The diagnostic management of suspected pulmonary embolism in special patient populations

Stals, M.A.M.

5

Performance of the 4-Level Pulmonary Embolism Clinical Probability Score (4PEPS) in the diagnostic management of pulmonary embolism: an external validation study

Milou A.M. Stals, Ludo F.M. Beenen, Michiel Coppens, Laura M. Faber, Herman M.A. Hofstee, Marcel M.C. Hovens, Menno V. Huisman, Tom van der Hulle, Karin A.H. Kaasjager, Marieke J.H.A. Kruip, Albert T.A. Mairuhu, Saskia Middeldorp, Marije ten Wolde, Frederikus A. Klok, Nick van Es, on behalf of the YEARS study group

Submitted

ABSTRACT

Background: The recently published 4-level Pulmonary Embolism Clinical Probability Score (4PEPS) integrates different aspects from currently available diagnostic strategies to further reduce imaging testing in patients with clinically suspected pulmonary embolism (PE).

Aim: To externally validate the performance of 4PEPS in an independent cohort.

Methods: In this post-hoc analysis of the prospective diagnostic management YEARS study, the primary outcome measures were discrimination, calibration, efficiency (proportion of imaging tests potentially avoided), and failure rate (venous thromboembolism (VTE) diagnosis at baseline or follow-up in patients with a negative 4PEPS algorithm). Multiple imputation was used for missing 4PEPS items. Based on 4PEPS, PE was considered ruled out in patients with a very low clinical pre-test probability (CPTP) without D-dimer testing, in patients with a low CPTP and D-dimer <1000 µg/L, and in patients with a moderate CPP and D-dimer below the age-adjusted threshold.

Results: Of the 3,465 patients, 474 (14%) were diagnosed with VTE at baseline or during 3-month follow-up. Discriminatory performance of the 4PEPS items was good (area under ROC-curve, 0.82; 95%CI, 0.80-0.84) as was calibration. Based on 4PEPS, PE could be considered ruled out without imaging in 58% (95%CI 57-60) of patients (efficiency), for an overall failure rate of 1.3% (95%CI 0.86-1.9). Compared to the YEARS algorithm, efficiency was higher (58% vs. 48%), as was the failure rate (1.3% vs. 0.42%).

Conclusion: In this retrospective external validation, 4PEPS appeared to safely rule out PE, with a higher efficiency than the originally used YEARS algorithm but at the cost of a 3-fold higher diagnostic failure rate.

INTRODUCTION

Correctly diagnosing pulmonary embolism (PE) is challenging as signs and symptoms of PE are not specific. Therefore, imaging tests are required to confirm the diagnosis.[1] However, the proportion of patients with confirmed PE among those with suspected PE is low (10-20%) and is decreasing steadily over recent decades.[2,3] Overtesting with imaging of patients with suspected PE can lead to unnecessary risks of radiation exposure and contrast medium induced reactions, but also to overdiagnosis of isolated small subsegmental PE, higher healthcare costs, and longer turnaround times in busy clinics.[1,4-7]

To reduce the number of imaging tests, the diagnostic management of suspected PE has evolved considerably over the past decades. Currently recommended diagnostic strategies for ruling out PE without imaging usually consist of standardized assessment of the clinical pre-test probability (CPTP) with validated clinical decision rules, e.g. the Wells rule, the revised Geneva score and the YEARS algorithm, in combination with D-dimer testing.[1] The combination of a non-high clinical probability and a D-dimer below the prespecified threshold safely rules out PE without imaging.[8-10] Since the specificity of D-dimer testing is low, modern strategies use D-dimer thresholds dependent on age or CPTP rather than a fixed threshold[10-14], which has decreased the need for imaging from about 70% to 40-50%.[10-14]

Recently, the 4-Level Pulmonary Embolism Clinical Probability Score (4PEPS) was developed with the aim to further decrease the need for imaging in patients with clinically suspected PE.[15] This score integrates different aspects from currently available diagnostic strategies, including the identification of very low risk patients in whom D-dimer testing can be withheld (as with the Pulmonary Embolism Rule-out Criteria (PERC) rule[16]) and the use of a CPTP-dependent D-dimer threshold (as with the YEARS algorithm). The derivation and validation study of 4PEPS, which was based on post-hoc analyses of large management studies, showed that the use of 4PEPS can lead to a substantial and safe reduction in imaging tests in patients with suspected PE.[15] However, a formal prospective management outcome study is lacking. We set out to externally validate the diagnostic performance of the 4PEPS strategy in an independent dataset by performing a post-hoc analysis of the YEARS study.[13]

METHODS

Patients and setting

The current study was a post-hoc analysis of the YEARS study[13], a prospective management study evaluating the YEARS algorithm in 3,465 patients with suspected PE. In the YEARS study, consecutive outpatients and inpatients with clinically suspected PE were included between 2013 and 2015 in twelve Dutch hospitals. Exclusion criteria were treatment with a therapeutic-dose anticoagulation initiated 24 hours or more before eligibility assessment, life expectancy less than 3 months, geographic inaccessibility precluding follow-up, pregnancy, allergy to intravenous contrast medium, and hemodynamic instability. The YEARS score, which consists of three clinical items (clinical signs of deep-vein thrombosis, hemoptysis, and clinical judgement whether PE is the most likely diagnosis), was calculated in all patients and combined with simultaneous assessment of D-dimer levels. D-dimer concentrations were measured with automated well validated high-sensitive quantitative D-dimer assays. According to the YEARS algorithm, PE was considered ruled out without imaging in patients with no YEARS items and a D-dimer level <1000 µg/L and in patients with one or more of the YEARS items and a D-dimer level <500 µg/L. All other patients were referred for CTPA to confirm or rule out the diagnosis of PE. Therapeutic anticoagulation was initiated in patients with confirmed PE, whereas patients with a negative diagnostic work-up were left untreated and followed for 3 months to evaluate the occurrence of symptomatic VTE. Suspicion of VTE during follow-up had to be confirmed by objective imaging tests or, in the case of death, by autopsy, by objective testing before death, or if PE could not be confidently excluded as a cause of death. An independent adjudication committee evaluated all episodes of suspected VTE and deaths during follow-up. based on the decision of an independent adjudication committee. For this post-hoc analysis, all 3,465 patients from the YEARS study were eligible for inclusion.

Study objective and outcomes

The primary aim of this study was to externally validate the discriminatory performance, calibration, safety, and efficiency of the 4PEPS in the diagnostic management of suspected PE. Safety was defined as the failure rate, which is the proportion of patients with confirmed VTE at baseline or during follow-up among those in whom PE was considered ruled out at baseline based on the strategy alone (as a measure of missed VTE events at baseline). 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 based on the following formula: 1.82+0.0053*prevalence (in %).[3] Based on a prevalence of 14% in the YEARS study, the accepted failure rate margin would be 1.89%. Efficiency was defined as the proportion of patients in whom PE would have been ruled out at baseline without imaging.

