Journal of Thrombosis and Haemostasis, **8**: 957–970 DOI: 10.1111/j.1538-7836.2010.03801.x

ORIGINAL ARTICLE

Clinical prediction rules for pulmonary embolism: a systematic review and meta-analysis

E. CERIANI,* C. COMBESCURE,† G. LE GAL,‡ M. NENDAZ,§ T. PERNEGER,† H. BOUNAMEAUX,* A. PERRIER§ and M. RIGHINI*

**Division of Angiology and Hemostasis, Geneva University Hospital and Faculty of Medicine, Geneva; †Division of Clinical Epidemiology, Geneva University Hospital and Faculty of Medicine, Geneva, Switzerland; ‡Department of Internal Medicine and Chest Diseases, Brest University Hospital, Brest, France; and §General Internal Medicine, Geneva University Hospital and Faculty of Medicine, Geneva, Switzerland*

To cite this article: Ceriani E, Combescure C, Le Gal G, Nendaz M, Perneger T, Bounameaux H, Perrier A, Righini M. Clinical prediction rules for pulmonary embolism: a systematic review and meta-analysis. *J Thromb Haemost* 2010; **8:** 957–70.

Summary. *Background:* Pretest probability assessment is necessary to identify patients in whom pulmonary embolism (PE) can be safely ruled out by a negative D-dimer without further investigations. *Objective:* Review and compare the performance of available clinical prediction rules (CPRs) for PE probability assessment. *Patients/methods:* We identified studies that evaluated a CPR in patients with suspected PE from Embase, Medline and the Cochrane database. We determined the 95% confidence intervals (CIs) of prevalence of PE in the various clinical probability categories of each CPR. Statistical heterogeneity was tested. *Results:* We identified 9 CPR and included 29 studies representing 31215 patients. Pooled prevalence of PE for three-level scores (low, intermediate or high clinical probability) was: low, 6% (95% CI, 4–8), intermediate, 23% (95% CI, 18–28) and high, 49% (95% CI, 43–56) for the Wells score; low, 13% (95% CI, 8–19), intermediate, 35% (95% CI, 31–38) and high, 71% (95% CI, 50–89) for the Geneva score; low, 9% (95% CI, 8–11), intermediate, 26% (95% CI, 24–28) and high, 76% (95% CI, 69–82) for the revised Geneva score. Pooled prevalence for two-level scores (PE likely or PE unlikely) was 8% (95% CI,6–11) and 34% (95% CI,29–40) for the Wells score, and 6% (95% CI, 3–9) and 23% (95% CI, 11–36) for the Charlotte rule. *Conclusion:* Available CPR for assessing clinical probability of PE show similar accuracy. Existing scores are, however, not equivalent and the choice among various prediction rules and classification schemes (three- versus two-level) must be guided by local prevalence of PE, type of patients considered (outpatients or inpatients) and type of D-dimer assay applied.

Correspondence: Marc Righini, Division of Angiology and Hemostasis, Department of Internal Medicine, Geneva University Hospital and Faculty of Medicine, 24 rue Micheli-du-Crest, CH-1211 Geneva 14, Switzerland.
Tel.: +41 22 372 92 94; fax: +41 22 372 92 99.
E-mail: Marc.Righini@hcuge.ch

Received 7 November 2009, accepted 1 February 2010

Keywords: clinical prediction rules, D-dimer, pulmonary embolism.

Introduction

Clinical assessment of the probability of pulmonary embolism (PE) is a crucial step in contemporary diagnostic strategies because the correct interpretation of test results depends on it. For example, the association of a low or intermediate clinical probability of PE with a normal D-dimer ELISA (enzyme-linked immunosorbent assay) test confidently rules out PE, reducing the need for further testing and costs [1]. However, in the presence of a high clinical probability, most clinicians feel that additional tests are necessary because the posttest probability of PE is still high (between 10 and 20%) despite a normal D-dimer result [2]. A normal computed tomographic pulmonary angiography (CTPA) may also safely rule out PE if clinical pretest probability is low or intermediate [2,3]. The negative predictive value of a negative CTPA is, however, low in patients with a high clinical probability. Therefore, clinical probability assessment is of utmost importance in the diagnostic approach of PE. Until recently, grouping patients into low, intermediate and high probability of PE was done implicitly using global clinical judgment. In the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) study [4], the prevalence of PE in the low, intermediate and high clinical probability categories was 9%, 30% and 68%, respectively. Although several studies confirmed the fair accuracy of implicit evaluation, it has been criticized, mainly because it is not standardized. Therefore, attempts have been made to standardize and render explicit the evaluation of clinical probability using statistically derived scores or clinical prediction rules (CPRs) that are able to provide estimates of the probability of PE based on clinical information. As these scores tend to be tailored to the characteristics of the patients used in the derivation model, they may be less accurate when applied to different sets of patients and must be externally validated. Ideally, outcome studies should demonstrate that patients may be safely managed on the basis of the assessment of the clinical

probability they provide [5,6]. The first prediction rule for PE was reported by Hoellerich *et al.* in 1986 [7]: this eight-item prediction rule was initially derived and locally validated in a small sample but the following steps of validation were never performed. Within the last two decades, many different rules have been described. The most widely used CPR is the Wells rule, which includes the physician's judgement of whether an alternative diagnosis is more likely than PE. This criterion, which carries a major weight in the score, is subjective and cannot be standardized. Therefore, several efforts were made to develop entirely objective scores such as the Geneva, Charlotte and Miniati rules. The first fully objective scores required diagnostic tests that are not always available, such as chest X-ray or blood gas analysis on room air. More recent rules are based only on clinical elements. Furthermore, in the last decade the most used rules were modified to increase their usefulness and acceptability for clinicians. Cut-off scores of the three-level rules that stratified patients into three levels of clinical probability (low, intermediate or high) were modified to obtain two-level rules classifying patients in only two categories ('PE likely/unlikely' for the Wells score or 'safe/unsafe' for the Charlotte rule). Existing scores were simplified by assigning one point to each item instead of the variable number of points per item in the initial rules, to simplify the memorization and computation of the score. This variety makes a rational choice among available scores for assessing clinical probability of PE difficult. The aim of this meta-analysis is to compare the accuracy of the principal CPRs for PE pretest probability estimation and review their level of validation.

Methods

Search strategy and study selection

We systematically searched Medline, Embase and the Cochrane Controlled Trials registry, using the following key words: *pulmonary embolism* AND (*decision tree* OR *clinical prediction rule* OR *clinical prediction score* OR *clinical decision rule* OR *clinical decision score* OR *management studies* OR *outcome studies* OR *D-dimer*). The search was performed for English, French, Italian, Spanish and German language, and completed on 15 July 2008. To ensure a comprehensive literature search, we examined reference lists from retrieved articles and reference literature (guidelines and systematic reviews) and questioned experts in PE diagnostic strategies for possible missing studies. Eligible studies were studies on the diagnosis of PE that used a CPR in the diagnostic work-up of patients with suspected PE. External validation in at least one study was required for inclusion of a given prediction rule in this systematic review. We accepted derivation and validation studies and retrospective and prospective studies. Studies in which the pretest probability was evaluated implicitly were excluded. If key data were missing, we contacted the study authors to request relevant data. Two investigators independently evaluated studies for possible inclusion (E.C., M.R.). For each study, two investigators who were blinded to the study authors and journal in which the studies were published assessed study quality independently and extracted the data on study design and patients characteristics. Disagreements about extracted data were resolved by consensus or by discussion with a third reviewer.

