Comparison of the Unstructured Clinician Gestalt, the Wells Score, and the Revised Geneva Score to Estimate Pretest Probability for Suspected Pulmonary Embolism

Andrea Penaloza, MD; Franck Verschuren, MD, PhD; Guy Meyer, MD; Sybille Quentin-Georget, MD; Caroline Soulie, MD; Frédéric Thys, MD, PhD; Pierre-Marie Roy, MD, PhD

Study objective: The assessment of clinical probability (as low, moderate, or high) with clinical decision rules has become a cornerstone of diagnostic strategy for patients with suspected pulmonary embolism, but little is known about the use of physician gestalt assessment of clinical probability. We evaluate the performance of gestalt assessment for diagnosing pulmonary embolism.

Methods: We conducted a retrospective analysis of a prospective observational cohort of consecutive suspected pulmonary embolism patients in emergency departments. Accuracy of gestalt assessment was compared with the Wells score and the revised Geneva score by the area under the curve (AUC) of receiver operating characteristic curves. Agreement between the 3 methods was determined by κ test.

Results: The study population was 1,038 patients, with a pulmonary embolism prevalence of 31.3%. AUC differed significantly between the 3 methods and was 0.81 (95% confidence interval [CI] 0.78 to 0.84) for gestalt assessment, 0.71 (95% CI 0.68 to 0.75) for Wells, and 0.66 (95% CI 0.63 to 0.70) for the revised Geneva score. The proportion of patients categorized as having low clinical probability was statistically higher with gestalt than with revised Geneva score (43% versus 26%; 95% CI for the difference of 17%=13% to 21%). Proportion of patients categorized as having high clinical probability was higher with gestalt than with Wells (24% versus 7%; 95% CI for the difference of 17%=14% to 20%) or revised Geneva score (24% versus 10%; 95% CI for the difference of 15%=13% to 21%). Pulmonary embolism prevalence was significantly lower with gestalt versus clinical decision rules in low clinical probability (7.6% for gestalt versus 13.0% for revised Geneva score and 12.6% for Wells score) and non–high clinical probability groups (18.3% for gestalt versus 29.3% for Wells and 27.4% for revised Geneva score) and was significantly higher with gestalt versus Wells score in high clinical probability groups (72.1% versus 58.1%). Agreement between the 3 methods was poor, with all κ values below 0.3.

Conclusion: In our retrospective study, gestalt assessment seems to perform better than clinical decision rules because of better selection of patients with low and high clinical probability. [Ann Emerg Med. 2013;xx:xxx.]

Please see page XX for the Editor's Capsule Summary of this article.

0196-0644/$-see front matter
Copyright © 2013 by the American College of Emergency Physicians.
http://dx.doi.org/10.1016/j.annemergmed.2012.11.002

SEE EDITORIAL, P. XXX.

INTRODUCTION

Background

With an estimated annual incidence of 70 cases per 100,000 population,[1,2] pulmonary embolism is a frequent, potentially life-threatening and difficult-to-diagnose condition. Clinical signs and symptoms are common and neither specific nor sensitive when used individually.[3,4] Nevertheless, when these are combined either with implicit physician assessment (gestalt or global clinical judgment) or with the use of a clinical decision rule, they allow the categorization of patients with suspected pulmonary embolism into pretest probability groups with increasing pulmonary embolism prevalence.

Importance

The most accurate way to assess clinical probability remains unknown. The accuracy of gestalt assessment has been confirmed in several studies[5-8] but has been criticized because it is not standardized, leading to development of explicit clinical decision rules. The Wells score and the revised Geneva score are probably the most validated clinical decision rules (Table 1).[9,10] Both are standardized. Whereas the revised Geneva score is fully explicit, the Wells score includes 6 explicit criteria and 1 subjective criterion, the physician's judgment of whether an

Editor's Capsule Summary

What is already known on this topic

Several clinical decision rules have been used to classify risk for pulmonary embolism. Most studies of clinical decision rules have not compared them with physician gestalt.

What question this study addressed

Do treating physicians assess pulmonary embolism risk as accurately as 2 widely used clinical decision rules?

What this study adds to our knowledge

In this retrospective analysis of 1,038 patients with suspected pulmonary embolism treated in 117 European emergency departments, the physicians identified low- and high-probability patients more effectively than the clinical decision rules.

How this is relevant to clinical practice

This article provides evidence that clinical decision rules, despite their current popularity, may not be as good as traditional physician judgment and also emphasizes that all research on clinical decision rules should test them against physician judgment.

alternative diagnosis is less likely than pulmonary embolism. Clinical decision rules are used in almost all recent diagnostic studies about pulmonary embolism but have not been proven to perform better than gestalt assessment.[5-8] Moreover, most emergency physicians still use gestalt assessment for suspected pulmonary embolism, but little is known about the respective accuracy and agreement of clinical decision rules and gestalt assessment for diagnosing pulmonary embolism.[11]

Goals of This Investigation

We hypothesized that gestalt assessment could compare favorably with clinical decision rules in the management of pulmonary embolism suspicion in emergency medicine. Therefore, we used the data of a large multicenter diagnostic study to compare performances and agreements of gestalt assessment, Wells score, and revised Geneva score.