Study algorithm

The 4PEPS strategy was applied in this study as in the original study paper.[15] The 4PEPS sum score was calculated in all patients based on the following scoring items that were prospectively collected data within the YEARS study: age (<50 years: -2 points; 50-64 years: -1 point), chronic respiratory disease (-1 point), heart rate <80 beats per minute (-1 point), chest pain and acute dyspnea (+1 point), male sex (+2 points), hormonal estrogenic treatment (+2 points), personal history of VTE (+2 points), syncope (+2 points), immobility within the last 4 weeks (+2 points), pulse oxygen saturation <95% (+3 points), calf pain and/or unilateral lower limb edema (+3 points), and PE is the most likely diagnosis (+5 points). Patients were subsequently classified as having a very low clinical probability (CPP; 4PEPS <0 points), low CPP (4PEPS 0-5 points), moderate CPP (4PEPS 6-12 points), or high CPP (4PEPS >12 points). PE was considered ruled out in patients with a very low CPP without D-dimer testing, in patients with a low CPP and a D-dimer <1000 µg/L, and in patients with a moderate CPP and a D-dimer below an age-adjusted threshold (i.e. age times 10 µg/L in those older than 50 years). Patients with a high CPP and/or abnormal D-dimer test were considered to require imaging to confirm or rule out the diagnosis of PE (**Table 1**).

Table 1. 4-Level Pulmonary Embolism Clinical Probability Score (4PEPS)

Variables of 4PEPS	Points	Corresponding variables in YEARS
Age		Age
<50	-2	
50-64	-1	
Chronic respiratory disease	-1	Known COPD disease
Heart rate <80 beats per minute	-1	Heart frequency
Chest pain and acute dyspnea	1	Dyspnea / PainResp / Pain
Male	2	Sex
Hormonal estrogenic treatment	2	BLhormones (BL stands for baseline)
Personal history of VTE	2	Prior history of VTE
Syncope	2	Syncope or near collaps
Immobility within the last 4 wk»	2	Immobility
Pulse oxygen saturation <95%	3	Saturation
Calf pain and/or unilateral lower limb edema	3	Clinical signs of DVT (YEARS item)
PE is the most likely diagnosis	5	PE most likely diagnosis (YEARS item)

VTE: venous thromboembolism; wk: weeks; PE: pulmonary embolism; COPD: chronic obstructive pulmonary disease; DVT: deep-vein thrombosis

Four levels of CPP:

1. Very low CPP (<2%; <0 points), allowing exclusion of PE on clinical criteria only (thus without a D-dimer).
2. Low CPP (2-20%; 0-5 points), allowing exclusion of PE with a D-dimer level <1000 µg/L.
3. Moderate CPP (20-65%; 6-12 points), allowing exclusion of PE with a D-dimer level less than the age-adjusted cutoff value (<500 µg/L in patients <50 years old and the patient's age times 10 µg/L in patients ≥50 years old).
4. High CPP (>65%; >12points), not allowing a safe exclusion of PE with D-dimer testing and requiring imaging testing, without preceding of the D-dimer test.

Statistical analysis

Patient characteristics were described using standard descriptive statistics. Missing 4PEPS variables were imputed twenty times using multiple imputation by chained equations (MICE) assuming a missing at random pattern. This pattern, unlike missing completely at random (MCAR) or missing not at random (MNAR), implies that missingness depends on observed variables for which imputation techniques can be used. Baseline information as well as outcome data were included in the imputation model. Rubin's rule was used to pool data across the imputed datasets. We also performed a complete case analysis.

Discriminatory performance of the 4PEPS, both with and without D-dimer testing, was evaluated by the area under the receiver operating characteristic (ROC)-curve (AUC) with 95% confidence intervals (CI) based on DeLong's method. We considered an AUC less than 0.60 as very poor, 0.60 to 0.69 as poor, 0.70 to 0.79 as fair, 0.80 to 0.89 as good, and more than or equal to 0.90 as excellent discrimination.[17] In addition, we performed a multivariable logistic regression model with the 4PEPS variables, with and without (categorical) D-dimer levels, as independent variables, and a diagnosis of VTE at baseline or during follow-up as the dependent variable. Odds ratios with 95% CIs were compared with the odds ratios reported in the original 4PEPS study paper.[15] Calibration was evaluated by comparing the estimated VTE probabilities based on the model with the observed proportion of VTE in a calibration plot using loess regression. In all analyses, patients lost to follow-up were excluded.

Estimates of the failure rate and efficiency with 95% CI were calculated by using the Clopper-Pearson method. We first determined the proportion of patients in whom PE would be considered ruled out without imaging, based on the different categories of the 4PEPS strategy (efficiency). We then calculated the diagnostic failure rates in patients managed without CTPA. Patients who received anticoagulation for indications other than VTE during follow-up or who were lost to follow-up were excluded from the failure rate analysis to be conservative. Safety and efficiency were calculated overall, separately for the four levels of the 4PEPS, and in the following subgroups: patients with cancer, patients ≥ 50 years of age, patients ≥ 75 years of age, patients with a history of VTE, and inpatients. Performance of the 4PEPS was compared to the performance of the originally applied YEARS algorithm by calculating the difference in efficiency and failure rate with 95% confidence intervals based on 250 bootstrap samples.

As a sensitivity analysis, a complete case analysis was performed by excluding patients with missing 4PEPS variables.

SPSS Statistics version 25.0 and R version 4.0.3 were used for data analysis.

Role of the Funding source

No funding was received to perform this study.

RESULTS

Patient characteristics and study outcomes

All 3,465 patients from the original YEARS study were included in the present post-hoc analysis. The mean age was 53 years (standard deviation (SD) 18), 38% of patients were male, and 87% were outpatients (**Table 2**). The median D-dimer level was 670 µg/L (interquartile range (IQR) 335-1500 µg/L). The 4PEPS individual scoring items were complete in a total of 1409 patients (41%), while one or more missing 4PEPS scoring items were imputed in the other 59%. Most missing values were encountered within the 4PEPS items of 'syncope' (missing in 57% of the patients), 'pulse oxygen saturation' (in 46% of the patients), and 'chest pain and dyspnea' (in 44% of the patients). 459 patients were diagnosed with PE at baseline (13%) and 15 (0.43%) were diagnosed with VTE during the 3-months follow-up period, resulting in an overall PE prevalence of 14%.

4PEPS without D-dimer testing

In the multiply imputed dataset, patients had a 4PEPS sum score between -4 and 18 points, with a median of 7 points (IQR, 1-13). Discriminatory performance of the 4PEPS (without D-dimer testing) was good, with an AUC of 0.82 (95%CI 0.80-0.84; **Figure 1**). The odds ratios from the individual 4PEPS variables in this study were in general comparable to the odds ratios reported in the original 4PEPS paper, except for the variables 'chronic respiratory disease' and 'calf pain and/or unilateral limb edema' which were respectively 0.25 (0.57 in original paper) and 8.9 (2.7 in original paper) (**Table 3**). Overall PE prevalence was higher with increasing 4PEPS sum scores (**Figure 2**) ranging from 0% in patients with -4 or -3 points to 100% in patients with 17 or 18 points. Prevalence of PE also increased with higher 4PEPS CPTP levels (**Figure 4**). The calibration plot (**Figure 3**) showed overall good agreement between the estimated probabilities based on the 4PEPS model and the prevalence of PE in the overall range of 0-100% (slope 1.05; 95% CI, 0.95-1.15). In the clinically relevant range of probabilities from 0-10%, the 4PEPS slightly underestimated the risk of PE. The complete case analysis showed consistent results (**Appendix Figure 1-4 and Table 3**).