Quality assessment and data extraction

Methodological quality of included studies was assessed independently by two observers (E.C. and M.R.) using the QUADAS tool (Table 1), which is a quality assessment tool specifically developed for systematic reviews of diagnostic accuracy studies [8]. Regarding the patient spectrum criterion, we considered as representative cohorts that included consecutive patients with a PE prevalence between 15 and 25%. The QUADAS standards also include clarity of the methodological description for the diagnostic tool evaluated (in our case a CPR), an explicit and valid reference criterion standard and blinded interpretation of both CPR and the criterion standard. We considered appropriate the following reference standard tests: negative D-dimer in the presence of low or intermediate clinical probability, ventilation-perfusion lung scan when used and interpreted according to well-accepted criteria, helical computed tomography and pulmonary angiography [9]. The description of the reference standard was considered appropriate if explained with sufficient detail to permit replication of the test. The time between reference standard and index test (CRP calculation) was considered acceptable if < 24 h. To assess the influence of study quality on the results of our analysis, we calculated a global QUADAS score for each study, consisting of the sum of the scores for each individual criterion (score of 1 for criterion met, score of 0 for each criterion not met or if it was unclear whether the criterion was met). Studies with a global score above the median score were considered high quality and those with a score below the median low quality.

We collected the following data for each study: year of publication, type of CPR evaluated, data collection methods (retrospective or prospective), setting (outpatient, inpatient or both), geographic location, demographic characteristics of studied population (mean age, percentage of women), type of reference standard applied, prevalence of chronic obstructive lung disease, heart failure and cancer, follow-up duration, prevalence of PE in the study population, distribution of patients in each group of pretest probability and prevalence of PE in each category of pretest probability. We considered as acceptable a PE definition which was in line with well-established diagnostic criteria reported in the literature. Venous thromboembolisms (VTE) detected during follow-up in patients in whom anticoagulation was withheld were included in the pooled proportion. As most studies used a 90-day follow-up (mean follow-up period was 106 days, median value was 90 days) we did not perform an adjustment for the difference in follow-up length.

Table 1 Summary of retrieved studies and their quality evaluation using the QUADAS tool [8]

Author, year	Score	1. Representative spectrum of patients	2. Selection criteria clearly described	3. Reference standard likely to correctly classify the target condition	4.Time between reference standard and index test short enough to be reasonably sure that the target condition did not change	5.Whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis	6.Patients receive the same reference standard regardless of the index test result	7.Reference standard independent of the index test (i.e. the index test did not form part of the reference standard)	8.Score assessment described in sufficient detail to permit its replication	9.Reference standard described in sufficient detail to permit its replication	10.Pretest probability scored without knowledge of the results of the reference standard	11.Reference standard results interpreted without knowledge of PP	12.Same clinical data available when test results were interpreted as would be available when the test is used in practice	13. Uninterpretable/ intermediate test results reported	14. Withdrawals from the study explained	Total *n*/14
Righini *et al.*, 2008 [25]	GENEVA RV	1	1	1	1	1	0	1	1	1	1	1	1	1	1	13
Klok *et al.*, 2008 [39]	GVA RV. SIMP.	1	1	1	1	1	0	0	1	1	1	1	1	1	1	12
Klok *et al.*, 2008 [26]	GENEVA RV; WELLS, 3-level	1	1	1	0	1	0	0	1	1	1	1	1	0	0	9
Gibson *et al.*, 2008 [40]	WELLS SIMP.	1	1	1	0	1	0	0	1	1	1	1	1	1	1	11
Goekoop *et al.*, 2007 [41]	WELLS, 2-level	1	1	1	0	1	0	0	1	0	1	0	1	0	1	8
Penaloza *et al.*, 2007 [18]	WELLS, 3-level; WELLS, 2-level	1	1	1	1	1	0	0	1	1	1	0	1	1	1	11
Christ Study Inv, 2006 [20]	WELLS, 2-level	1	1	1	1	1	0	1	1	1	1	1	1	1	1	13
Kearon *et al.*, 2006 [42]	WELLS, 3-level	1	1	1	0	1	0	1	1	1	1	0	1	1	1	11
Kline *et al.*, 2006 [15]	WELLS, 3-level; CHARLOTTE	0	1	1	0	1	0	0	1	1	1	1	1	0	1	9
Le Gal *et al.*, 2006 [24]	GENEVA RV	1	1	1	1	1	0	1	1	1	0	0	1	1	1	11
Ollengerg *et al.*, 2006 [23]	WELLS, 3-level; GENEVA	0	0	1	0	1	1	1	1	1	0	0	1	0	0	7
Kline *et al.*, 2006 [43]	WELLS, 3-level	0	1	1	0	1	1	1	1	1	1	1	1	0	0	10
Miniati *et al.*, 2005 [16]	MINIATI, 3-level, WELLS, 3-level, GENEVA	0	1	1	1	1	1	1	0	1	1	0	1	0	1	10
Perrier *et al.*, 2005 [21]	GENEVA	1	1	1	1	1	0	1	1	1	1	1	1	1	1	13
Steeghs *et al.*, 2005 [44]	WELLS, 2-level	1	1	1	0	1	0	0	1	1	1	0	1	1	1	10
Runyon *et al.*, 2005 [45]	WELLS, 3-level, CHARLOTTE	0	0	1	0	1	1	1	1	0	1	1	1	0	0	8
Bosson *et al.*, 2005 [28]	WELLS, 3-level	0	1	1	0	1	0	0	1	0	1	1	1	1	1	9

Table 1 (Continued)

Author, year	Score	1. Representative spectrum of patients	2. Selection criteria clearly described	3. Reference standard likely to correctly classify the target condition	4.Time between reference standard and index test short enough to be reasonably sure that the target condition did not change	5.Whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis	6.Patients receive the same reference standard regardless of the index test result	7.Reference standard independent of the index test (i.e. the index test did not form part of the reference standard)	8.Score assessment described in sufficient detail to permit its replication	9.Reference standard described in sufficient detail to permit its replication	10.Pretest probability scored without knowledge of the results of the reference standard	11.Reference standard results interpreted without knowledge of PP	12.Same clinical data available when test results were interpreted as would be available when the test is used in practice	13. Uninterpretable/ intermediate test results reported	14. Withdrawals from the study explained	Total *n*/14
Kabrhel *et al.*, 2005 [17]	WELLS, 2-level, WELLS, 3-level	0	0	1	0	1	1	1	1	0	1	0	1	0	0	7
Wolf *et al.*, 2004 [19]	WELLS, 2-level, WELLS, 3-level	0	0	1	0	1	0	1	1	0	1	0	1	1	1	8
Kline *et al.*, 2004 [46]	CHARLOTTE	1	0	1	0	1	0	0	1	0	1	0	1	1	1	8
Perrier *et al.*, 2004 [1]	GENEVA	1	1	1	1	1	0	1	1	1	1	1	1	1	1	13
Miniati *et al.*, 2003 [47]	MINIATI, 4-level	0	1	0	1	1	0	1	1	0	1	1	1	1	1	10
Miniati *et al.*, 2003 [48]	MINIATI, 4-level	0	0	1	0	1	0	1	0	0	1	0	1	0	1	6
Aujesky *et al.*, 2003 [22]	GENEVA	1	1	1	0	1	0	0	1	1	1	0	1	1	1	10
Kline *et al.*, 2002 v	CHARLOTTE	0	1	1	0	1	0	0	1	0	1	0	1	1	0	7
Wells *et al.*, 2001 [13]	WELLS, 3-level	1	1	1	1	1	0	0	1	1	1	1	1	1	1	12
Wicki *et al.*, 2001 [30]	GENEVA	1	1	1	1	1	0	1	1	1	1	0	1	1	0	11
Sanson *et al.*, 2000 [14]	WELLS, 3-level	1	1	1	1	1	1	1	1	1	1	0	1	0	0	11
Wells *et al.*0, 2000 [29]	WELLS, 3-level; WELLS, 2-level	1	1	1	1	1	0	1	1	1	1	0	1	1	0	11

PP, pretest probability.

Primary outcome of the study was to estimate the pooled prevalence of PE in each group of pretest probability for retrieved CPR.

