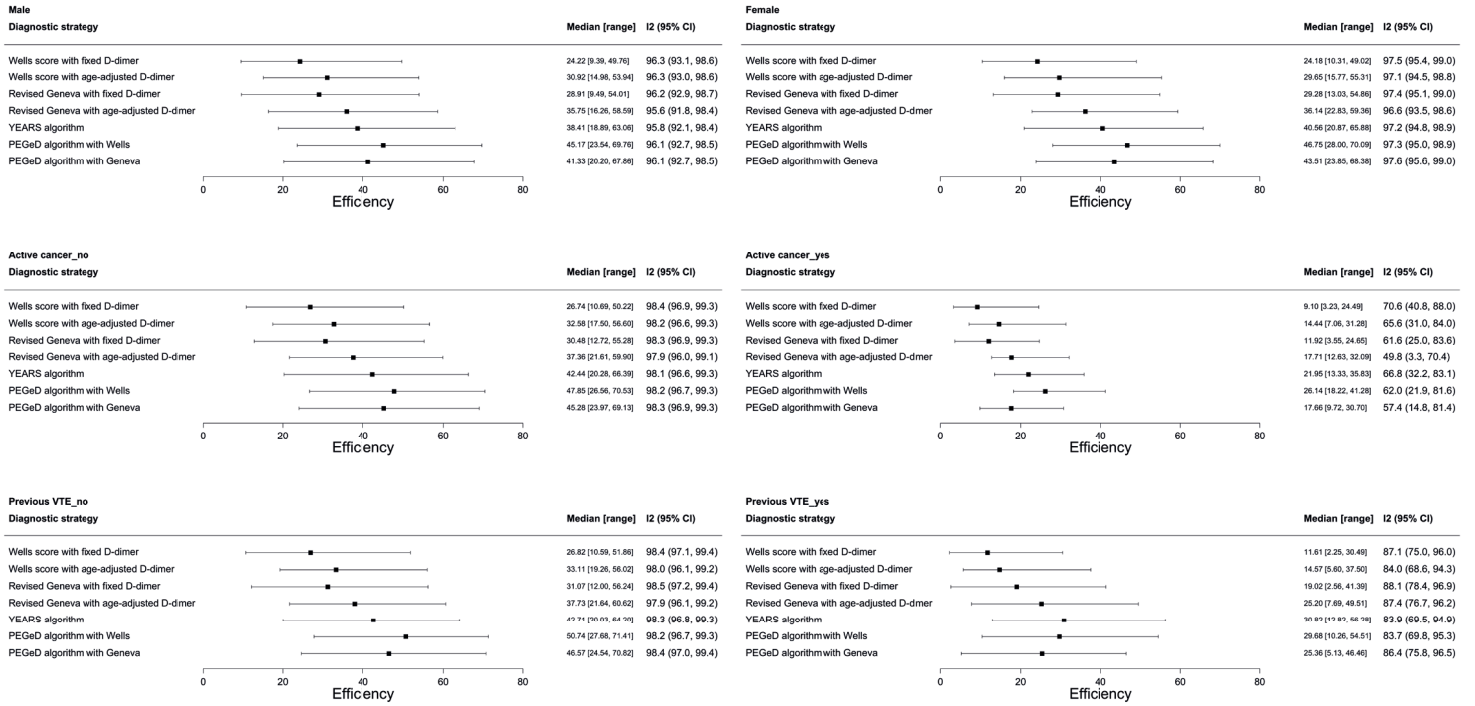
n= number; BMI= body mass index; VTE = venous thromboembolism; ImmoSurg= immobility or surgery in the previous 4 weeks (Wells item); DVT = deep-vein thrombosis; PE = pulmonary embolism



Appendix Figure 5. Crosshair figures of safety versus efficiency of the clinical decision rules and D-dimer testing in clinically relevant patient subgroups (according to main analysis)
The intersection of the two lines indicates the point estimate of the predicted failure rate and efficiency, and the 95% CI is indicated by the length of both lines.
CDR: clinical decision rule; VTE: venous thromboembolism.



Appendix Figure 6. Forest plot with I-square statistics for failure rate of the different diagnostic strategies across patient subgroups
Note: CI: confidence interval; VTE: venous thromboembolism



Appendix Figure 7. Forest plot with I-square statistics for efficiency of the different diagnostic strategies across patient subgroups
CI: confidence interval; VTE: venous thromboembolism

Appendix Table 2. Failure rate* of the CDRs and D-dimer testing in excluding PE, overall and in clinically relevant patient subgroups (according to sensitivity analysis: only including VTE events diagnosed at baseline)