Table 2. Baseline characteristics of the complete study group

Characteristics	**	Missing (%)
Participants, n	3,465	NA
Age, y, mean (SD)	53 (18)	0 (0)
Active cancer, n (%)	336 (9.7)	5 (0.1)
Outpatients, n (%)	2995 (87)	1 (0.0)
4PEPS variables:		
Age <50, n (%)	1448 (42)	0 (0)
50-64, n (%)	973 (28)	0 (0)
Chronic respiratory disease, n (%)	423 (12)	0 (0)
Heart rate <80 beats per minute, n (%)	1186 (35)	66 (1.9)
Chest pain and acute dyspnea, n (%)	896 (47)	1537 (44)
Male, n (%)	1311 (38)	0 (0)
Hormonal estrogenic treatment, n (%)	337 (9.8)	35 (1.0)
Personal history of VTE, n (%)	359 (10)	2 (0.1)
Syncope, n (%)	104 (6.9)	1966 (57)
Immobility within the last 4 wk, n (%)	407 (12)	5 (0.1)
Pulse oxygen saturation <95%, n (%)	373 (20)	1583 (46)
Calf pain and/or unilateral lower limb edema, n (%)	112 (3.2)	0 (0)
PE is the most likely diagnosis, n (%)	1625 (47)	0 (0)
4PEPS classification:		
-Very low CPP (<0 points), n (%) ^	256/1,409 (18)	2056 (59)
-Low CPP (0-5 points), n (%) ^	699/1,409 (50)	2056 (59)
-Moderate CPP (6-12 points), n (%) ^	443/1,409 (31)	2056 (59)
-High CPP (>12 points), n (%) ^	11/1,409 (0.8)	2056 (59)
D-dimer, µg/L, median (IQR)	670 (335-1500)	12 (0.3)
-D-dimer level between 0 µg/L to age-adjusted value, n (%)	1490 (43)	12 (0.3)
-D-dimer level between age-adjusted value to 1000 µg/L, n (%)	685 (20)	12 (0.3)
-D-dimer level ≥ 1000 µg/L), n (%)	1255 (36)	12 (0.3)
PE prevalence, n (%)	474 (14)	0 (0)

n: number; y: years; SD: standard deviation; VTE: venous thromboembolism; wk: weeks; PE: pulmonary embolism; CPP: clinical probability; IQR: interquartile range; NA: not applicable

** Percentage was calculated by dividing the number of patients by the total number of patients in the study group minus number of missing values

^ Percentage was calculated by dividing the number of patients by 1409 (total number of patients in whom 4PEPS classification could be calculated)

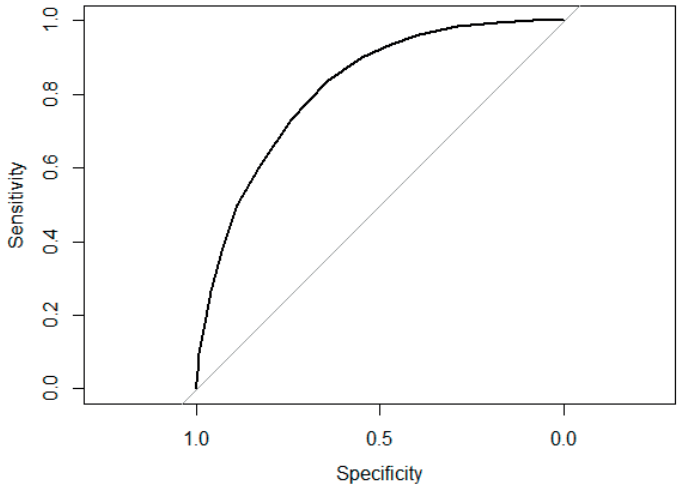


Figure 1. Receiver operating characteristic curve of 4PEPS without D-dimer testing*
AUC: 0.82 (95%CI 0.80-0.84)
*After multiple imputation

Table 3. Regression model 4PEPS without and with D-dimer testing*

4 PEPS items	Univariable current study OR (95% CI)	Multivariable original study OR	Multivariable current study (without D-dimer) OR (95% CI)	Multivariable current study (with D-dimer) OR (95% CI)
Age, y				
<50	0.42 (0.33-0.53)	0.37	0.37 (0.27-0.50)	0.70 (0.50-0.98)
50-64	0.78 (0.62-0.98)	0.52	0.80 (0.61-1.05)	1.1 (0.83-1.5)
Chronic respiratory disease	0.36 (0.24-0.55)	0.57	0.25 (0.16-0.41)	0.30 (0.19-0.50)
Heart rate <80 bpm	0.64 (0.51-0.79)	0.67	0.66 (0.51-0.86)	0.79 (0.60-1.0)
Chest pain and acute dyspnea	1.1 (0.84-1.5)	1.3	1.3 (0.90-1.8)	1.2 (0.84-1.7)
Male	1.5 (1.2-1.8)	1.6	1.6 (1.3-2.1)	1.5 (1.2-1.9)
Hormonal estrogenic treatment	1.2 (0.85-1.6)	1.8	2.4 (1.6-3.6)	2.1 (1.3-3.3)
Personal history of VTE	3.3 (2.5-4.2)	2.0	3.1 (2.3-4.1)	3.2 (2.3-4.5)
Syncope	0.90 (0.51-1.6)	1.7	0.90 (0.48-1.7)	0.81 (0.41-1.6)
Immobility within the last 4wk	3.4 (2.6-4.3)	1.5	2.3 (1.8-3.1)	1.6 (1.2-2.2)
Pulse oxygen saturation <95%	2.1 (1.6-2.7)	2.3	2.0 (1.5-2.8)	1.8 (1.2-2.5)
Calf pain and/or unilateral limb edema	12 (8.2-18)	2.7	8.9 (5.6-14)	6.3 (3.8-10)
PE is the most likely diagnosis	8.1 (6.2-10)	6.4	6.0 (4.6-8.0)	4.2 (3.1-5.7)
D-dimer (in categories: 1) 0 µg/L to age-adjusted; 2) age-adjusted to 1000 µg/L and 3) ≥ 1000 µg/L)	2: 14 (6.4-29) 3: 92 (46-187)	- -	- -	2: 8.1 (3.7-17) 3: 48 (23-99)

y: years; bpm: beats per minute; VTE: venous thromboembolism; wk: weeks; PE: pulmonary embolism; n: number; OR: odds ratio; CI: confidence interval
*After multiple imputation