Data analysis

We determined the 95% confidence intervals (CIs) of the prevalence of PE in the various clinical probability categories using the exact method (see Appendix S1).The prevalence of VTE in each level of clinical probability was separately assessed using the method of the inverse variance on the arcsine transformed proportions [10]. Heterogeneity was tested with the Cochran Q statistic and also quantified by the indicator I^2 (ranging from 0% for perfect homogeneity to 100% for extreme heterogeneity) [11]. In case of heterogeneity (Cochran Q test with a *P*-value < 0.10 or $I^2 > 50\%$) a random effect model was used [12]. The heterogeneity factors were explored graphically by plotting the prevalence in each level against the factors. For the most important heterogeneity factors, a sub-group analysis was performed. The pooled results were given for each sub-group. Heterogeneity factors were explored only if the number of studies was greater than five (two- and three-level Wells score and Geneva score). Potential factors were: prospective/retrospective design, multicenter or single center study, blinded or unblinded readers, consecutive patients vs. other selection mechanism, location (North America or Europe), type of reference standard (imaging, follow-up or both), mean age, inpatients and outpatients/outpatients only recruitment and overall prevalence of PE. For the last point a cut off of 20% was chosen because it is the mean and median value of PE prevalence among selected studies. Moreover, PE prevalence among European studies is 20–25%, whereas in American and Canadian studies it is usually between 10 and 20%; thus this cut-off of 20% could highlight different local management of suspected PE.

Publication bias was explored by funnel plots and Egger's test. Potential bias concerned the low risk group, so we plotted the logarithm of the odds ratio of PE (between low group and the other groups) against its standard error.

Results

Selection

The literature search yielded 3752 articles, of which 130 were potentially relevant for our purpose and, therefore, screened in detail. Of these 130 articles, 104 were discarded for various reasons (Fig. 1), and 26 were included. Three additional articles

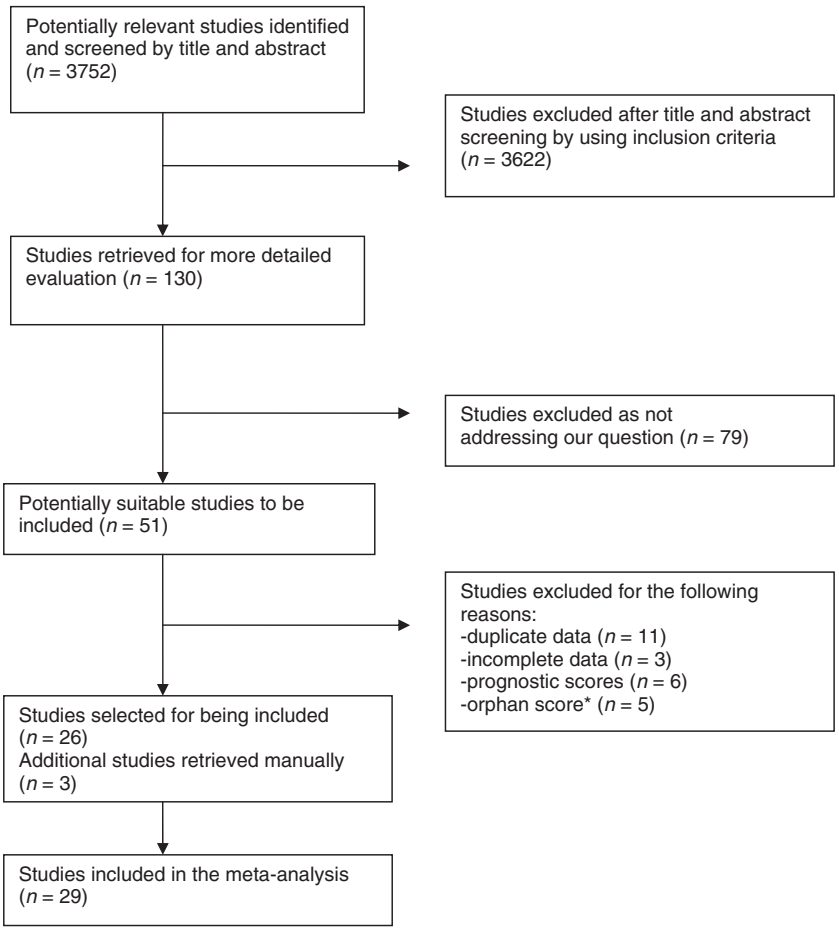

Fig. 1. Study flow diagram. *score that was derived in a single study, but never validated

were manually retrieved by asking experts in VTE or by reviewing the reference list of retrieved articles. Twenty-nine studies were finally included in this systematic review (Fig. 1).

Some studies evaluated more than one CPR. Nine different CPRs were identified: the Wells score, three-level (low, intermediate and high clinical probability) or two-level ('PE likely', 'PE unlikely'), the Wells Simplified score, the Geneva score, the Revised Geneva score, the Simplified Revised Geneva score, the Miniati score at three or four levels of probability and the Charlotte rule.

The three-level Wells score was examined in 14 studies, of which 11 were prospective and three were retrospective. The two-level Wells score was tested in seven studies, of which six were prospective and one was retrospective. The Geneva score was examined in six studies, of which four were prospective and two were retrospective. The number of studies assessing the performance of the other rules was as follows: Charlotte rule, four; Revised Geneva score, three; four-level Miniati score, two; three-level Miniati score, one; Simplified Revised Geneva score, one; and Simplified Wells score, one. Table 2 displays in detail the retrieved CPRs, while Tables 1 and 3 outline the characteristics of the studies in which these rules were evaluated. Pooled results of PE prevalence in each pretest probability for different clinical prediction models are displayed in Table 4.

Three-level Wells score

The prevalence of PE ranged from 1.3% [13] to 27.9% [14] in the low clinical probability category, from 8.6% [15] to 54.2% [16] in the intermediate probability category and from 33.3% [17] to 100% [18] in the high probability category. Pooled prevalence was 5.7% in the low, 23.2% in the intermediate and 49.3% in the high clinical probability groups. Random effects models were used because heterogeneity was detected in each of the three levels ($P < 0.001$ for each level, Cochran's test).

Two-level Wells score

The prevalence of PE ranged from 3.4% [19] to 12.1% [20] in the 'PE unlikely' group and from 22.8% [7] to 48.8% [18] in the 'PE likely' group. The pooled prevalence was 8.4% in the 'PE unlikely' group and 34.4% in the 'PE likely' one. Random effects models were used because heterogeneity was detected ($P < 0.001$ in both levels).

Geneva score

The prevalence of PE ranged from 6.5% [1] to 50.0% [16] in the low probability group, from 30.0% [21] to 41.4% [22] in the intermediate group and from 43.7% [23] to 96.3% [21] in the high probability group. The prevalence of PE in the low probability subgroup was exceptionally high in the study from Miniati (50.0%). In the other studies, this prevalence ranged from 6.5% [1] to 18.4%[23]. The pooled prevalence of PE (including the study from Miniati) was 12.8% in the low probability group, 34.7% in the moderate and 71.1% in the high probability group. Random effects models were also used because heterogeneity was detected in the three categories ($P < 0.001$ in low and high and 0.05 in intermediate, Cochran's test).

Revised Geneva score

The prevalence of PE ranged from 7.9% [24] to 9.4% [25] in the low probability group, from 22.8% [26] to 28.5% [24] in the intermediate level and from 71.4% [26] to 84.0% [25] in the high probability category. The pooled prevalence of PE was 9.0% in the low probability level, 26.2% in the intermediate one and 75.7% in the high one. Heterogeneity was not detected in the intermediate level ($P < 0.001$) nor in the low and high levels (respectively $P = 0.92$, $P = 0.31$ and $P = 0.43$ in the low, intermediate and high classes respectively).

Charlotte Rule

The prevalence ranged from 3.9% [15] to 13.3% [27] in the 'safe' category and from 12.0% [15] to 42.1% [27] in the 'unsafe' group. The pooled prevalence was 5.9% in the safe level and 22.5% in the unsafe one. Heterogeneity was detected in both levels and was strong ($P < 0.001$ in both levels) but not explored because of the small number of studies ($n = 4$).

Study quality

Table 1 shows the assessment of study quality (median QUADAS score, 10 points). The performances of the scores for PE did not differ between high- and low-quality studies.