MATERIALS AND METHODS

Selection of Participants

This study is based on the retrospective analysis of a prospective cohort study designed to assess the appropriateness of diagnostic criteria used in routine practice to rule in or rule out pulmonary embolism in 116 emergency departments (EDs) in France and 1 in Belgium (n=1,529).[12] During a 5-week period, consecutive patients presenting to ED with clinical suspicion of pulmonary embolism and for whom any diagnostic testing for this suspicion was performed (including D-dimer) were included. Patients were excluded if (1) the diagnosis of thromboembolic disease was documented before admission; (2) pulmonary embolism was suspected among inpatients (hospital stay of more than 2 days' duration; or (3) diagnostic testing was cancelled for ethical reasons, because of rapid death, or because the patient decided to leave the hospital against medical advice or declined testing. Patients were managed by a large number of emergency physicians with heterogeneous training level, including (as in daily practice) young postgraduates having various levels of experience and confirmed emergency physicians. Physicians prospectively completed a standardized form to report patients' characteristics and their own diagnostic hypotheses, and before any diagnostic testing they were invited to give their gestalt assessment to estimate the clinical probability of pulmonary embolism (low, intermediate, or high). All clinical data used in Wells score and revised Geneva score were recorded in clinical record form but the physicians were not forced to calculate these scores. They were free to manage patients as usual, according to local practice. How patients were handled as regards the exclusion or confirmation of pulmonary embolism was presented and discussed in the original study.[12]

Patients, relatives, or their general practitioners were interviewed at the end of a 3-month follow-up period about the possible occurrence of a venous thromboembolic event, and all suspected events or deaths caused by pulmonary embolism were assessed by a central adjudication committee, constituted by a panel of 3 experts.[12]

Table 1. Clinical decision rules.

Score	Points
Wells score*	
Clinical signs of DVT (edema and pain)	3.0
Alternative diagnosis less likely than PE	3.0
Pulse rate >100 beats/min	1.5
Recent surgery or immobilization (within 4 wk)	1.5
Previous PE or DVT	1.5
Hemoptysis	1.0
Cancer	1.0
Revised Geneva score†	
Age >65 y	1
Previous PE or DVT	3
Surgery (under general anesthesia) or fracture (lower limb) within 4 wk	2
Active malignancy	2
Unilateral lower limb pain	3
Hemoptysis	2
Pulse rate 75–94 beats/min	3
Pulse rate ≥95 beats/min	5
Pain on lower-limb deep vein palpation and unilateral edema	4

DVT, Deep venous thrombosis; *PE*, pulmonary embolism.
*Greater than 2: low clinical probability; 2 to 6: moderate clinical probability; greater than 6: high clinical probability.
†Zero to 3: low clinical probability; 4 to 10: moderate clinical probability; greater than or equal to 11: high clinical probability.

Data Collection and Processing

For the present study, we collected the prospectively documented clinical gestalt assessment and retrospectively calculated the Wells score and the revised Geneva score. Three assessment methods categorized clinical probability of patients suspected of having pulmonary embolism in 3 levels: low, moderate, and high. We considered as a final diagnosis of pulmonary embolism (1) a pulmonary embolism diagnosis ruled in at the end of the initial diagnostic evaluation; (2) a thromboembolic event (pulmonary embolism or deep venous thrombosis) occurring during follow-up among patients for whom the diagnosis of pulmonary embolism was initially ruled out; or (3) death adjudicated as related or possibly related to pulmonary embolism.

The original study was designed to measure the appropriateness of diagnostic criteria used in routine practice to rule in or rule out pulmonary embolism. Results showed that for confirmed pulmonary embolism, diagnostic criteria were appropriate according to guidelines in 92% of cases, whereas they were inappropriate in 43% of cases for excluded pulmonary embolism. To quantify a possible misclassification bias, besides the main analysis we performed a subgroup analysis comparing the performances of the 3 assessment methods in patients for whom diagnostic strategies were validated as appropriate in the original study. In this subgroup analysis, we included only confirmed pulmonary embolism documented by a positive angiography result, a positive computed tomography (CT) scan result, a high-probability ventilation/perfusion scan result or a positive ultrasonographic result during initial evaluation, and the documented thromboembolic events or sudden deaths probably related to pulmonary embolism during follow-up.

All statistical analyses were performed with SPSS (version 15.0; SPSS, Inc., Chicago, IL). Accuracy of categorization of patients into pretest clinical probability groups (low, moderate, and high) was assessed by calculating pulmonary embolism prevalence. The accuracy of the clinical probability assessment methods was compared by using the AUC of receiver operating characteristic curves.[13] The receiver operating characteristic curve illustrates the performance of a binary classifier system (pulmonary embolism versus no pulmonary embolism) as the discrimination threshold of a continuous or discrete variable. In our study, because of 3-level categorization of gestalt assessment, we used discrete 3-level variables (1 low, 2 moderate, and 3 high). For the 3 methods (gestalt, Wells, and revised Geneva score), we compared AUC in receiver operating characteristic analyses of the 3-level classification scheme. This method was used for pulmonary embolism clinical probability assessment methods by several authors.[14-16] Agreement between clinical assessment methods was determined by κ test. The statistical analysis consisted of a comparison by the χ^2 test (for categorical variables) or Mann-Whitney *U* test (for continuous variables). Significance was accepted at $P<.05$.