Variable	*Overall*	*Sex*		*Age*				*Active cancer*		*History VTE*	
		Male	*Female*	*<40y*	*40-59y*	*60-79y*	*≥80y*	*No*	*Yes*	*No*	*Yes*
Wells score (fixed D-dimer), %	0.30	0.31	0.29	0.15	0.29	0.57	NE~	0.30	NE~	0.27	0.45
95% CI	0.16-0.55	0.12-0.83	0.14-0.58	0.04-0.54	0.12-0.71	0.22-1.5	NE	0.17-0.55	NE	0.15-0.50	0.08-2.5
95% PI	0.11-0.79	0.09-1.1	0.10-0.82	0.03-0.72	0.09-0.97	0.16-2.0	NE	0.12-0.79	NE	0.11-0.66	0.06-3.2
N	20,553	8,391	12,162	4,229	6,854	7,104	2,366	18,334	2,219	17,611	2,942
Tau (τ)	0.33	0.33			0.33			0.33		0.27	
Wells score (age-adjusted D-dimer), %	0.65	0.45	0.79	0.15	0.43	1.1	2.2	0.63	1.1	0.60	0.98
95% CI	0.43-0.98	0.21-0.94	0.52-1.2	0.05-0.48	0.21-0.88	0.61-2.0	1.0-4.9	0.42-0.96	0.31-3.5	0.40-0.88	0.26-3.7
95% PI	0.29-1.5	0.16-1.3	0.35-1.8	0.04-0.62	0.15-1.2	0.44-2.8	0.76-6.4	0.28-1.4	0.24-4.5	0.29-1.2	0.20-4.6
N	20,553	8,391	12,162	4,229	6,854	7,104	2,366	18,334	2,219	17,611	2,942
Tau (τ)	0.32	0.31			0.31			0.32		0.27	
Revised Geneva score (fixed D-dimer), %	0.48	0.50	0.46	0.30	0.37	1.0	NE~	0.46	0.70	0.39	0.92
95% CI	0.30-0.77	0.25-0.99	0.26-0.80	0.13-0.72	0.18-0.74	0.52-2.0	NE	0.29-0.74	0.11-4.2	0.24-0.64	0.37-2.3
95% PI	0.18-1.2	0.17-1.5	0.17-1.2	0.09-1.0	0.12-1.1	0.35-3.0	NE	0.18-1.2	0.08-5.6	0.18-0.88	0.29-2.9
N	20,553	8,391	12,162	4,229	6,854	7,104	2,366	18,334	2,219	17,611	2,942
Tau (τ)	0.37	0.37			0.37			0.37		0.28	
Revised Geneva score (age-adjusted D-dimer), %	0.96	0.83	1.0	0.30	0.53	1.7	3.0	0.89	2.1	0.74	2.2
95% CI	0.68-1.3	0.49-1.4	0.72-1.5	0.13-0.70	0.30-0.96	1.1-2.6	1.7-5.4	0.63-1.3	0.82-5.0	0.53-1.0	1.2-3.9
95% PI	0.41-2.2	0.32-2.1	0.44-2.4	0.09-0.99	0.20-1.4	0.70-4.0	1.1-7.9	0.38-2.1	0.58-7.0	0.37-1.5	0.93-5.1
N	20,553	8,391	12,162	4,229	6,854	7,104	2,366	18,334	2,219	17,611	2,942
Tau (τ)	0.36	0.36			0.35			0.36		0.28	

Appendix Table 2. Failure rate* of the CDRs and D-dimer testing in excluding PE, overall and in clinically relevant patient subgroups (according to sensitivity analysis: only including VTE events diagnosed at baseline) (*continued*)

Variable	*Overall*	*Sex*		*Age*				*Active cancer*		*History VTE*	
		Male	*Female*	*<40y*	*40-59y*	*60-79y*	*≥80y*	*No*	*Yes*	*No*	*Yes*
YEARS algorithm, %	1.6	1.9	1.4	0.99	1.7	2.1	2.8	1.6	2.1	1.3	3.2
95% CI	1.2-2.2	1.4-2.8	1.0-2.0	0.62-1.6	1.2-2.4	1.4-3.1	1.3-5.8	1.2-2.1	0.91-4.8	1.0-1.8	2.1-4.8
95% PI	0.72-3.7	0.83-4.5	0.62-3.3	0.39-2.5	0.72-4.0	0.86-4.9	0.91-8.2	0.70-3.6	0.64-6.7	0.70-2.5	1.5-6.4
N	20,553	8,391	12,162	4,229	6,854	7,104	2,366	18,334	2,219	17,611	2,942
Tau (τ)	0.36	0.36			0.36			0.36		0.27	
PEGeD algorithm (Wells), %	2.6	3.1	2.2	2.0	2.5	3.1	3.3	2.6	2.5	2.4	3.2
95% CI	2.1-3.2	2.4-4.1	1.8-2.8	1.5-2.8	1.9-3.3	2.3-4.2	1.8-5.8	2.1-3.2	1.3-4.8	2.0-3.0	2.1-4.8
95% PI	1.3-5.2	1.5-6.4	1.1-4.5	0.94-4.4	1.2-5.2	1.5-6.5	1.3-8.1	1.3-5.2	0.94-6.6	1.3-4.4	1.5-6.4
N	20,553	8,391	12,162	4,229	6,854	7,104	2,366	18,334	2,219	17,611	2,942
Tau (τ)	0.32	0.32			0.33			0.32		0.26	
PEGeD algorithm (Geneva), %	2.6	3.2	2.3	2.0	2.8	3.1	2.9	2.6	2.1	2.5	2.6
95% CI	2.1-3.2	2.5-4.1	1.8-2.9	1.5-2.8	2.2-3.6	2.3-4.2	1.5-5.6	2.2-3.3	0.86-5.1	2.1-3.1	1.5-4.3
95% PI	1.3-5.2	1.6-6.4	1.1-4.6	0.94-4.3	1.3-5.7	1.5-6.5	1.1-7.6	1.3-5.3	0.64-6.7	1.4-4.5	1.2-5.6
N	20,553	8,391	12,162	4,229	6,854	7,104	2,366	18,334	2,219	17,611	2,942
Tau (τ)	0.32	0.32			0.32			0.32		0.26	

CDR: clinical decision rule; PE: pulmonary embolism; CI: confidence interval; PI = prediction interval; N: number of patients included in the analysis; VTE = venous thromboembolism; NE: not estimable

* Defined as the predicted probability of VTE in patients with a low score on the CDR combined with a negative D-dimer test result.

~ No failures were observed in patients with a low score on the CDR combined with a negative D-dimer test result.