Exploration of heterogeneity

We present only the results for the low probability category. The results of the sub-groups analysis are shown in Table 4. The overall prevalence of PE in the studies was identified as a factor of heterogeneity for the three-level Wells score: the prevalence ranged from 1.3% [13] to 5.4% [28] among the studies with an overall PE prevalence of less than 20% ($n = 11$) and the pooled prevalence was 3.4%, whereas it ranged from 12.5% [16] to 27.9% [14] in the studies with a higher overall PE prevalence ($n = 4$), with a pooled prevalence of 15.6%. But the heterogeneity was still significant even when studies with an overall prevalence of PE below and above 20% were pooled separately ($P < 0.01$ for both groups of studies, see I^2 values in Table 4).

The location of the studies was identified as a source of heterogeneity for the two-level Wells score: the prevalence ranged from 3.4% [19] to 7.8% [29] among the North American studies ($n = 4$) with a pooled prevalence of 6.5%, and from 9.2% [18] to 12.1% [20] among the European studies ($n = 4$), with a pooled prevalence of 11.5% (see Table 4). Heterogeneity was no longer significant in both American and

Table 2 Clinical scores for predicting the pretest probability of pulmonary embolism

Geneva	Points	Revised Geneva	Points	Wells	Points	Miniati	Coefficient
Recent Surgery	3.0	Age > 65 years old	1.0	Clinical signs of DVT	3.0	Male sex	0.81
Previous DVT or PE	2.0	Previous history of PE or DVT	3.0	Recent surgery or immobilization	1.5	*Age*	
Heart rate > 100 bpm	1.0	Surgery or fracture within 1 month	2.0	Heart rate > 100 bpm	1.5	63–72 years old	0.59
Age		Active malignancy	2.0	Previous history of PE or DVT	1.5	≥ 73 years old	0.92
60–79 years old	1.0	*Heart rate (bpm)*		Hemoptysis	1.0	*Preexisting disease*	
≥ 80 years old	2.0	75–94	3.0	Malignancy	1.0	Cardiovascular	0.56
Chest Radiograph		≥ 95	5.0			Pulmonary	−0.97
Atelectasis	1.0	Pain on leg venous palpation and unilateral edema	4.0	Alternative diagnosis less likely than PE	3.0	History of thrombophlebitis	0.69
Elevated hemidiaphram	1.0	Unilateral leg pain	3.0			Dyspnea (sudden onset)	1.29
PaO_2		Hemoptysis	2.0			Chest pain	0.64
< 49 mm Hg (6.5 kPa)	4.0					Fever > 38°	−1.17
49–59 mm Hg (6.5–7.99 kPa)	3.0					ECG signs of acute right ventricular overload	1.53
60–71 mm Hg (8–9.49 kPa)	2.0					Chest radiograph	
72–82 mmHg (9.5–10.99 kPa)	1.0			*3 levels*		Oligoemia	3.86
$PaCO_2$				*Low*	*< 2*	Amputation of the hilar artery	3.92
< 36 mmHg (4.8 kPa)	2.0			*Intermediate*	*2–6*	Consolidation (infarction)	3.55
36–38·9 mmHg (4.8–5.2 kPa)	1.0			*High*	*> 6*	Consolidation (no infarction)	−1.23
Low	0–4	*Low*	0–3	*2 levels*		Pulmonary edema	−2.83
Intermediate	5–8	*Intermediate*	4–10	*PE unlikely*	*≤ 4*	Constant	−3.26
High	≥ 9	*High*	≥ 11	*PE likely*	*> 4*		

Charlotte Rule		Simp. Rev. Geneva	Points	Simp. Wells	Points		
> 50 yers old		Age > 65 years old	1.0	Clinical signs of DVT	1.0		
Heart rate > systolic blood pressure		Previous history of PE or DVT	1.0	Recent surgery or immobilization	1.0		
Unexplained hypoxemia (O_2 < 95%)		Surgery or fracture within 1 month	1.0	Heart rate > 100 beats per minutes	1.0		
Recent surgery (previous 4 weeks)		Active malignancy	1.0	Previous history of PE or DVT	1.0		
Hemoptysis		*Heart rate (bpm)*		Hemoptysis	1.0	*Pretest probability (%) = 1/[1 + exp(-sum)]*	
Unilateral leg swelling		75–94	1.0	Malignancy	1.0	*4 levels*	
		≥ 95	1.0	Alternative diagnosis less likely than PE	1.0	*Low*	*≤ 10%*
		Pain on leg deep vein palpation	1.0			*Intermediate*	*10–50%*
		Unilateral leg pain	1.0			*Moderately High*	*50–90%*
		Hemoptysis	1.0			*High*	*> 90%*
Any of the previous present	*Unsafe*	*Low*	*0–1*	*Unlikely*	*≤ 1*	*3 levels*	
All of the previous absent	Safe	*Intermediate*	*2–4*	*Likely*	*> 1*	*Low*	*≤ 10%*
		High	*≥ 5*			*Intermediate*	*10–90*
						High	*> 90%*

Table 3 Summary of the study characteristics

Studies	Characteristics*	*N*	Mean age, years	Sex (%F)	COPD (%)	HF (%)	Cancer (%)	Setting	Follow up (d)	Prevalence (%)			
										Study	In each level, % (95%CI)		
WELLS, 3-level											Low	Intermediate	High
Wells *et al.*, a [29]	P; D	972	NA	NA	NA	NA	NA	In-out	90	17	4 (2–6)	21 (17–24)	67 (54–78)
Wells *et al.*, b [29]	P; V	247	NA	NA	NA	NA	NA	In-out	90	16	2 (0–7)	19 (12–27)	50 (27–73)
Wells *et al.* [13]	P; V	930	50.5	62.7	NA	NA	7.2	Out	90	9	1 (0–3)	16 (13–21)	38 (26–51)
Wolf *et al.* [19]	P; V	134	58†	54	NA	NA	NA	Out	90	12	2 (0–9)	15 (7–26)	43 (18–71)
Kabrhel *et al.* [17]	P; V	607	47.9	74	NA	NA	NA	Out	90	10	4 (2–7)	14 (10–19)	33 (20–48)
Bosson *et al.* [28]	P; V	1528	67	54.1	NA	NA	13	In-out	0	20	5 (4–7)	28 (24–31)	47 (39–55)
Runyon *et al.* [45]	R; V	2477	45	70	NA	NA	8	Out	45	6	3 (2–4)	12 (9–15)	33 (20–48)
Kline *et al.* [43]	P; V	178	48	NA	NA	NA	NA	Out	90	14	3 (1–8)	24 (13–37)	62 (32–86)
Penaloza *et al.* [18]	P; V	185	56	58.9	NA	NA	11	Out	90	18	3 (1–9)	31 (21–43)	100 (47–100)
Kearon *et al.* [42]	P; V	1126	57	65	NA	NA	14	In-out	180	15	5 (4–7)	26 (22–31)	55 (43–67)
Kline *et al.* [15]	P; V	2302	44.7	69	13	7	8	Out	90	5	3 (2–4)	9 (6–11)	26 (13–42)
Sanson *et al.* [14]	P; V	414	51	58	NA	NA	10	In-out	0	29	28 (21–36)	30 (25–36)	38 (9–76)
Miniati *et al.* [16]	P; V	215	70	64	NA	NA	17	NA	365	43	13 (6–23)	54 (45–63)	64 (45–80)
Ollengerg *et al.* [23]	R; V	1359	55	55	12	18	15	In-out	365	29	17 (15–21)	36 (33–40)	55 (46–65)
Gibson *et al.* [40]	R; V	3298	NA	NA	NA	NA	NA	In-out	90	21	7 (6–9)	26 (24–27)	58 (50–65)
WELLS, 2-level											Unlikely		Likely
Wells *et al.*, a [29]	P; D	964	NA	NA	NA	NA	NA	In-out	90	18	8 (6–10)		41 (35–47)
Wells *et al.*, b [29]	P; V	247	NA	NA	NA	NA	NA	In-out	90	15	5 (2–9)		39 (28–52)
Wolf *et al.* [19]	P; V	134	58†	54	NA	NA	NA	Out	90	12	3 (1–10)		28 (16–44)
Kabrhel *et al.* [17]	P; V	607	47.9	74	NA	NA	NA	Out	90	10	6 (4–8)		23 (17–31)
Steeghs *et al.* [44]	P; V	331	51	61.9	13.3	3	0.9	Out	90	14	11 (7–15)		31 (19–45)
Christopher Inv. [20]	P; V	3306	53	57.4	10.3	7.4	14	In-out	90	20	12 (11–14)		37 (34–40)
Penaloza *et al.* [18]	R; V	185	56	58.9	NA	NA	11	Out	90	18	9 (5–15)		49 (33–65)
Goekoop *et al.* [41]	P; V	879	51	62.6	13	20	1.4	Out	90	13	11 (8–13)		29 (20–39)
WELLS SIMP.													
Gibson *et al.* [40]	R; V	3298	NA	NA	NA	NA	NA	In-out	90	21	11 (10–12)		36 (33–38)
GENEVA											Low	Intermediate	High
Wicki *et al.* [30]	P; D	986	62†	55	12	NA	13	Out	90	27	10 (8–13)	38 (33–43)	81 (69–90)
Aujesky *et al.* [22]	P; V	259	63†	58	16	10	13	Out	90	30	9 (4–15)	41 (32–52)	59 (43–74)
Perrier *et al.* [1]	P; V	965	61	58	10	10	9	Out	90	23	7 (5–9)	34 (29–39)	85 (75–92)
Perrier *et al.* [21]	P; V	756	60	60	10	8	10	Out	90	26	7 (5–10)	30 (25–36)	96 (90–99)
Miniati *et al.* [16]	R; V	215	70	64	NA	NA	17	NA	365	43	50 (30–70)	39 (31–48)	49 (36–62)
Ollengerger *et al.* [23]	R; V	998	55	55	12	18	15	In-out	365	29	18 (14–23)	31 (27-35)	44 (36-51)
GENEVA REVISED											Low	Intermediate	High
Le Gal *et al.*, a [24]	P; D	956	60.6	58.2	10.3	9.8	9.2	Out	90	23	9 (6-13)	28 (24-31)	72 (58-83)
Le Gal *et al.*, b [24]	P; V	749	NA	NA	NA	NA	NA	Out	90	26	8 (5-12)	29 (24-33)	74 (60-85)
Klok *et al.* [26]	R; V	300	NA	NA	NA	NA	NA	In-out	90	16	8 (5-14)	23 (16-31)	71 (29-96)
Righini *et al.* [25]	P; V	1693	59.3	55,3	NA	NA	NA	Out	90	18.2	9 (7-12)	25 (22-28)	84 (71-93)