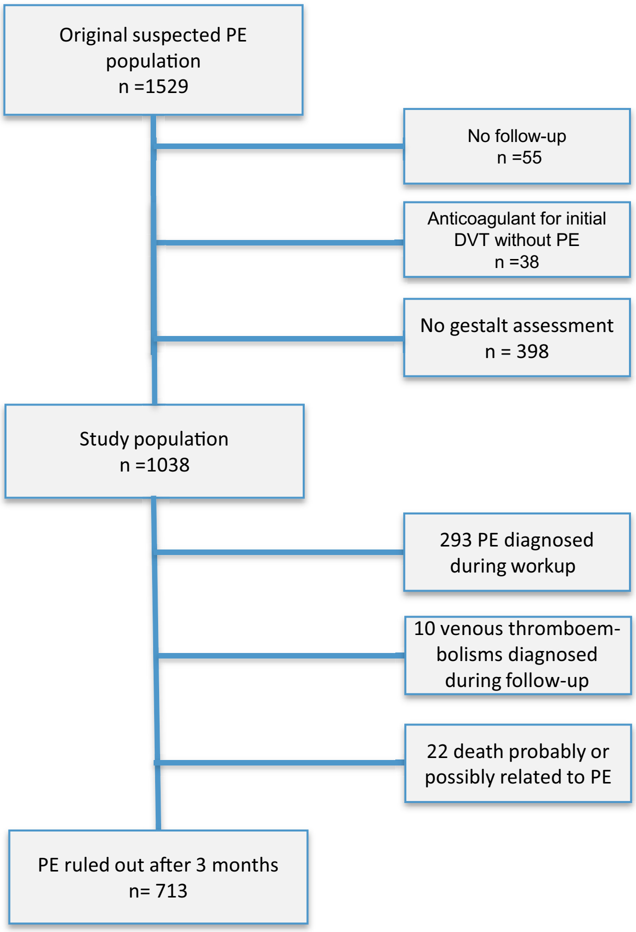

Figure 1. Study population.

RESULTS

In the total population (n=1,529), we excluded patients for whom the follow-up was not obtained (n=55) and patients receiving anticoagulation for an initial diagnosis of deep venous thrombosis without pulmonary embolism (n=38). In the remaining population, the clinical probability assessment was available for 1,038 patients with the 3 methods (Figure 1).

The baseline characteristics of the study population are shown in Table 2. The prevalence of pulmonary embolism in the study population was 31.3%. Table 3 summarizes the method by which the final pulmonary embolism diagnosis was confirmed. Table 4 shows the proportion of patients classified in each category and pulmonary embolism prevalence in each level of clinical probability for the 3 assessment methods. Each clinical probability assessment method was discriminant for patient categorization because pulmonary embolism prevalence significantly differed between the 3 groups of clinical probability. Nevertheless, there were differences between assessment methods in pulmonary embolism prevalence in each clinical probability group. Pulmonary embolism prevalence in patients with low clinical probability was significantly lower

Table 2. Baseline characteristics of the study population.*

Patient Characteristics	Excluded Population, n=491	Study Population, n=1,038
Mean age, y (95% CI)	65 (63–67)	64 (63–65)
Female sex	288 (58.7; 54.3–63.0)	644 (62.0; 59.1–65.0)
Cancer	41 (8.4; 6.1–11.1)	69 (6.6; 5.3–8.3)
Personal history VTE	118 (24.0; 20.4–28.0)	237 (22.8; 20.4–25.5)
Past surgery <1 mo	27 (5.5; 3.7–7.8)	39 (3.8; 2.7–5.1)
Fracture	11 (2.2; 1.2–3.9)	31 (3.0; 2.1–4.2)
Palpation pain and lower limb edema	68 (13.8; 11.0–17.1)	136 (13.1; 11.2–15.3)
Final PE	149 (30.4; 26.4–34.5)	325 (31.3; 28.5–34.2)

VTE, Venous thromboembolism.
*Data are presented as percentage (95% CI) unless otherwise indicated.

Table 3. Diagnostic method for 325 final pulmonary embolism diagnoses.

Final PE (n=325)	n
CT	173
High probability V/Q lung scan	63
Proximal DVT on leg US	37
Angiography	1
Echocardiography and high CP	5
VTE during FU	10
Death during FU possibly or probably PE related	22
Inappropriate strategy	14

CP, Clinical probability; *FU*, follow-up.

with gestalt assessment than the revised Geneva score (7.6% versus 13.0%; 95% confidence interval [CI] for the difference of 5.3%=0.6% to 10.0%) than with the Wells score (7.6% versus 12.6%; 95% CI for the difference of 4.9%=1.1% to 8.8%). Pulmonary embolism prevalence in patients with high clinical probability was higher with gestalt assessment than with the Wells score (72.1% versus 58.1%; 95% CI for the difference of 14.0%=1.5% to 26.5%) and nonsignificantly higher than with the revised Geneva score (68.7%). The prevalence of pulmonary embolism in the non–high probability group (combining low and moderate) was significantly lower with gestalt assessment than with Wells score (18.3% versus 29.3%; 95% CI for the difference of 11.0%=7.0% to 14.9%) or revised Geneva score (18.3% versus 27.4%; 95% CI for the difference of 9.1%=5.1% to 13.0%). Moreover, the proportion of patients categorized in each clinical probability group differed among methods. Proportion of patients categorized as having low clinical probability was similar with gestalt assessment (43%) and the Wells score (47%) but was larger with the gestalt assessment than with revised Geneva score (43% versus 26%; 95% CI for the difference of 17%=13% to 21%). The proportion of patients categorized as having high clinical probability was higher with gestalt assessment than with Wells score (24% versus 7%; 95% CI for the difference of 17%=14% to 20%) or revised Geneva score (24% versus 10%; 95% CI for the difference of 15%=13% to 21%). The AUC was 0.81 (95% CI 0.78 to 0.84), 0.71 (95% CI 0.68 to 0.75), and 0.66 (95% CI 0.63 to 0.70) for gestalt assessment and Wells and revised Geneva score, respectively (Figure 2). The same analysis was performed in a subgroup of patients (n=651) for whom diagnostic strategies were validated as appropriate in the original study. The results of this analysis were in line with those of the main analysis, showing a significantly higher AUC for gestalt assessment than for clinical decision rules: AUC of 0.89 (95% CI 0.87 to 0.92) for gestalt, 0.76 (95% CI 0.72 to 0.79) for Wells score, and 0.72 (95% CI 0.68 to 0.76) for revised Geneva score (see detailed results in Appendix E1, available online at http://www.annemergmed.com).