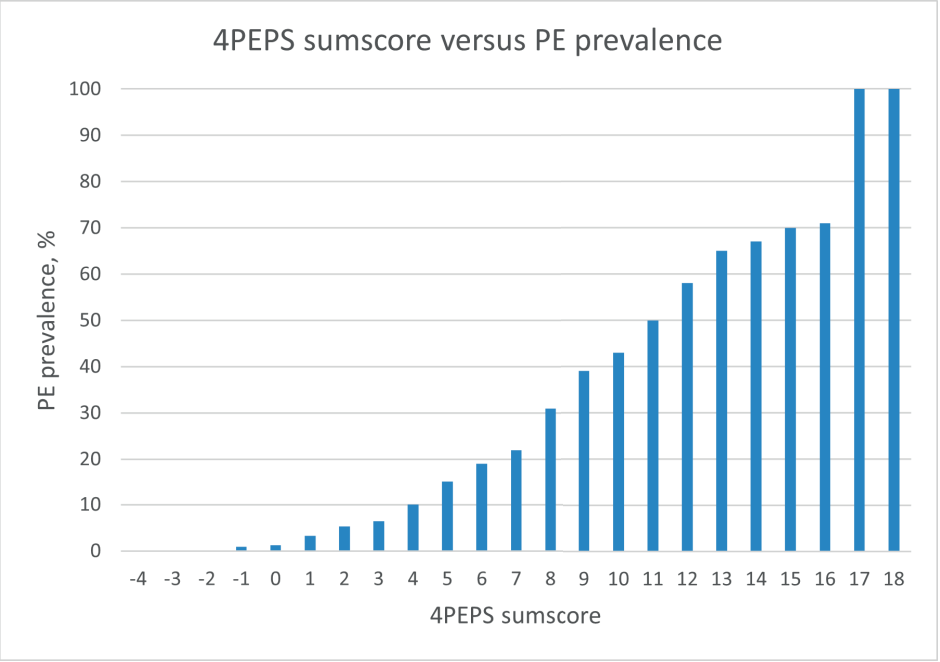


Figure 2. 4PEPS sumscore versus prevalence of PE*****
The proportion of patients in the 4PEPS sumscore groups of -4 to 18 points was as follows:
-4 points: 0.058%; -3 points: 1.7%; -2 points: 5.4%; -1 points: 8.6%; 0 points: 9.5%; 1 points: 9.4%; 2 points: 6.6%; 3 points: 7.7%; 4 points: 9.0%; 5 points: 9.8%; 6 points: 9.2%; 7 points: 6.4%; 8 points: 5.6%; 9 points: 3.8%; 10 points: 3.3%; 11 points: 1.8%; 12 points: 1.2%; 13 points: 0.63%; 14 points: 0.21%; 15 points: 0.08%; 16 points: 0.02%; 17 points: 0.02%; 18 points: 0.001%
*After multiple imputation

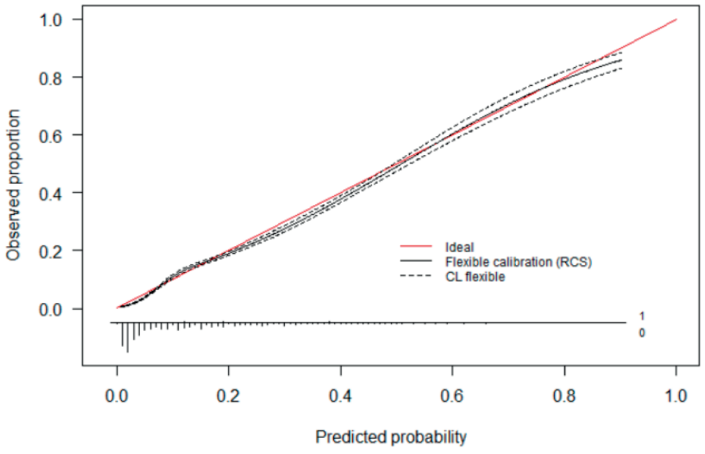


Figure 3. Calibration plot*
Legend: the red line describes the ideal correlation between predicted probabilities and observed proportion of VTE, while the black line describes the correlation between predicted probabilities and observed proportion of VTE based on the 4PEPS model in our study.
*After multiple imputation

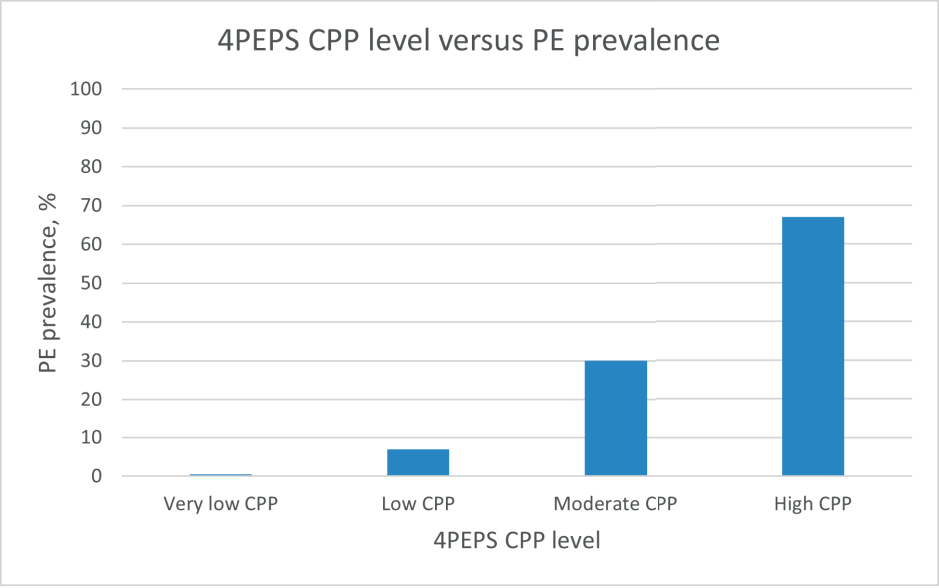


Figure 4. 4PEPS CPP level versus prevalence of PE*
CPP: clinical pre-test probability; PE: pulmonary embolism
*After multiple imputation

4PEPS in combination with D-dimer testing

When patients were retrospectively classified by the 4PEPS strategy, 16% were defined as having a very low CPP (4PEPS <0 points), 52% as having a low CPP (4PEPS 0-5 points), 31% as having a moderate CPP (4PEPS 6-12 points), and 1% as having a high CPP (4PEPS >12 points). With 4PEPS, PE could be excluded without the use of imaging in 58% (95%CI 57-60) of the patients (efficiency). The overall 3-month failure rate in patients in whom PE was considered ruled out without imaging based on 4PEPS was 1.3% (95%CI 0.86-1.9). Failure rates were higher in patients with cancer, aged ≥50 years, aged ≥75 years, and with a history of VTE (**Table 4**). In these subgroups, the proportion of patients that could be ruled out from having PE without imaging was decreased by 9 to 32%. Based on the YEARS algorithm, PE was ruled out without imaging in 48% (95%CI 46-49) of the patients, with an overall 3-months failure rate of 0.42% (95%CI 0.20-0.89), which was not statistically different compared to the performance of 4PEPS (**Table 5**).