Table 3 (Continued)

GENEVA REVISED SIMPLIFIED											Low	Intermediate	High	
Klok *et al.* [39]	R;V	1049	NA	NA	NA	NA	In-out	90	23	8 (5-11) Unsafe	29 (26-33)	64 (48-78) Safe	
CHARLOTTE														
Kline *et al.* [27]	P; D	934	52	68	21	NA	17	Out	180	19	13 (11–16)		42 (35–49)	
Kline *et al.* [46]	P; V	1339	44	69	NA	NA	NA	Out	90	6	4 (3–6)		22 (15–30)	
Kline *et al.* [15]	P; V	2302	44.7	69	13	7	8	Out	90	5	4 (3–5)		12 (8–17)	
Runyon *et al.* [45]	P; V	2477	45	70	NA	NA	8	Out	45	6	4 (3–5)		17 (13–22)	
MINIATI, 4-level											Low	Intermediate	Moderately-high	High
Miniati *et al.* [47]	P; V	390	68.5	58	NA	NA	14	In-out	365	41	3,1 (1–7,1)	24 (16–34)	83 (66–93)	100 (96–100)
Miniati *et al.* [48]	P; D	1100	68[†]	55	NA	NA	14.7	In-out	180	40	4 (3–7)	22 (17–27)	74 (62–83)	98 (95–99)
MINIATI, 3-level											Low	Intermediate	High	
Miniati *et al.* [16]	P; V	215	70	64	NA	NA	17	NA	365	43	5 (1–12)	42 (31–53)	98 (91–100)	

NA, not specifically recorded; HF, heart failure; COPD, chronic obstructive pulmonary disease. [*]Study characteristics refers to study design (P, prospective; R, retrospective; V, validation; D, derivation). [†]Median age.

European studies when analyzed separately ($P = 0.24$ and $P = 0.12$, respectively).

The retrospective or prospective design of the studies was identified as a source of heterogeneity for the Geneva score: the prevalence of PE ranged from 6.5% [1] to 10.1% [30] among the prospective studies ($n = 5$) with a pooled prevalence of 8.0%, and was 18.4% [23] to 50.0% [16] in the retrospective studies ($n = 2$) with a pooled prevalence of 32.1% (see Table 4). Heterogeneity was detected in the retrospective studies ($P < 0.001$) but not in the prospective studies ($P = 0.19$).

The type of population tested (in- and outpatients vs. outpatients only) was also a source of heterogeneity. In the studies including only outpatients, heterogeneity was not detected for the three-level Wells score ($P = 0.28$) and the Geneva score ($P = 0.19$). For these studies, the pooled prevalence was 2.9% (three-level Wells score) and 8.0% (Geneva score).

Heterogeneity was similar in the low (≤ 10) and high-quality (> 10) studies as assessed by the QUADAS score.

Publication

A publication bias for the three-level Wells score was suspected ($P = 0.09$, Egger's test): a study with a small sample size was more likely to be published if it showed an important contrast between the low level and the other levels (see Appendix S1 for funnel plot). For the other scores, no publication bias was identified.

Discussion

In this meta-analysis, we studied the accuracy of nine different CPRs for the evaluation of PE pretest probability (Table 2). Overall, our data show that these rules have comparable accuracy, which is reassuring for clinicians (Table 4). However, several important differences should be pointed out. First, they have not all been validated to the same extent and some validation studies were of lower quality. The most extensively validated rules are the three-level and two-level Wells scores, the Geneva score, the Revised Geneva score and the Charlotte rule. Of the 29 articles retrieved for this analysis, 25 used various versions of the Wells score or the Geneva score, representing 86% of the 31215 patients included in our meta-analysis. These scores were associated with the highest level of validation according to methodological standards developed for CPRs (Table 5) [31]. Furthermore, each of these rules has been evaluated in outcome studies, demonstrating that patients can be safely managed based on clinical assessment by these scores. In contrast, the Miniati score was validated in a relatively small number of patients, in a local context characterized by a very high prevalence of PE and mainly hospitalized patients, making its performance in other populations unpredictable. Second, the CPRs included in this analysis have been applied to various patient populations. All rules have been validated in outpatients, but only the Wells rule and the Miniati

Table 4 Pooled prevalence of PE in each pretest probability level for the different retrieved CPR