Table 5 shows overall agreement in risk stratification between the 3 assessment methods. Agreement was poor, with κ coefficients of 0.21 (95% CI 0.17 to 0.25), 0.25 (95% CI 0.20 to 0.29), and 0.26 (95% CI 0.21 to 0.31) for gestalt versus revised Geneva score, gestalt versus Wells score, and revised Geneva score versus Wells score, respectively. The percentage of discordant assessments was 52%, 48%, and 45% for gestalt versus revised Geneva score, gestalt versus Wells score, and revised Geneva score versus Wells score, respectively. Major disagreement (patients classified in the low clinical probability group by one method and in the high clinical probability group by another) occurred in 3% in gestalt versus revised Geneva score, 4% in gestalt versus Wells score, and 1% in revised Geneva score versus Wells score.

LIMITATIONS

Our study has some limitations. First, it was a secondary retrospective analysis of a prospective study, which was not designed or powered to compare clinical probability assessment methods. The original study was designed to measure the appropriateness of diagnostic criteria used in routine practice to rule in or rule out pulmonary embolism, ie, whether physicians in EDs routinely used evidence-based diagnostic criteria for managing patients with suspected pulmonary embolism. Physicians with various levels of training were asked to manage patients as usual and did not systematically assess the clinical probability. Gestalt assessment was missing in 25% of the patients who were excluded from the present study. Nevertheless, baseline characteristics and pulmonary embolism prevalence (Table 2) were similar between excluded and included patients. Because diagnostic criteria were not standardized, some pulmonary embolism may have been over- or underdiagnosed, and the exact incidence of pulmonary embolism remains unknown. However, the risk was shared by all 3 clinical probability assessment methods (gestalt, Wells, or revised Geneva score). Moreover, in the initial study 92% of confirmed pulmonary embolisms were appropriately diagnosed, whereas diagnostic criteria were inappropriate in 43% of excluded pulmonary embolism. Latter possible bias was partially resolved by the 3-month follow-up, allowing us to recover most of the clinically relevant thromboembolic events in patients initially considered as not having pulmonary embolism.

Table 4. Comparison of assessment methods: proportion of patient categorized and pulmonary embolism prevalence in each clinical probability level.

CP Assessment (n=1,038) (%, 95% CI)	Gestalt	Wells Score	RGS
Number of patients			
Low	445 (43; 40–46)	486 (47; 44–50)	270 (26; 23–29)
Moderate	342 (33; 30–36)	478 (46; 43–49)	669 (65; 62–67)
High	251 (24; 22–27)	74 (7; 6–9)	99 (10; 8–11)
Number of pulmonary embolisms			
Low	34 (7.6; 5.4–10.4)	61 (12.6; 9.8–15.4)	35 (13; 9.3–17.4)
Moderate	110 (32.2; 27.4–37.3)	221 (42.6; 41.8–50.7)	222 (33.2; 29.7–36.8)
High	181 (72.1; 66.3–77.4)	43 (58.1; 46.7–68.9)	68 (68.7; 59.1–77.2)

CP, Clinical probability; *RGS*, revised Geneva score.

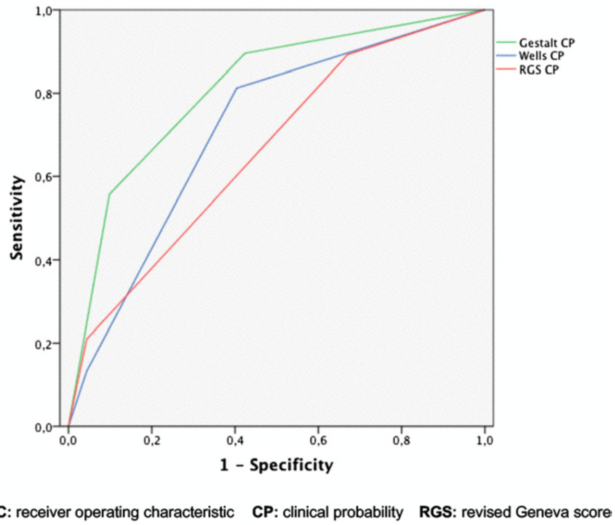

Figure 2. Comparison of the predictive accuracy of the 3 clinical probability assessment methods in patients with suspected pulmonary embolism, using clinical probability groups (low, moderate, and high).

As is usual in outcome studies on pulmonary embolism, all patients with pulmonary embolism ruled out during initial diagnostic evaluation who experience a venous thromboembolic episode during follow-up were considered as having pulmonary embolism. Nevertheless, we cannot exclude that some patients were misclassified in the present analysis. To address this potential bias, we performed a subgroup analysis in patients (n=651) for whom the diagnosis of pulmonary embolism was based on validated methods and obtained results in line with those of the main analysis (Appendix E1, available online at http://www.annemergmed.com).

Second, clinical decision rules were retrospectively calculated. All the criteria were prospectively collected, allowing us such analyses. However, concerning the Wells score, the likelihood of an alternative diagnosis was prospectively assessed but not used in the diagnostic strategy, which could result in misevaluation of this criterion.

Third, the overall prevalence in our study was high (31.3%), as classically described in European studies.[17-19] Our results were obtained from an unselected population issuing from a large number of EDs (117) and managed as in daily practice. Therefore, our results are likely to provide an accurate sample of the European population, in which pulmonary embolism prevalence among suspected patients is high. Application of these results to other emergency medicine practices, in particular in an area with low pulmonary embolism prevalence such as North America, could be questionable. In a recent study comparing populations with suspected pulmonary embolism from Europe and the United States, we suggested that pulmonary embolism prevalence differences observed among patients with suspected pulmonary embolism in EDs between the 2 continents could be explained by differences in physician suspicion threshold.[20] Prevalence differences may reflect a lower suspicion threshold in the United States than in Europe. However, in North American studies with low pulmonary embolism prevalence, the accuracy of gestalt assessment also compares favorably with the Wells score, and physicians use gestalt assessment more often than a clinical decision rule in their daily practice.[9,21] In addition, gestalt has some advantages: it is always available, can be used in addition to clinical decision rules, and allows consideration of all clinical signs and symptoms in more detail than clinical decision rules. Nevertheless, this needs to be confirmed in a prospective study including EDs with a low prevalence of pulmonary embolism among suspected cases.