The complete case analysis yielded slightly higher point estimates for the failure rate of the 4PEPS strategy, while point estimates for efficiency were comparable (**Appendix Table 3 and 4**). In this analysis, 14 of the 825 patients (1.7%) classified as 'PE ruled out' based on 4PEPS (thus not requiring imaging) received a VTE diagnosis at baseline (n=12) or during follow-up (n=2). Of the patients classified as 'PE ruled out' by the YEARS algorithm, 2 of the 692 patients (0.29%) were diagnosed with VTE at baseline (n=1) or during

follow-up (n=1). The diagnosis in these two patients was also missed by 4PEPS. Of the remaining twelve patients with a missed diagnosis (failures) based on 4PEPS, 1 patient had a DVT and 11 patients had PE, of which 9 segmental PEs and 2 isolated subsegmental PEs. A comparison of the baseline characteristics of patients in the complete case analyses versus patients in whom one or more 4PEPS items were missing is presented in **Appendix Table 5**.

Table 4. Failure rate and efficiency 4PEPS overall and across different subgroups *

	Overall	Very low CPP	Low CPP	Moderate CPP	High CPP
Failure rate, % (95% CI)	1.3 (0.86-1.9)	0.50 (0.12-2.0)	1.5 (0.94-2.4)	1.9 (0.72-5.0)	NA
Efficiency, % (95% CI)	58 (57-60)	100	70 (68-72)	20 (18-23)	0

	No malignancy	Malignancy	Aged <50 years	Aged ≥ 50 years	Aged <75 years	Aged ≥ 75 years
Failure rate, % (95% CI)	1.2 (0.78-1.8)	3.1 (0.91-10)	0.88 (0.45-1.7)	1.8 (1.1-2.9)	1.3 (0.81-1.9)	1.7 (0.48-6.0)
Efficiency, % (95% CI)	62 (60-63)	30 (25-35)	73 (71-76)	48 (46-50)	63 (61-64)	34 (29-38)

	No history of VTE	History of VTE	Outpatients	Inpatients
Failure rate, % (95% CI)	1.3 (0.83-2.0)	1.5 (0.38-5.7)	1.3 (0.85-2.0)	1.1 (0.29-4.3)
Efficiency, % (95% CI)	60 (59-62)	42 (37-47)	60 (58-61)	51 (46-56)

CI: confidence interval; pts: patients; CPP: clinical probability; NA: not applicable/available; VTE: venous thromboembolism
*After multiple imputation

Table 5. Failure rate and efficiency 4PEPS compared to YEARS diagnostic strategy*

	4PEPS	YEARS	Absolute difference
Failure rate, % (95% CI)	1.3 (0.86-1.9)	0.42 (0.2-0.89)	0.87 (-8.4 ; 10)
Efficiency, % (95% CI)	58 (57-60)	48 (46-49)	11 (-3.6 ; 25)
NNT	10	NA	NA
NNH	114	NA	NA

CI: confidence interval; pts: patients; NNT: number needed to prevent one CT-scan; NNH: number needed to miss a PE diagnosis
*After multiple imputation

DISCUSSION

In this post-hoc analysis of the prospective diagnostic management YEARS study, the newly derived 4PEPS diagnostic strategy for ruling out PE was externally validated. Based on this strategy, PE would have been ruled out without imaging in 58% (95%CI 57-60) of patients with an overall failure rate of 1.3% (95%CI 0.86-1.9). Compared to the YEARS algorithm, efficiency was higher (58% vs. 48%), as was the failure rate (1.3% vs. 0.42%).

Compared to currently used algorithms that use a specific strategy of D-dimer testing, the 4PEPS strategy integrates different aspects from currently available diagnostic strategies. 4PEPS identifies very low risk patients in whom PE is ruled out without D-dimer testing (similar to PERC), low risk patients in whom PE is ruled out based on a D-dimer <1000 µg/L (similar to YEARS), and moderate risk patients in whom PE is ruled out based on a D-dimer below the age-adjusted D-dimer threshold (similar to ADJUST-PE).[11,13,16] This new strategy has the potential to reduce the need for imaging tests at an acceptably low diagnostic failure rate.[15] In the derivation and validation study of 4PEPS, the first external validation cohort, with a PE prevalence of 22%, showed an efficiency of 54% and an overall failure rate of 0.71% for 4PEPS. In the second external validation cohort, with a PE prevalence of 12%, efficiency was 68% and the overall failure rate 0.89%.[15] In the present study, with a PE prevalence of 14%, the efficiency of 4PEPS (58%) was generally in line with the one reported in the 4PEPS derivation and validation study. However, we observed a higher failure rate (1.3%), which was three-fold higher than that of the originally used YEARS algorithm. Moreover, as the 4PEPS score includes 12 items in combination with different D-dimer thresholds, its complexity could hamper adherence to the strategy in busy clinics. Computer or smartphone applications could maybe (partially) overcome this problem in the nearby future, but are not available yet.

To differentiate the failure rate of the 4PEPS strategy from the failure rate of the originally used YEARS algorithm, we outlined the failures of both strategies, also defined as ‘missed VTE events’. In the complete case analysis, twelve more patients would have been missed with the 4PEPS strategy, in comparison with the YEARS algorithm. These patients with ‘missed VTE events’ were probably not identified by the 4PEPS strategy as PE can still be ruled out by 4PEPS without imaging in patients that score the item of PE most likely diagnosis (5 points) but have a D-dimer <1000 µg/L, while the D-dimer threshold is 500 µg/L for these patients within the YEARS algorithm.[13,15] Indeed, from the twelve patients, ten patients scored the item of PE most likely diagnosis and all ten had a D-dimer level between 500 and 1000 µg/L. As a result, these patients were indicated as high risk within the YEARS algorithm and referred for imaging. Given the retrospective

design of the present analysis, the clinical course of these 'missed events' if imaging had not been performed is unknown.

Strengths of our study include the large sample size with more than 3,500 patients and the calculation of the 4PEPS based on prospectively collected data within the YEARS diagnostic management study. Other strengths include the near complete follow-up and independent adjudication of VTE events and deaths within the YEARS study.

Our study also has limitations. The most important limitation is that this external validation was performed retrospectively. Therefore, more patients received imaging than would have been the case when the 4PEPS strategy was applied in a prospective management study, potentially resulting in an overestimation of the failure rate.[18-20] Another limitation was that one or more 4PEPS items were missing in 59% of patients, and that the characteristics and prevalence of PE in these patients was different than that of patients in whom 4PEPS could be calculated. Therefore, to reduce the bias associated with missing data, we used multiple imputation based on a model including all baseline variables as well as the outcome, which is in line with statistical recommendations. We assumed a missing at random pattern, which may have been incorrect but cannot be compared statistically to a missing not at random pattern. Reassuringly, discrimination was comparable in the complete case analysis, although calibration was poor, possibly as a result of the difference in PE prevalence between the complete case and imputed datasets. In addition, there were small differences in the definitions of the 4PEPS variables and corresponding variables within the YEARS study, for instance corresponding variables for the 4PEPS items 'chronic respiratory disease' and 'chest pain and acute dyspnea' were 'known COPD disease' and 'Dyspnea and Pleuritic chest pain and/or Pain' in the YEARS database.