Scores	Studies	Levels					
Wells 3-level		Low	I^2 (%)	Intermediate	I^2 (%)	High	I^2 (%)
	All (n = 14)	5.7 (3.7–8.2)	94	23.2 (18.3–28.4)	95	49.3 (42.6–56.0)	71
	With prev < 0.20	3.4 (2.7–4.3)	57	18.8 (14.3–23.8)	92	46.8 (38.3–55.5)	73
	With prev > 0.20	15.6 (7.6–25.7)	96	35.6 (26.5–45.3)	90	56.9 (51.5–62.2)*	0
	In-Out patients	8.5 (4.6–13.4)	95	26.6 (22.8–30.6)	87	54.4 (50.5–58.4)*	40
	Out patients	2.9 (2.4–3.5)*	19	15.8 (11.6–20.4)	84	40.8 (30.4–51.6)	61
Wells 2-level		Unlikely				Likely	
	All (N = 8)	8.4 (6.4–10.6)	81			34.4 (29.4–39.7)	70
	North American studies	6.5 (5.3–7.9)*	30			32.9 (23.2–43.3)	82
	European studies	11.5 (10.5–12.6)*	49			36.7 (34.0–39.3)*	0
Geneva		Low		Intermediate		High	
	All (N = 6)	12.8 (7.9–18.7)	91	34.7 (31.3–38.2)	55	71.1 (49.6–88.5)	96
	Prospective studies	8.0 (6.7–9.4)*	37	35.2 (31.0–39.6)	57	82.1 (65.8–94.0)	90
	Retrospective studies	32.1 (7.1–64.8)	91	34.2 (26.6–42.2)	67	45.1 (38.9–51.5)*	0
	In-out patients	18.4 (14.4-23.0)		30.9 (26.8–35.2)		43.7 (36.2–51.4)	
	Out patients	8.0 (6.7–9.4)*	37	35.2 (31.0–39.6)	57	82.1 (65.8–94.0)	90
Revised Geneva		Low		Intermediate		High	
	All (N = 6)	9.0 (7.6–10.6)*	0	26.2 (24.4–28.0)*	15	75.7 (69.0–81.8)*	0
Revised Geneva Simplified		Low		Intermediate		High	
	All (N = 1)	7.7 (5.2–10.8)		29.3 (25.7–33.0)		64.3 (48.0–78.4)	
Charlotte		Safe				Unsafe	
	All (N = 4)	5.9 (3.3–9.3)	96			22.5 (11.4–36.2)	95

*Heterogeneity was not detected and a fixed effects model was used. The other results were derived from a random effect model (resulting in an increased confidence interval to take into account the heterogeneity).

Table 5 Level of validation of the retrieved CPRs, according to reference [31]

Scores	Level of validation
Wells, 3 levels	1
Wells, 2 levels	1
Wells, 2 levels Simplified	3
Geneva	1
Revised Geneva	1
Revised Geneva Simplified	3
Miniati score	3
Charlotte rule	1

Level 4: derivation: identification of factors with predictive power. Level 3: narrow validation: Application of rule in a similar clinical setting and population as in step 1. Level 2: broad validation: Application of rule in multiple clinical settings with varying prevalence and outcomes of disease. Level 1: impact analysis: evidence that rule changes physician behaviour and improves patient outcomes and/or reduces costs.

rule have also been applied to hospitalized patients, albeit on small numbers of patients. Therefore, inpatients with suspected PE should probably be evaluated by the Wells or Miniati rule. The Charlotte rule has been validated on a large sample of patients. However, all patients above 50 years of age must be considered as 'unsafe', making the rule less useful in daily life.

Third, point estimates for the prevalence of PE in clinical probability groups vary according to the rule and classification scheme (three- vs. two-level). The point of using prediction rules is to obtain an accurate post-test probability of PE that can be used for clinical decision-making. Figure 2 shows that such variations may be clinically relevant. Assuming that a post-test probability of PE below 3% is necessary to safely rule out PE [32], Fig. 2 displays the range of pretest probabilities across which a given D-dimer assay may safely rule out PE. A highly sensitive D-dimer (sensitivity, 97%) would allow ruling out PE up to a pretest probability of 29%. Hence, using such an assay, PE could be ruled out by a normal D-dimer and a low or intermediate clinical probability or a 'PE unlikely' classification as assessed by the Wells two- or three-level scheme or the Revised Geneva rule, or a 'safe' classification by the Charlotte rule. In contrast, the post-test probability of PE in patients with an intermediate clinical probability as assessed by the original Geneva rule would be too high to safely rule out PE, unless one accepts that a post-test probability up to 5% would also be acceptable in intermediate probability patients. If a less sensitive D-dimer was used (sensitivity, 90%), PE could only be ruled out up to a pretest probability of 13%. Hence, PE would be satisfactorily ruled out only in patients with a negative less sensitive D-dimer and low clinical probability of PE using the Geneva and the Revised Geneva score, or a 'safe' Charlotte classification. Furthermore, the Wells rule would only be safe in populations with a low overall prevalence of PE, for example below 20%. In contrast, ruling out PE with a less sensitive D-dimer assay would be unsafe in patients with an

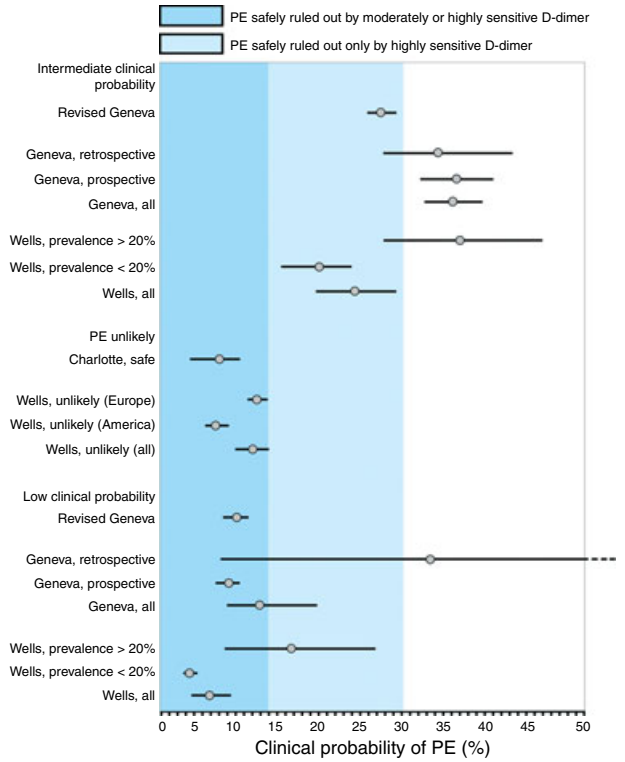

Fig. 2. Ruling out pulmonary embolism by combining clinical assessment by prediction rules and D-dimer measurement. The light blue and hatched blue zones represent the range of pretest probabilities across which a highly sensitive D-dimer assay (sensitivity, 97%; specificity, 40%) [38] may safely rule out PE. Safety is defined as a post-test probability of pulmonary embolism below 3%. Applying the Bayes rule, the 3% post-test probability threshold would be crossed above a pretest probability of 29%. The hatched blue zone represents the range of pretest probabilities across which a moderately sensitive D-dimer assay (sensitivity, 90%; specificity, 55%) [38] may safely rule out PE. Safety is defined as a post-test probability of pulmonary embolism below 3%. Applying the Bayes rule, the 3% post-test probability threshold would be crossed above a pretest probability of 13%.

intermediate clinical probability whichever rule was used. Therefore, we believe that this meta-analysis is useful to guide clinicians to choose rational and safe combinations between available scores and available D-dimer tests.

Fourth, accuracy for predicting the prevalence of PE in a given clinical probability group is not the only feature of a prediction rule that impacts on clinical decision-making. Indeed, the proportion of all patients classified in a given category is also important because it determines the proportion of patients in whom a D-dimer test can be applied. D-dimer should not be performed in patients with a high clinical probability or a PE likely classification, as its negative predictive value is lower in these subgroups.[33] Highly sensitive D-dimer assays may rule out PE in combination with a low or intermediate clinical probability, which represents around 90% of all patients with suspected PE in most cohorts. They can also be used in patients categorized as 'PE unlikely' by the two-level Wells rule, but these patients represent a smaller proportion of patients with suspected PE (around 70% when compared with the 90% of patients classified as having a low or an intermediate clinical probability in three-level rules). Therefore, it would make sense to apply three-level rules in combination with highly sensitive D-dimer assays to increase the proportion of patients in whom D-dimer can be measured and, hence, to increase the proportion of patients in whom PE could be ruled out by this simple biological test. Conversely, the diagnostic yield could be increased using a two-level rule when a less sensitive D-dimer is available. Indeed, this would allow performing D-dimer in the 70% of patients with a PE unlikely rating instead of only 50% of patients with a low clinical probability of PE classification by a three-level rule, as we have seen that the combination of a negative less sensitive D-dimer and an intermediate clinical probability is unsafe for ruling out PE.