DISCUSSION

To our knowledge, is the first study comparing gestalt to the revised Geneva score and the Wells score in a large population. Our results confirm that gestalt compares favorably with clinical decision rules for estimating the clinical probability of pulmonary embolism. Even if the 3 methods were discriminant for patient categorization into clinical probability groups, gestalt performance appears better than that of the Wells and revised Geneva scores in our study, mainly because of differences in proportion and pulmonary embolism prevalence in the low and high clinical probability groups.

Table 5. Agreement on the 3 clinical probability assessment methods in patients with suspected pulmonary embolism, compared 2 by 2.

		RGS			
PE Prevalence %, n	**Probability**	**Low**	**Moderate**	**High**	**Total**
Gestalt assessment*	**Low**	188 (4.3)	246 (9.8)	11 (18.2)	445 (7.6)
	Moderate	62 (24.2)	253 (31.6)	27 (55.6)	342 (32.2)
	High	20 (60.0)	170 (69.4)	61 (83.6)	251 (72.1)
	Total	270 (13.0)	669 (33.2)	99 (68.7)	1,038
		Gestalt Assessment			
PE Prevalence %, n	**Probability**	**Low**	**Moderate**	**High**	**Total**
Wells†	**Low**	314 (4.8)	137 (19.0)	35 (57.1)	486 (12.6)
	Moderate	120 (15.8)	186 (40.9)	172 (73.3)	478 (42.6)
	High	11 (0)	19 (42.1)	44 (79.5)	74 (58.1)
	Total	445 (7.6)	342 (32.2)	251 (72.1)	1,038
		Wells			
PE Prevalence %, n	**Probability**	**Low**	**Moderate**	**High**	**Total**
RGS‡	**Low**	192 (5.2)	73 (34.2)	5 (0)	270 (13.0)
	Moderate	286 (15.4)	360 (47.2)	23 (34.8)	669 (33.2)
	High	8 (87.6)	45 (57.8)	46 (76.1)	99 (68.7)
	Total	486 (12.6)	478 (42.6)	74 (58.1)	1,038

RGS, Revised Geneva score.
*κ Coefficient=0.21 (95% CI 0.17 to 0.25).
†κ Coefficient=0.25 (95% CI 0.20 to 0.29).
‡κ Coefficient=0.26 (95% CI 0.21 to 0.31).

Comparison of pulmonary embolism prevalence in each category of clinical probability assessed by gestalt, Wells score, and revised Geneva score showed that the 3 methods meaningfully categorize patients with suspected pulmonary embolism into clinical probability groups. These results are concordant with 2 recent meta-analyses in which accuracy of different assessment methods was shown to be similar by comparing pulmonary embolism prevalence in each category of clinical probability.[20,22] Nevertheless, the receiver operating characteristic curve approach is a more appropriate technique to determine the potential performance of a score.[13,23] In our study, AUCs differ significantly, increasing from revised Geneva score to Wells score and to gestalt assessment. Increase in AUCs appears to be parallel to the integration of subjective criteria in the assessment method. Indeed, revised Geneva score is fully explicit, the Wells score has a subjective criterion, and gestalt assessment is completely based on global clinical judgment (ie, implicit). These results are concordant with those of previous studies showing that the subjective criterion (physician's judgment of whether an alternative diagnosis is less likely than pulmonary embolism) is the most informative criterion of the Wells score[21,24] and that the score performs better than the simplified revised Geneva score.[25]

Gestalt assessment categorizes a larger proportion of patients as having low clinical probability than the revised Geneva score and a larger proportion of patients as having high clinical probability than both clinical decision rules. Moreover, the prevalence of pulmonary embolism in the low clinical probability group was lower when gestalt assessment was used than when clinical decision rules were used. When low and moderate probability in the non–high probability group was combined, the prevalence of pulmonary embolism was also lower when gestalt assessment was used than when both clinical decision rules were used. These results suggest that gestalt assessment performs better in selecting nonhigh (low and moderate) and high pretest probability patients than standardized clinical decision rules. This is an important point in the diagnostic approach of pulmonary embolism. Indeed, lower pulmonary embolism prevalence in the low clinical probability and non–high clinical probability group should lead to a higher proportion of patients with normal D-dimer levels in whom pulmonary embolism could be excluded without further examinations.[26,27] Conversely, patients with high clinical probability should have not a D-dimer test but a CT scan directly and receive an anticoagulant treatment and await results.[7,28] This would reduce time-wasting in our crowded EDs. Indeed, classifying such subgroups of patients as having low or moderate clinical probability leads to performance of a D-dimer test, whose results most of the time will be positive, requiring a subsequent CT scan. Moreover, a negative D-dimer test result in this group with high pulmonary embolism prevalence carries a theoretical high risk of false-negative test results.[29] As predicted by Bayes' theorem, the posttest probability and the risk of false test results is highly dependent on the baseline prevalence of disease.