What are the clinical consequences of the present analysis? In our study, the 4PEPS strategy does not exceed the failure rate margin of 1.89%, as recommended by the International Society on Thrombosis and Haemostasis based on a prevalence of 14%[3], and confirms the efficiency of the 4PEPS strategy as imaging could have been withheld in 58% of patients with suspected PE. Nevertheless, as the observed failure rate in our analysis appeared to be higher than with YEARS, a formal prospective management study is needed before its use can be recommended by guidelines and integrated in clinical practice. The failure rate of 4PEPS may be lower in such a management study due to the verification bias in the present analysis, i.e. patients with a negative 4PEPS algorithm outcome having received imaging.

Various new diagnostic algorithms and strategies for suspected PE have been proposed over the past decade, including ADJUST, YEARS, PEGeD, and now 4PEPS, which all aim to provide a safe and efficient diagnostic strategy for clinically suspected PE. As a consequence, the decision which algorithm to use in practice has not become more simple, as performance of these algorithms is in part dependent on PE prevalence. Higher efficiency is almost inevitably accompanied by a higher failure rate, although this may include identification of less relevant smaller clots.[20] Physicians may let simplicity prevail or choose more complex algorithms that require calculators to avoid calculation or interpretation errors. However, such complexity could hamper adherence and thereby performance in busy clinics. The ultimate answer regarding accuracy may come from randomized diagnostic trials, although showing superiority of one strategy over the other will likely require a very large sample size.

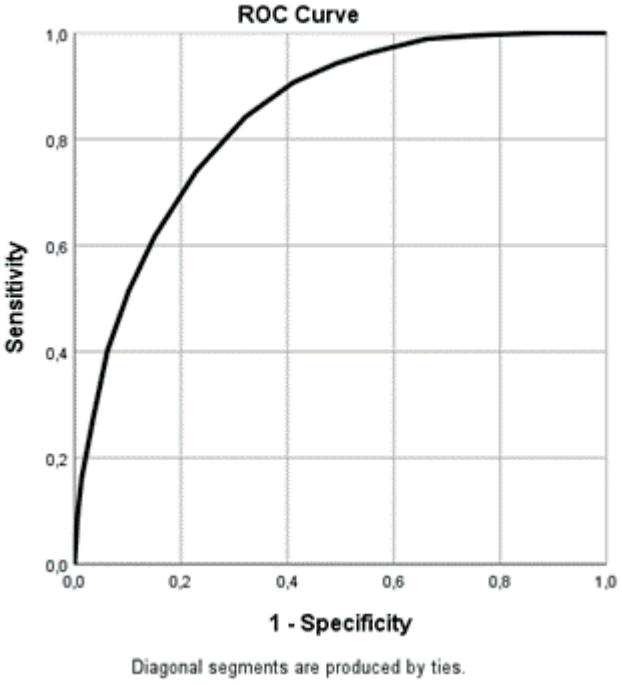
In summary, the 4PEPS strategy appeared to safely rule out PE, with a non-significantly 10% higher efficiency than the originally used YEARS algorithm but at the cost of a 3-fold higher diagnostic failure rate. External validation of the 4PEPS strategy in a prospective outcome study is needed.

REFERENCES

1. Huisman MV, Barco S, Cannegieter SC, et al. Pulmonary embolism. Nat Rev Dis Primers. 2018; 4: 18028.
2. Le Gal G, Bounameaux H. Diagnosing pulmonary embolism: running after the decreasing prevalence of cases among suspected patients. J Thromb Haemost. 2004; 2: 1244-6.
3. Dronkers CEA, van der Hulle T, Le Gal G, et al. Towards a tailored diagnostic standard for future diagnostic studies in pulmonary embolism: communication from the SSC of the ISTH. J Thromb Haemost. 2017; 15: 1040-3.
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. 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: 725-33.
7. van der Hulle T, Dronkers CE, Huisman MV, Klok FA. Current standings in diagnostic management of acute venous thromboembolism: Still rough around the edges. Blood reviews. 2016; 30: 21-6.
8. 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: 253-61.
9. 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.
10. Stals MAM, Takada T, Kraaijpoel N, et al. 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. Ann Intern Med. 2021. 10.7326/m21-2625.
11. 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.
12. 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.
13. van der Hulle T, Cheung WY, Kooij S, et al. Simplified diagnostic management of suspected pulmonary embolism (the YEARS study): a prospective, multicentre, cohort study. Lancet. 2017; 390: 289-97.
14. Kearon C, de Wit K, Parpia S, et al. Diagnosis of Pulmonary Embolism with d-Dimer Adjusted to Clinical Probability. N Engl J Med. 2019; 381: 2125-34.
15. Roy PM, Friou E, Germeau B, et al. Derivation and Validation of a 4-Level Clinical Pretest Probability Score for Suspected Pulmonary Embolism to Safely Decrease Imaging Testing. JAMA Cardiol. 2021; 6: 669-77.
16. Kline JA, Mitchell AM, Kabrhel C, et al. Clinical criteria to prevent unnecessary diagnostic testing in emergency department patients with suspected pulmonary embolism. J Thromb Haemost. 2004; 2: 1247-55.

17. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982; 143: 29-36.
18. Carrier M, Klok FA. Symptomatic subsegmental pulmonary embolism: to treat or not to treat? Hematology Am Soc Hematol Educ Program. 2017; 2017: 237-41.
19. 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: 1251-61.
20. 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: 629-35.

APPENDIX

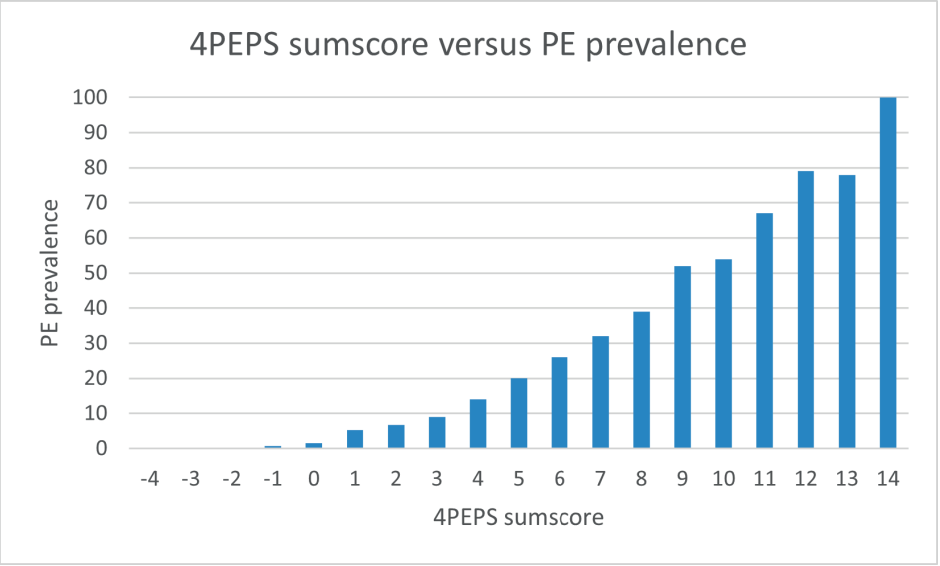


Appendix Figure 1. Receiver operating characteristic (ROC) curve 4PEPS score without D-dimer testing (according to complete case analysis)
AUC: 0.84 (95%CI 0.82-0.87)