The proportion of patients classified in a given category is also important because it impacts the proportion of patients who should be treated by anticoagulants while awaiting a definitive diagnostic confirmation. This is recommended by the latest ACCP consensus conference for any patient with a high clinical probability using a three-level rule or classified as PE likely by a two-level rule [34]. Applying this recommendation would result in submitting a higher proportion of patients (30%) to the risk of anticoagulants when using a two-level rule than a three-level rule (only 10% of patients or less with a high clinical probability). Although the risk of a short course of anticoagulants (typically < 24 h) is very low, it might become significant in patients at high risk of bleeding. Because of the PE prevalence in the intermediate clinical probability category, anticoagulants might also be recommended during the diagnostic process in these patients.

Strengths and weaknesses of this study

Our study has some limitations. First, we were not able to identify all sources of heterogeneity. For the three-level Wells score, heterogeneity was more marked in studies with a higher overall prevalence of PE (more than 20%), reflecting that score performances vary according to the population in which the score is used. For the Geneva score, there was less heterogeneity in prospective than in retrospective studies. Patients with missing data (mostly arterial blood gas analysis) were excluded from the retrospective validation studies. Therefore, those analyses may be biased towards inclusion of a sicker patient population of patients [23]. The Revised Geneva score showed heterogeneity only in the intermediate clinical probability category. Second, we included prospective and retrospective studies for the sake of completeness. Although retrospective studies allow estimation of the prevalence of PE in the various pretest probability groups, they may be subject to multiple biases. The Geneva score performed better in prospective studies and displayed less heterogeneity, which is reassuring (Fig. 2). Likewise, performances of the Revised Geneva score were similar in retrospective and prospective series. Third, the quality and the design of included studies were variable. In older studies, PE was ruled out by imaging studies, whereas in

more recent studies PE was excluded in many patients by various combinations of clinical probability and test results (mainly D-dimer and CTPA) and an uneventful 3-month follow-up. Length of follow-up could be a source of heterogeneity, possibly increasing the number of PE rate not related to the index event. However, duration of follow-up was similar (three months) in most studies (mean follow-up period was 106 days, while median value was 90 days). Another possible source of heterogeneity could be the difference in imaging tests performed, as angiography, CTPA or V/Q scan were used in the selected studies. Obviously these tests have a different diagnostic accuracy and their interpretation may vary with radiologist expertise. Fourth, our analysis is based only on published studies and publication bias may be a concern. However, we found evidence of publication bias only for one score. Our meta-analysis also has some strengths. It fulfils the criteria defined by the Quorum Statement for systematic reviews [35]. It provides clinicians with realistic estimates of the prevalence of PE in the various clinical probability subgroups defined by the best validated prediction rules. Hopefully, this might encourage clinicians to use CPRs for PE, which is still not the case in everyday clinical settings. A recent study showed that only around 50 to 75% of physicians were familiar with a clinical score for PE and that only one half of those used them in more than 50% of applicable cases [36]. This may not be without consequences for patient management. In a recent prospective cohort study performed in 116 emergency departments, the lack of clinical scoring was an independent risk factor for inappropriate management and worse outcome, including higher mortality in patients in whom PE had been inappropriately ruled out [37].

Which recommendations may be drawn from our analysis? First, available prediction rules for assessing clinical probability of PE have a similar accuracy and there is no clear winner. Although we could not demonstrate the superiority of a single CPR, our results show that combinations of the various prediction rules and the various D-dimer tests do not always result in a post-test probability of less than 3%. Therefore, three-level classification models should be best used with highly sensitive D-dimer tests, as exclusion of PE in patients with an intermediate clinical probability will only be safe with these tests. Alternatively, two-level models may be used with less sensitive tests. Second, North-American CPRs seem to be less reliable if the prevalence of PE exceeds 20%; in that case, European CPRs should be preferred. Third, the Geneva scores were prospectively validated in outpatients only, and therefore it is wise to use the Wells scores for inpatients. Fourth, scores with a more extensive validation, such as Wells scores, Geneva and Revised Geneva scores, should be preferred in common clinical practice.

Addendum

Author Contributions: M. Righini, E. Ceriani and C. Combescure had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: M. Righini, H. Bounameaux, A. Perrier. Acquisition of data: E. Ceriani, M. Righini. Analysis and interpretation of data: M. Righini, E. Ceriani, C. Combescure, G. Le Gal, T. Perneger, H. Bounameaux, A. Perrier. Drafting of the manuscript: M. Righini, E. Ceriani, C. Combescure. Critical revision of the manuscript for important intellectual content: G. Le Gal, M. Nendaz, T. Perneger, H. Bounameaux, A. Perrier, M. Righini. Statistical analysis: C. Combescure, T. Perneger.

Disclosure of Conflicts of Interests

The authors state that they have no conflict of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Forests plots for the scores with 3 levels.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

1 Perrier A, Roy PM, Aujesky D, Chagnon I, Howarth N, Gourdier AL, Leftheriotis G, Barghouth G, Cornuz J, Hayoz D, Bounameaux H. 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.

2 Kruip MJ, Slob MJ, Schijen JH, van der Heul C, Buller HR. Use of a clinical decision rule in combination with D-dimer concentration in diagnostic workup of patients with suspected pulmonary embolism: a prospective management study. *Arch Intern Med* 2002; **162**: 1631–5.

3 Musset D, Parent F, Meyer G, Maitre S, Girard P, Leroyer C, Revel MP, Carette MF, Laurent M, Charbonnier B, Laurent F, Mal H, Nonent M, Lancar R, Grenier P, Simonneau G. Diagnostic strategy for patients with suspected pulmonary embolism: a prospective multicentre outcome study. *Lancet* 2002; **360**: 1914–20.

4 The PIOPED Investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). *JAMA* 1990; **263**: 2753–9.

5 Wasson JH, Sox HC. Clinical prediction rules. Have they come of age?. *JAMA* 1996; **275**: 641–2.

6 Wyatt JC, Altman AD. Prognostic models: clinically useful or quickly forgotten? *BMJ* 1995; **311**: 1539–41.

7 Hoellerich VL, Wigton RS. Diagnosing pulmonary embolism using clinical findings. *Arch Intern Med* 1986; **146**: 1699–704.

8 Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; **3**: 25.

9 Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1994; **271**: 389–91.

