Received: 24 January 2024 | Revised: 2 August 2024 | Accepted: 5 August 2024

DOI: 10.1002/hsr2.70003

ORIGINAL RESEARCH

Health Science Reports Open Access WILEY

Comparisons of clinical scoring systems among suspected pulmonary embolism patients presenting to emergency department

Luojia Tang[1] | Yundi Hu[2] | Min Min[1] | Jianyong Gu[1] | Dong Pan[3] | Xiaolei Lin[2] | Chaoyang Tong[1]

[1]Emergency Department of Zhongshan Hospital, Fudan University, Shanghai, China

[2]School of Data Science, Fudan University, Shanghai, China

[3]Department of Information and Intelligence Development, Zhongshan Hospital, Fudan University, Shanghai, China

Correspondence

Xiaolei Lin, School of Data Science, Fudan University, Shanghai, China.
Email: xiaoleilin@fudan.edu.cn

Chaoyang Tong, Emergency Department of Zhongshan Hospital, Fudan University, Shanghai, China.
Email: tong.chaoyang@zs-hospital.sh.cn

Funding information

National Key Research and Development Program of China, Grant/Award Number: 2021YFC2501800; Science and Technology of Shanghai Committee, Grant/Award Number: 21MC1930400

Abstract

Introduction: Pulmonary embolism (PE) is among the most severe cardiovascular disorders worldwide. Timely and appropriate diagnosis of PE remains an important step in reducing PE related mortality and morbidity.

Methods: In this retrospective single-center cohort study, we comprehensively compared the screening performances of several clinical scoring systems (Wells score [WS], Revised Geneva score [RGS], WS + D-Dimer [D-D], RGS + D-D, WS + PE rule-out criteria [PERC] and RGS + PERC) among PE suspected patients. Failure rates across different PE severity grades as well as overall sensitivity/specificity were considered in evaluating each screening strategy.

Results: A total of 3437 patients were included in this study and 698 of them were diagnosed with PE. Patients with and without PE were similar in demographics, while significantly different in respiration-related characteristics. Compared with WS or RGS alone, Integrating PERC or D-D with WS or RGS significantly decreased the failure rates across all PE severity grades, and increased the overall sensitivity from 88.5% and 87.2% to 96.3% and 94.8% (D-D) to 99.4% and 99.6% (PERC), respectively. However, compared with other four scoring approaches, using WS or RGS alone increased the specificity from 8.3% and 7.2%, 38.3% and 21.3%, to 63.5% and 34.8%, respectively, and increased the AUC from 0.54 to 0.54, 0.70 and 0.69, to 0.8 and 0.76, respectively. In general, all screening approaches achieved better performances among PE patients with respiratory distress compared to those without respiratory distress.

Conclusion: Combining PERC or D-D with WS or RGS, and the presence of respiratory distress provide significantly better PE rule-out performances.

KEYWORDS

D-Dimer, pulmonary embolism, pulmonary embolism rule out criteria, respiratory distress, revised Geneva score, Wells score

Luojia Tang and Yundi Hu contributed equally.

1 | INTRODUCTION

Pulmonary embolism (PE) is among the most severe cardiovascular disorders and is responsible for over 50,000 annual deaths in the United States.[1] Most PE is caused by emboli that originated from venous thrombi and traveled to the arteries of lung, with various clinical presentations and symptoms, including arrhythmia, hemodynamic collapse and shock.[2] Undiagnosed PE can be life-threatening, and patients surviving delayed PE diagnosis are at increased risk for chronic right ventricular dysfunction and decreased quality of life.[3] Therefore, timely diagnosis and appropriate therapeutic management of PE remains essential for reducing the mortality and morbidity associated with PE.