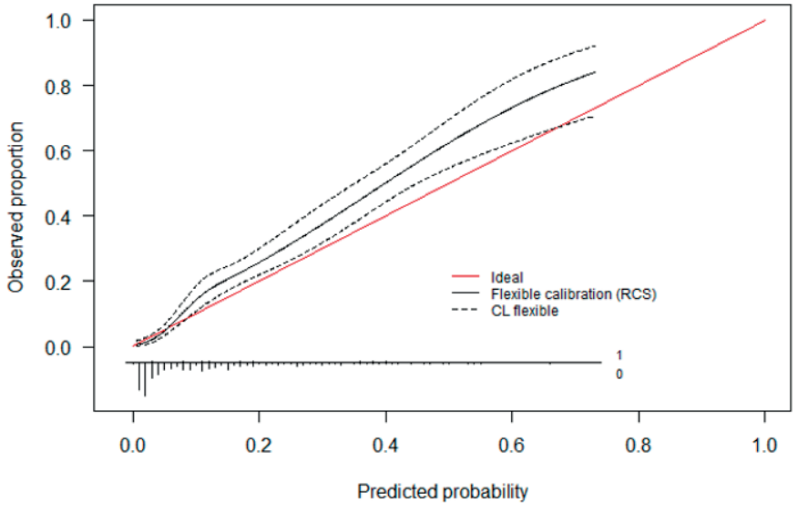
Appendix Table 2. Regression model 4PEPS score with and without D-dimer testing (according to complete case analysis)

4 PEPS items	Univariable current study	Multivariable original study	Multivariable current study (without D-dimer)	Multivariable current study (with D-dimer)
	OR (95% CI)	OR	OR (95% CI)	OR (95% CI)
Age, y				
<50	0.39 (0.28-0.55)	0.37	0.32 (0.20-0.49)	0.68 (0.41-1.1)
50-64	0.83 (0.60-1.2)	0.52	0.75 (0.51-1.1)	1.1 (0.68-1.7)
Chronic respiratory disease	0.41 (0.24-0.71)	0.57	0.30 (0.16-0.57)	0.38 (0.19-0.76)
Heart rate <80 bpm	0.72 (0.53-0.97)	0.67	0.80 (0.55-1.1)	1.0 (0.67-1.6)
Chest pain and acute dyspnea	1.3 (0.97-1.7)	1.3	1.5 (1.1-2.1)	1.4 (0.94-2.0)
Male	1.8 (1.4-2.4)	1.6	1.8 (1.3-2.5)	1.5 (1.1-2.3)
Hormonal estrogenic treatment	1.1 (0.73-1.8)	1.8	2.6 (1.4-4.8)	2.2 (1.0-4.5)
Personal history of VTE	3.0 (2.1-4.3)	2.0	2.4 (1.6-3.7)	2.4 (1.5-4.0)
Syncope	1.8 (1.1-2.8)	1.7	1.8 (1.0-3.2)	1.5 (0.78-2.7)
Immobility within the last 4wk	2.9 (2.1-4.2)	1.5	2.1 (1.4-3.1)	1.3 (0.81-2.0)
Pulse oxygen saturation <95%	2.4 (1.8-3.2)	2.3	2.2 (1.5-3.3)	1.8 (1.2-2.8)
Calf pain and/or unilateral limb edema	19 (10-37)	2.7	15 (7.3-33)	11 (4.8-26)
PE is the most likely diagnosis	10 (7.1-15)	6.4	7.2 (4.8-11)	5.4 (3.4-8.4)
D-dimer (in categories: 1) 0 µg/L to age-adjusted; 2) age-adjusted to 1000 µg/L and 3) ≥ 1000 µg/L)	-	-	-	2: 8.3 (2.7-25) 3: 67 (24-187)

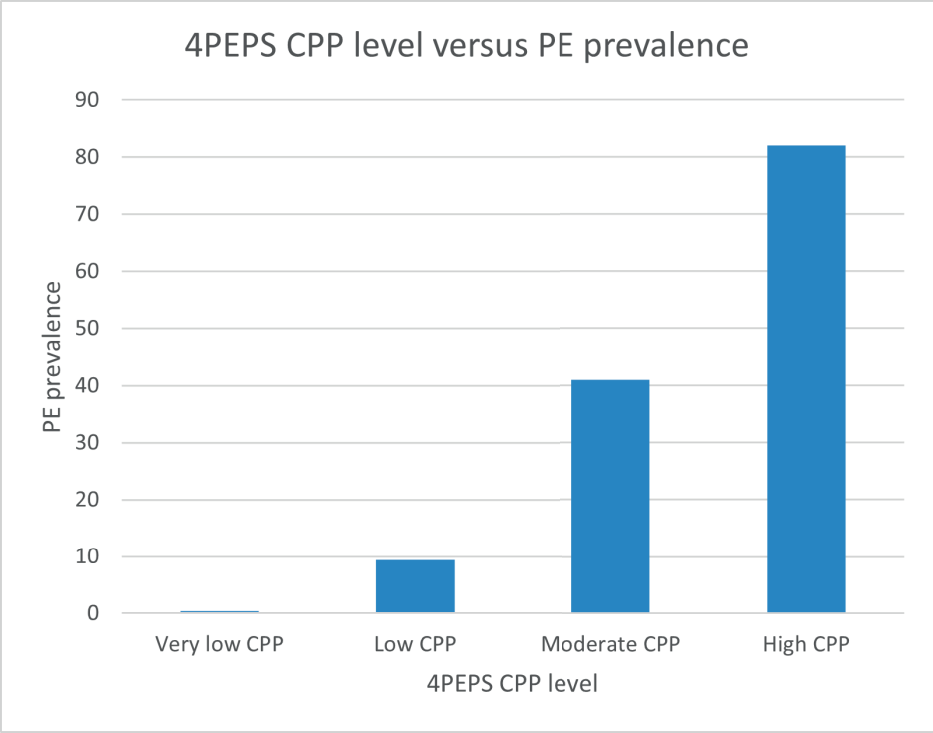
y: years; bpm: beats per minute; VTE: venous thromboembolism; wk: weeks; PE: pulmonary embolism; n: number; OR: odds ratio; CI: confidence interval



Appendix Figure 2. 4PEPS sumscore versus prevalence of PE (according to complete case analysis)



Appendix Figure 3. Calibration plot (according to complete case analysis)
The red line describes the ideal correlation between predicted probabilities and observed proportion of VTE, while the black line describes the correlation between predicted probabilities and observed proportion of VTE based on the 4PEPS model in our study.