The better performance of gestalt over clinical decision rules is not surprising. Clinical decision rules are simplified models constructed from most frequent dichotomized or categorized data. Conversely, gestalt assessment allows us to consider and integrate all clinical signs and symptoms in more detail; for example, a history of venous thromboembolism can be weighted by its spontaneous or provoked character. Such assessment can also take into account rare or moderate risk factors for pulmonary embolism; for example, long air travel.[30]

Gestalt assessment has been criticized for its lack of standardization and the difficulty in teaching it.[7] However, Runyon et al[31] showed that the accuracy of gestalt assessment remains good independently of physician's training level and compares favorably with the Wells score, even in terms of interobserver agreement (κ 0.6 versus 0.47). The same authors observed that if a majority of clinicians (68%) are familiar with a clinical decision rule for pulmonary embolism, only half of them use it in their daily practice. Conversely, 57% of physicians reported that they used gestalt assessment for suspected pulmonary embolism.[11] The specific work rhythm in EDs with high patient turnover may explain this preference. Our results confirm Runyon's work, suggesting that gestalt assessment may play a central role in pulmonary embolism diagnostic approach in EDs. Even if clinical decision rules are good didactic tools and useful ways in clinical research to homogenize clinical probability assessment among clinicians and centers, gestalt assessment seems the best way in everyday emergency medicine practice.

Our analysis needs to be weighted because some physicians (n=168) calculated one of the clinical decision rules (Wells or revised Geneva score) besides gestalt assessment, even if it was not required in the initial study. Therefore, some physicians might have been influenced by one of them. However, even for these patients, gestalt assessment performed better than clinical decision rules (see Appendix E2, available online at http://www.annemergmed.com). Moreover, comparison of assessment methods 2 by 2 showed quite poor agreement: none of the κ tests were above 0.3 (Table 5). Whatever the comparison made, the percentage of discordant assessment was approximately 50%, with major disagreements occurring in up to 4% of cases (most frequently between gestalt and clinical decision rule). In cases of disagreement, gestalt assessment appears more accurate than revised Geneva score or Wells score. For patients categorized as having low probability by gestalt assessment and as having high probability with a clinical decision rule, pulmonary embolism prevalence was 5.2% for Wells and 18.2% for revised Geneva score. For patients categorized as having high probability by gestalt assessment and as having low probability by one of the clinical decision rules, pulmonary embolism prevalence was greater than or equal to 55%. These results, and the fact that gestalt seemed to perform better than a clinical decision rule (even in patients for whom the physician reported calculating a clinical decision rule besides gestalt assessment; see Appendix E2, available online at http://www.annemergmed.com), led us to hypothesize that an overriding of clinical decision rules by a physician's implicit clinical judgment may improve clinical decision rules' performance. This override possibility was tested by Chagnon et al[32] on the original Geneva score, with a significant increase of AUC for the "overridden Geneva score" (0.82 versus 0.69 for Geneva score alone). This hypothesis seems promising and deserves to be specifically assessed in a larger prospective study.

Our results confirm that gestalt compares favorably with the Wells score and the revised Geneva score. In our retrospective study, there was poor agreement between the 3 methods. The overall performance of gestalt appeared to be superior to that of the 2 scores, with better selection of patients with a low or high clinical probability of pulmonary embolism. The improved performance seems to be related to subjective assessment, which allows the physician to take into account all signs, symptoms, and risk factors of patients with suspected pulmonary embolism. (Tables E1 and E2, Figures E1 and E2).

The authors acknowledge the Fondation Saint-Luc for providing a research grant to Dr. Penaloza.

Supervising editor: Allan B. Wolfson, MD

Author affiliations: From the Emergency Department, Cliniques Universitaires St-Luc, Université Catholique de Louvain, Belgium (Penaloza, Verschuren, Thys); the L'UNAM Université; the Emergency Department, CHU Angers, Université d'Angers, France (Penaloza, Quentin-Georget, Soulie, Roy); and the Department of Respiratory Diseases, Hôpital européen Georges Pompidou, Assistance Publique Hôpitaux de Paris, and the Université Paris Descartes, Sorbonne Paris Cité, Paris, France (Meyer).

Author contributions: AP and P-MR had full access to all data study, take responsibility for the integrity of the data and the accuracy of the data analysis, and were responsible for study concept and design. FV, GM, and P-MR were responsible for data acquisition. AP, FV, GM, SQ-G, CS, FT, and PM-R were responsible for analysis and interpretation of data, critical revision of the article for important intellectual content, and study supervision. AP, FV, GM, and P-MR were responsible for drafting the article. AP was responsible for statistical analysis. AP and PMR take responsibility for the paper as a whole.

Funding and support: By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist.

Publication dates: Received for publication May 24, 2012. Revisions received July 20, 2012; October 8, 2012; and October 27, 2012. Accepted for publication November 2, 2012.

Address for correspondence: Andrea Penaloza, MD, E-mail andrea.penaloza@uclouvain.be.