10 Stuart A, Ord JK. *Kendall's Advanced Theory of Statistics*.Vol. 1, *Distribution Theory*, 6th edn. London: Edward Arnold, 1994.
11 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**: 557–60.
12 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**: 177–88.
13 Wells PS, Anderson DR, Rodger M, Stiell I, Dreyer JF, Barnes D, Forgie M, Kovacs G, Ward J, Kovacs MJ. Excluding pulmonary embolism at the bedside without diagnostic imaging: Management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and D-dimer. *Ann Intern Med* 2001; **135**: 98–107.
14 Sanson BJ, Lijmer JG, MacGillavry MR, Turkstra F, Prins MH, Buller HR. 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.
15 Kline JA, Runyon MS, Webb WB, Jones AE, Mitchell AM. Prospective study of the diagnostic accuracy of the simplify D-dimer assay for pulmonary embolism in emergency department patients. *Chest* 2006; **129**: 1417–23.
16 Miniati M, Bottai M, Monti S. Comparison of 3 clinical models for predicting the probability of pulmonary embolism. *Medicine (Baltimore)* 2005; **84**: 107–14.
17 Kabrhel C, McAfee AT, Goldhaber SZ. The contribution of the subjective component of the Canadian Pulmonary Embolism Score to the overall score in emergency department patients. *Acad Emerg Med* 2005; **12**: 915–20.
18 Penaloza A, Melot C, Dochy E, Blocklet D, Gevenois PA, Wautrecht JC, Lheureux P, Motte S. Assessment of pretest probability of pulmonary embolism in the emergency department by physicians in training using the Wells model. *Thromb Res* 2007; **120**: 173–9.
19 Wolf SJ, McCubbin TR, Feldhaus KM, Faragher JP, Adcock DM. Prospective validation of wells criteria in the evaluation of patients with suspected pulmonary embolism. *Ann Emerg Med* 2004; **44**: 503–10.
20 van Belle A, Buller HR, Huisman MV, Huisman PM, Kaasjager K, Kamphuisen PW, Kramer MH, Kruip MJ, Kwakkel-van Erp JM, Leebeek FW, Nijkeuter M, Prins MH, Sohne M, Tick LW. Effectiveness of managing suspected pulmonary embolism using an algorithm combining clinical probability, D-dimer testing, and computed tomography. *JAMA* 2006; **295**: 172–9.
21 Perrier A, Roy PM, Sanchez O, Le Gal G, Meyer G, Gourdier AL, Furber A, Revel MP, Howarth N, Davido A, Bounameaux H. Multidetector-row computed tomography in suspected pulmonary embolism. *N Engl J Med* 2005; **352**: 1760–8.
22 Aujesky D, Hayoz D, Yersin B, Perrier A, Barghouth G, Schnyder P, Bischof-Delaloye A, Cornuz J. Exclusion of pulmonary embolism using C-reactive protein and D-dimer: A prospective comparison. *Thromb Haemost* 2003; **90**: 1198–203.
23 Ollenberger GP, Worsley DF. Effect of patient location on the performance of clinical models to predict pulmonary embolism. *Thromb Res* 2006; **118**: 685–90.
24 Le Gal G, Righini M, Roy PM, Sanchez O, Aujesky D, Bounameaux H, Perrier A. Prediction of pulmonary embolism in the emergency department: The revised Geneva score. *Ann Intern Med* 2006; **144**: 165–71.
25 Righini M, Le Gal G, Aujesky D, Roy PM, Sanchez O, Verschuren F, Rutschmann O, Nonent M, Cornuz J, Thys F, Le Manach CP, Revel MP, Poletti PA, Meyer G, Mottier D, Perneger T, Bounameaux H, Perrier A. 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.
26 Klok FA, Kruisman E, Spaan J, Nijkeuter M, Righini M, Aujesky D, Roy PM, Perrier A, Le Gal G, Huisman MV. Comparison of the revised Geneva score with the Wells rule for assessing clinical probability of pulmonary embolism. *Journal of Thrombosis and Haemostasis* 2008; **6**: 40–4.
27 Kline JA, Nelson RD, Jackson RE, Courtney DM. 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.
28 Bosson JL, Barro C, Satger B, Carpentier PH, Polack B, Pernod G. Quantitative high D-dimer value is predictive of pulmonary embolism occurrence independently of clinical score in a well-defined low risk factor population. *Journal of Thrombosis and Haemostasis* 2005; **3**: 93–9.
29 Wells PS, Anderson DR, Rodger M, Ginsberg JS, Kearon C, Gent M, Turpie AGG, Bormanis J, Weitz J, Chamberlain M, Bowie D, Barnes D, Hirsh J. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: Increasing the models utility with the SimpliRED D-dimer. *Thromb Haemost* 2000; **83**: 416–20.
30 Wicki J, Perneger TV, Junod AF, Bounameaux H, Perrier A. Assessing clinical probability of pulmonary embolism in the emergency ward: a simple score. *Arch Intern Med* 2001; **161**: 92–7.
31 McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA* 2000; **284**: 79–84.
32 Kruip MJ, Leclercq MG, van der Heul C, Prins MH, Buller HR. Diagnostic strategies for excluding pulmonary embolism in clinical outcome studies. A systematic review. *Ann Intern Med* 2003; **138**: 941–51.
33 Torbicki A, Perrier A, Konstantinides S, Agnelli G, Galie N, Pruszczyk P, Bengel F, Brady AJ, Ferreira D, Janssens U, Klepetko W, Mayer E, Remy-Jardin M, Bassand JP, Vahanian A, Camm J, De Caterina R, Dean V, Dickstein K, Filippatos G, *et al*. Guidelines on the diagnosis and management of acute pulmonary embolism: the Task Force for the Diagnosis and Management of Acute Pulmonary Embolism of the European Society of Cardiology (ESC). *Eur Heart J* 2008; **29**: 2276–315.
34 Kearon C, Kahn SR, Agnelli G, Goldhaber S, Raskob GE, Comerota AJ. Antithrombotic therapy for venous thromboembolic disease: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 2008; **133**: S454–545.
35 Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 1999; **354**: 1896–900.
36 Runyon MS, Richman PB, Kline JA. Emergency medicine practitioner knowledge and use of decision rules for the evaluation of patients with suspected pulmonary embolism: variations by practice setting and training level. *Acad Emerg Med* 2007; **14**: 53–7.
37 Roy PM, Meyer G, Vielle B, Le Gall C, Verschuren F, Carpentier F, Leveau P, Furber A. Appropriateness of diagnostic management and outcomes of suspected pulmonary embolism. *Ann Intern Med* 2006; **144**: 157–64.
38 Righini M, Perrier A, De Moerloose P, Bounameaux H. D-Dimer for venous thromboembolism diagnosis: 20 years later. *J Thromb Haemost* 2008; **6**: 1059–71.
39 Klok FA, Mos IC, Nijkeuter M, Righini M, Perrier A, Le Gal G, Huisman MV. Simplification of the revised Geneva score for assessing clinical probability of pulmonary embolism. *Arch Intern Med* 2008; **168**: 2131–6.
40 Gibson NS, Sohne M, Kruip MJHA, Tick LW, Gerdes VE, Bossuyt PM, Wells PS, Buller HR. Further validation and simplification of the Wells clinical decision rule in pulmonary embolism. *Thromb Haemost* 2008; **99**: 229–34.
41 Goekoop RJ, Steeghs N, Niessen RW, Jonkers GJ, Dik H, Castel A, Werker-van Gelder L, Vlasveld LT, van Klink RC, Planken EV, Huisman MV. Simple and safe exclusion of pulmonary embolism in outpatients using quantitative D-dimer and Wells' simplified decision rule. *Thromb Haemost* 2007; **97**: 146–50.
42 Kearon C, Ginsberg JS, Douketis J, Turpie AG, Bates SM, Lee AY, Crowther MA, Weitz JI, Brill-Edwards P, Wells P, Anderson DR,

Kovacs MJ, Linkins LA, Julian JA, Bonilla LR, Gent M. An evaluation of D-dimer in the diagnosis of pulmonary embolism: a randomized trial. *Ann Intern Med* 2006; **144**: 812–21.

43 Kline JA, Hogg M. Measurement of expired carbon dioxide, oxygen and volume in conjunction with pretest probability estimation as a method to diagnose and exclude pulmonary venous thromboembolism. *Clin Physiol Funct Imaging* 2006; **26**: 212–9.

44 Steeghs N, Goekoop RJ, Niessen RW, Jonkers GJ, Dik H, Huisman MV. C-reactive protein and D-dimer with clinical probability score in the exclusion of pulmonary embolism. *Br J Haematol* 2005; **130**: 614–9.

45 Runyon MS, Webb WB, Jones AE, Kline JA. Comparison of the unstructured clinician estimate of pretest probability for pulmonary embolism to the Canadian score and the Charlotte rule: A prospective observational study. *Acad Emerg Med* 2005; **12**: 587–93.

46 Kline JA, Webb WB, Jones AE, Hernandez-Nino J. Impact of a rapid rule-out protocol for pulmonary embolism on the rate of screening, missed cases, and pulmonary vascular imaging in an urban US emergency department. *Ann Emerg Med* 2004; **44**: 490–502.

47 Miniati M, Monti S, Bauleo C, Scoscia E, Tonelli L, Dainelli A, Catapano G, Formichi B, Di Ricco G, Prediletto R, Carrozzi L, Marini C. A diagnostic strategy for pulmonary embolism based on standardised pretest probability and perfusion lung scanning: a management study. *Eur J Nucl Med Mol Imaging* 2003; **30**: 1450–6.

48 Miniati M, Monti S, Bottai M. A structured clinical model for predicting the probability of pulmonary embolism. *Am J Med* 2003; **114**: 173–9.