Appendix Figure 4. 4PEPS CPP level versus prevalence of PE (according to complete case analysis)

Appendix Table 3. Failure rate and efficiency 4PEPS overall and across different subgroups (according to complete case analysis)

	Overall	Very low CPP	Low CPP	Moderate CPP	High CPP
Failure rate, % (95% CI) n/n	1.7 (1.0-2.9) 14/817	0.39 (0.01-2.4) 1/255	2.3 (1.2-4.1) 11/480	2.4 (0.15-9.0) 2/82	NA
Efficiency, % (95% CI) n/n	59 (56-61) 825/1409	100 (98-100) 256/256	69 (66-72) 483/699	19 (16-23) 86/443	0 0/11
Number of pts in analysis	1409	256	699	443	11

	No malignancy	Malignancy	Aged <50 years	Aged ≥ 50 years	Aged <75 years	Aged ≥ 75 years
Failure rate, % (95% CI) n/n	1.8 (1.0-3.0) 14/785	0 (0.0-14) 0/29	1.4 (0.56-3.1) 6/434	2.1 (0.99-4.1) 8/383	1.8 (1.0-3.0) 13/739	1.3 (0.01-7.6) 1/78
Efficiency, % (95% CI) n/n	62 (60-65) 793/1270	21 (15-29) 29/136	75 (71-78) 438/585	47 (44-50) 387/824	63 (60-66) 747/1182	34 (28-41) 78/227
Number of pts in analysis	1270	136	585	824	1182	227

	No history of VTE	History of VTE	Outpatients	Inpatients
Failure rate, % (95% CI) n/n	1.8 (0.97-2.9) 13/759	1.8 (0.01-10) 1/58	1.8 (0.98-3.0) 13/755	1.6 (0.01-9.4) 1/62
Efficiency, % (95% CI) n/n	61 (58-64) 764/1250	38 (31-46) 61/159	59 (56-61) 763/1301	57 (48-66) 62/108
Number of pts in analysis	1250	159	1301	108

CI: confidence interval; pts: patients; CPP: clinical probability; NA: not applicable/available; VTE: venous thromboembolism

Appendix Table 4. Failure rate and efficiency 4PEPS compared to YEARS diagnostic strategy (according to complete case)

	4PEPS	YEARS
Failure rate, % (95% CI) n/n	1.7 (1.0-2.9) 14/817	0.29 (0.01-1.1) 2/685
Efficiency, % (95% CI) n/n	59 (56-61) 825/1409	49 (47-52) 692/1409
NNT	10	NA
NNH	71	NA
Number of pts in analysis	1409	1409

CI: confidence interval; pts: patients; NNT: number needed to prevent one CT-scan; NNH: number needed to miss a PE diagnosis

Appendix Table 5. Comparison of the baseline characteristics of patients in the complete case analysis versus patients in whom one or more 4PEPS items were missing

	Complete case analysis		Other patients*		Comparing two groups
Characteristics	******	**Missing (%)**	******	**Missing (%)**	**p-value**
Participants, n	1409	NA	2056	NA	NA
Age, y, mean (SD)	54 (19)	0 (0)	53 (18)	0 (0)	0.54
Active cancer, n (%)	136 (9.7)	3 (0.2)	200 (9.7)	2 (0.1)	0.95
Outpatients, n (%)	1301 (92)	0 (0)	1694 (82)	1 (0.05)	0.00
Duration of symptoms in days, median (IQR)	3 (1-9)	3 (0.2)	3 (1-7)	22 (1.1)	0.035
Active smoking, n (%)	322 (24)	36 (2.6)	508 (26)	81 (3.9)	0.14
History of rheumatic or auto-immune disorder, n (%)	67 (7.9)	558 (40)	53 (9.7)	1510 (73)	0.23
On antiplatelet treatment at time of presentation, n (%)	140 (16)	550 (39)	96 (18)	1519 (74)	0.44
Hemoptysis, n (%)	48 (3.4)	0 (0)	89 (4.3)	0 (0)	0.17
Renal insufficiency (GFR <30 ml/min) at presentation, n (%)	14 (1.0)	24 (1.7)	35 (1.8)	81 (3.9)	0.07
C-reactive protein level at presentation, mg/L, median (IQR)	8 (3-32)	49 (3.5)	8 (2-28)	161 (7.8)	0.00
4PEPS variables:					
Age <50, n (%) 50-64, n (%)	585 (42) 387 (28)	0 (0)	863 (42) 586 (29)	0 (0)	0.79 0.53
Chronic respiratory disease, n (%)	174 (12)	0 (0)	249 (12)	0 (0)	0.83
Heart rate <80 beats per minute, n (%)	460 (33)	0 (0)	726 (37)	66 (3.2)	0.02
Chest pain and acute dyspnea, n (%)	656 (47)	0 (0)	240 (46)	1537 (75)	0.90
Male, n (%)	520 (37)	0 (0)	791 (39)	0 (0)	0.35
Hormonal estrogenic treatment, n (%)	134 (9.5)	0 (0)	203 (10)	35 (1.7)	0.61
Personal history of VTE, n (%)	159 (11)	0 (0)	200 (9.7)	2 (0.1)	0.14
Syncope, n (%)	99 (7.0)	0 (0)	5 (5.6)	1966 (96)	0.60
Immobility within the last 4 wk, n (%)	168 (12)	0 (0)	239 (12)	5 (0.2)	0.81
Pulse oxygen saturation <95%, n (%)	274 (19)	0 (0)	99 (21)	1583(77)	0.48
Calf pain and/or unilateral lower limb edema, n (%)	56 (4.0)	0 (0)	56 (2.7)	0 (0)	0.041
PE is the most likely diagnosis, n (%)	681 (48)	0 (0)	944 (46)	0 (0)	0.16

Appendix Table 5. Comparison of the baseline characteristics of patients in the complete case analysis versus patients in whom one or more 4PEPS items were missing (*continued*)

	Complete case analysis		Other patients*		Comparing two groups
Characteristics	****	**Missing (%)**	****	**Missing (%)**	**p-value**
4PEPS classification:					
-Very low CPP (<0 points), n (%)	256 (18)	0 (0)	NA	2056 (100)	NA
-Low CPP (0-5 points), n (%)	699 (50)	0 (0)	NA	2056 (100)	NA
-Moderate CPP (6-12 points), n (%)	443 (31)	0 (0)	NA	2056 (100)	NA
-High CPP (>12 points), n (%)	11 (0.8)	0 (0)	NA	2056 (100)	NA
D-dimer, µg/L, median (IQR)	650 (300-1680)	0 (0)	680 (370-1414)	12 (0.6)	0.07
-D-dimer level between 0 µg/L to age-adjusted value, n (%)	656 (47)	0 (0)	834 (41)	12 (0.6)	0.001
-D-dimer level between age-adjusted value to 1000 µg/L, n (%)	223 (16)	0 (0)	485 (24)	12 (0.6)	0.00
-D-dimer level ≥ 1000 µg/L), n (%)	530 (38)	0 (0)	725 (36)	12 (0.6)	0.20
PE prevalence, n (%)	258 (18)	0 (0)	216 (11)	0 (0)	0.00

n: number; y: years; SD: standard deviation; VTE: venous thromboembolism; wk: weeks; PE: pulmonary embolism; CPP: clinical probability; IQR: interquartile range; NA: not applicable

* Patients in whom one or more 4PEPS items were missing

** Percentage was calculated by dividing the number of patients by the total number of patients in the study group minus number of missing values