REFERENCES

1. Oger E. Incidence of venous thromboembolism: a community-based study in Western France. EPI-GETBP Study Group. Groupe d'Etude de la Thrombose de Bretagne Occidentale. *Thromb Haemost*. 2000;83:657-660.
2. Silverstein MD, Heit JA, Mohr DN, et al. Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study. *Arch Intern Med*. 1998;158:585-593.
3. Hildner FJ, Ormond RS. Accuracy of the clinical diagnosis of pulmonary embolism. *JAMA*. 1967;202:567-570.
4. West J, Goodacre S, Sampson F. The value of clinical features in the diagnosis of acute pulmonary embolism: systematic review and meta-analysis. *QJM*. 2007;100:763-769.
5. Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED). The PIOPED Investigators. *JAMA*. 1990;263:2753-2759.
6. Chunilal SD, Eikelboom JW, Attia J, et al. Does this patient have pulmonary embolism? *JAMA*. 2003;290:2849-2858.
7. Torbicki A, Perrier A, Konstantinides S, 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;29:2276-2315.
8. Roy P, Coroller-Bec C, Dambrine S, et al. Comparison of the unstructured clinician gestalt, the Wells score and the revised Geneva score to estimate the pretest probability for suspected pulmonary embolism. *J Thromb Haemost*. 2009;7(suppl 2): Abstract OC-MO-038.
9. Wells PS, Anderson DR, Rodger M, et al. 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;83:416-420.
10. Le Gal G, Righini M, Roy PM, et al. Prediction of pulmonary embolism in the emergency department: the revised Geneva score. *Ann Intern Med*. 2006;144:165-171.
11. 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;14:53-57.
12. Roy PM, Meyer G, Vielle B, et al. Appropriateness of diagnostic management and outcomes of suspected pulmonary embolism. *Ann Intern Med*. 2006;144:157-164.
13. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837-845.
14. Chagnon I, Bounameaux H, Aujesky D, et al. Comparison of two clinical prediction rules and implicit assessment among patients with suspected pulmonary embolism. *Am J Med*. 2002;113:269-275.
15. Klok FA, Mos IC, Nijkeuter M, et al. Simplification of the revised Geneva score for assessing clinical probability of pulmonary embolism. *Arch Intern Med*. 2008;168:2131-2136.
16. Douma RA, Mos IC, Erkens PM, et al. Performance of 4 clinical decision rules in the diagnostic management of acute pulmonary embolism: a prospective cohort study. *Ann Intern Med*. 2011;154:709-718.
17. Righini M, Goehring C, Bounameaux H, et al. Effects of age on the performance of common diagnostic tests for pulmonary embolism. *Am J Med*. 2000;109:357-361.
18. Perrier A, Roy PM, Aujesky D, et al. Diagnosing pulmonary embolism in outpatients with clinical assessment, D-dimer measurement, venous ultrasound, and helical computed tomography: a multicenter management study. *Am J Med*. 2004; 116:291-299.
19. Perrier A, Roy PM, Sanchez O, et al. Multidetector-row computed tomography in suspected pulmonary embolism. *N Engl J Med*. 2005;352:1760-1768.
20. Lucassen W, Geersing GJ, Erkens PM, et al. Clinical decision rules for excluding pulmonary embolism: a meta-analysis. *Ann Intern Med*. 2011;155:448-460.
21. 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;12:915-920.
22. Ceriani E, Combescure C, Le Gal G, et al. Clinical prediction rules for pulmonary embolism: a systematic review and meta-analysis. *J Thromb Haemost.* 2010;8:957-970.
23. Pepe MS, Janes H, Longton G, et al. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol*. 2004;159:882-890.
24. Klok FA, Kruisman E, Spaan J, et al. Comparison of the revised Geneva score with the Wells rule for assessing clinical probability of pulmonary embolism. *J Thromb Haemost*. 2008;6:40-44.
25. Penaloza A, Melot C, Motte S. Comparison of the Wells score with the simplified revised Geneva score for assessing pretest probability of pulmonary embolism. *Thromb Res.* 2011;127:81-84.
26. Di Nisio M, Squizzato A, Rutjes AW, et al. Diagnostic accuracy of D-dimer test for exclusion of venous thromboembolism: a systematic review. *J Thromb Haemost*. 2007;5:296-304.
27. Righini M, Perrier A, De Moerloose P, et al. D-Dimer for venous thromboembolism diagnosis: 20 years later. *J Thromb Haemost*. 2008;6:1059-1071.
28. Roy PM, Colombet I, Durieux P, et al. Systematic review and meta-analysis of strategies for the diagnosis of suspected pulmonary embolism. *BMJ*. 2005;331:259.
29. Penaloza A, Kline J, Verschuren F, et al. European and American suspected and confirmed pulmonary embolism populations: comparison and analysis. *J Thromb Haemost.* 2012;10:375-381.
30. Lapostolle F, Surget V, Borron SW, et al. Severe pulmonary embolism associated with air travel. *N Engl J Med*. 2001;345: 779-783.
31. Runyon MS, Webb WB, Jones AE, et al. 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; 12:587-593.
32. Chagnon I, Bounameaux H, Aujesky D, et al. Comparison of two clinical prediction rules and implicit assessment among patients with suspected pulmonary embolism. *Am J Med*. 2002;113:269-275.

APPENDIX: E1.

Subgroup analysis: 651 patients for whom diagnostic strategies were validated as appropriate in the original study.[10]

Table E1. Comparison of assessment methods: proportion of patient categorized and pulmonary embolism prevalence in each clinical probability level.

CP Assessment (N=651) (%; 95% CI)	Gestalt	Wells Score	RGS
Number of Patients			
Low	303 (47; 43–50)	269 (41; 37–45)	190 (29; 26–33)
Moderate	164 (25; 22–29)	324 (50; 46–54)	383 (59; 55–63)
High	184 (28; 25–32)	58 (9; 7–11)	78 (12; 10–15)
Number of PEs			
Low	22 (7.3; 4.7–10.6)	35 (13.0; 9.4–17.4)	27 (14.2; 9.8–19.7)
Moderate	83 (50.6; 43.0–58.2)	197 (60.8; 55.4–66.0)	182 (47.5; 42.6–52.5)
High	169 (91.8; 87.2–95.2)	42 (72.4; 59.9–82.7)	65 (83.3; 73.8–90.4)

RGS, Revised Geneva score.

Table E2. Agreement on the 3 clinical probability assessment methods in patients with suspected pulmonary embolism, compared 2 by 2.

		RGS			
PE Prevalence %, n	**Probability**	**Low**	**Moderate**	**High**	**Total**
Gestalt assessment*	**Low**	152 (3.9)	142 (9.9)	9 (18.2)	303
	Moderate	24 (41.7)	123 (49.6)	17 (70.6)	164
	High	14 (78.6)	118 (90.7)	52 (98.1)	184
	Total	190	383	78	651

		Gestalt Assessment			
PE Prevalence %, n	**Probability**	**Low**	**Moderate**	**High**	**Total**
Wells†	**Low**	202 (3.0)	47 (27.7)	20 (80.0)	269
	Moderate	93 (17.2)	104 (60.6)	127 (92.9)	324
	High	8 (0)	13 (53.8)	37 (94.6)	58
	Total	303	164	184	651

		Wells			
PE Prevalence %, n	**Probability**	**Low**	**Moderate**	**High**	**Total**
RGS‡	**Low**	130 (3.1)	57 (40.4)	3 (0)	190
	Moderate	132 (18.9)	236 (63.6)	15 (46.7)	383
	High	7 (85.7)	31 (77.4)	40 (87.5)	78
	Total	269	324	58	651

RGS, Revised Geneva score.
*κ Coefficient=0.27 (95% CI 0.22 to 0.32).
†κ Coefficient=0.28 (95% CI 0.23 to 0.33).
‡κ Coefficient=0.35 (95% CI 0.28 to 0.41).

APPENDIX: E2.

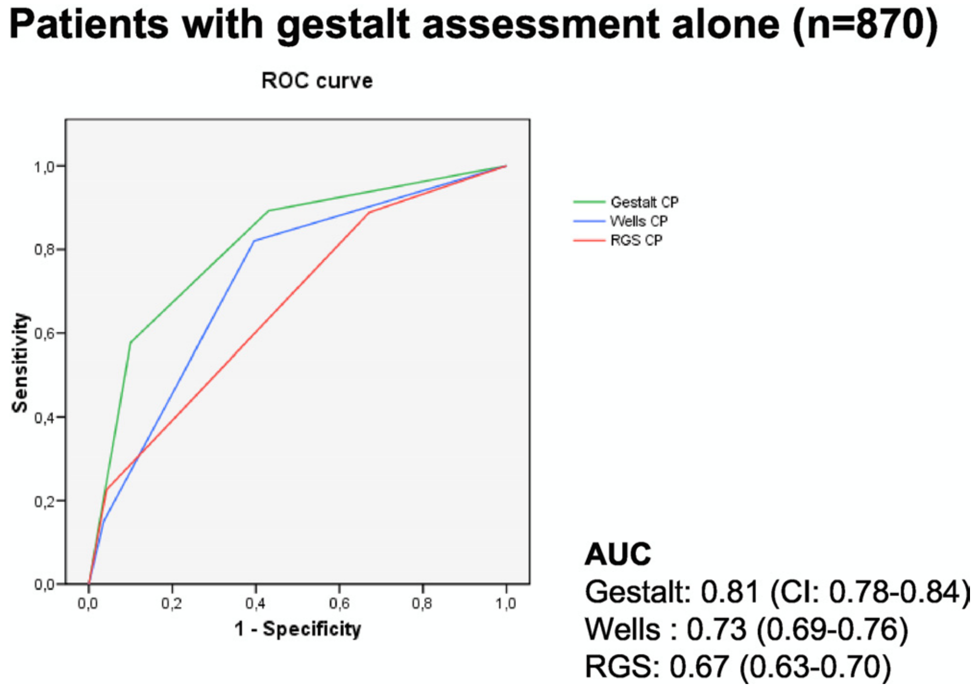

Figure E1. ROC curves: Comparison of the predictive accuracy of the 3 clinical probability assessment methods in patients with suspected pulmonary embolism, using clinical probability groups (low, moderate, and high). Patients with gestalt assessment alone (n=870). *AUC*, Area under the curve; *ROC*, receiver operating characteristics; *CP*, clinical probability; *RGS*, revised Geneva score.

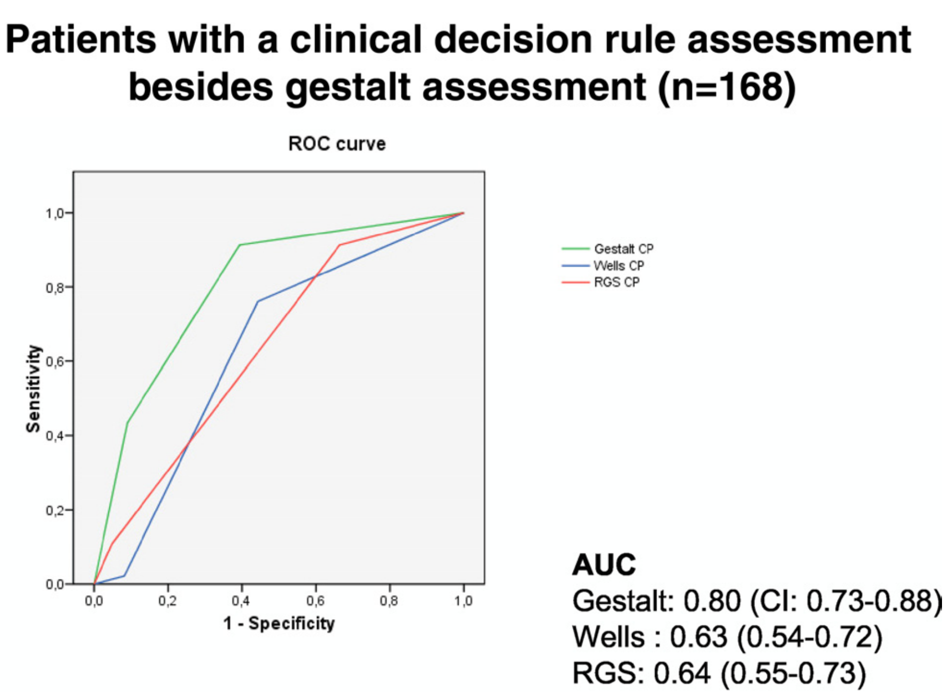

Figure E2. ROC curves: Comparison of the predictive accuracy of the 3 clinical probability assessment methods in patients with suspected pulmonary embolism, using clinical probability groups (low, moderate, and high). Patients with a clinical decision rule assessment besides gestalt assessment (n=168). *AUC*, Area under the curve; *ROC*, receiver operating characteristics; *CP*, clinical probability; *RGS*, revised Geneva score.