There have been several important improvements regarding the clinical diagnosis of PE. Among them, computed tomographic pulmonary angiography (CTPA), which generates three-dimensional images demonstrating the presence and location of pulmonary emboli within arteries in the lung, remains the gold standard.[4] However, CTPA is costly, time-consuming and often exposes patients to extra radiation. These limitations largely restrict their applications in emergency medical settings.[5] In the past 20 years, several clinical scoring systems, such as Well's score (WS), revised Geneva score (RGS), and pulmonary embolism rule-out criteria (PERC) have proven to be an alternative screening tool for aiding the timely diagnosis of PE.[6–9] These scoring systems evaluate the likelihood of PE via a combination of diagnostic modalities, including clinical signs, symptoms, and medical history, and have been validated by clinical trials and prospective cohort studies.[10–12] Unlike WS and RGS which both intend to stratify patients with suspected PE into different risk categories, PERC is designed to rule out PE in low-risk patients. Therefore, PERC is rarely used alone in predicting PE pretest probability, but rather integrated with existing pretest scores to improve their clinical performance. In addition, the plasma level of D-Dimer (D-D), which is often elevated in the presence of acute thrombosis due to simultaneous activation of coagulation and fibrinolysis, is also used to rule out PE.[13,14]

Existing research often utilizes these screening scores in multiple stages rather than integrate them into a single screening approach, which had limited their clinical applicability. In addition, it is well established that pulmonary artery embolism and multiple pulmonary arteriole embolism require a much more stringent diagnosis and treatment window than single pulmonary arteriole embolism. However, existing research rarely investigated the performances of these clinical scoring systems across different PE severity grades. Therefore, in this study, we aim to compare the screening performances of several well-known PE clinical prediction rules: WS alone, RGS alone, WS integrated with D-Dimer (WS + D-D), RGS integrated with D-Dimer (RGS + D-D), WS integrated with PE rule-out criteria (WS + PERC), and RGS integrated with PERC (RGS + PERC) among patients suspected of PE from our single medical center. Screening performances are evaluated with respect to different PE severity grades to detect the consistency of their clinical performance. Further, since respiratory distress is a typical symptom of PE and to investigate its clinical significance in constructing scoring system, we evaluated the performances of multiple screening approaches among patient subgroups with and without respiratory distress.

2 | MATERIALS AND METHODS

2.1 | Data collection

In this retrospective single-center cohort study, we collected data from patients presenting to the emergency department of Zhongshan Hospital (tertiary hospital affiliated to Fudan University) from July 2018 to October 2022. During this period, an average of 295,608 patients attended the emergency department of Zhongshan Hospital annually. Patients were included if they were suspected of PE or could not be ruled out of PE by the attending physicians. Since the attending physicians were well-educated and the clinical diagnosis of PE followed standard protocol, we define patients suspected of PE or could not be ruled out of PE as those who performed CTPA according to the medical order of the attending physicians. Patients were excluded if they were younger than 18 years old, or could not be confirmed of PE status by CTPA. Electronic medical records, CTPA, laboratory test results and vital signs were collected and reviewed.

2.2 | PE status and severity grades

The status of PE was determined based on CTPA. PE was divided into five severity grades: (1) pulmonary artery embolism; (2) multiple pulmonary arteriole embolism; (3) pulmonary artery branch embolism; (4) single pulmonary arteriole embolism; and (5) no PE. Supporting Information: Text S2 described how these severity grades were assessed in detail.

2.3 | Study design

The performances of four PE screening approaches were compared among patients included in the analysis, including WS, RGS, WS + D-D, RGS + D-D, WS + PERC, and RGS + PERC. Table 1 shows a detailed description of the six screening approaches. Detailed indicators of WS, RGS, and PERC can be found in Supporting Information: Tables S1–S3. Comparisons were conducted comprehensively in terms of screening accuracy, efficiency and the concordance of clinical probabilities. In addition, since respiratory distress was a typical symptom of PE, subgroup analyses were conducted among patients with respiratory distress. We defined patients who have shortness of breath, breathing difficulty, or oxyhemoglobin saturation less than 95% as respiratory distress.

2.4 | Statistical analysis

Statistical analysis was performed using R statistical software (version 4.0.3). Continuous characteristics were described using mean ±

TABLE 1 Detailed description of four screening approaches.

Screening approach	Content
WS	If Wells score ≤4, patients were diagnosed as no PE, and other patients needed to undergo CTPA.
RGS	If Geneva score ≤3, patients were diagnosed as no PE, and other patients needed to undergo CTPA.
WS + D-D	If Wells score ≤4 and D-D ≤ 2.5 mg/L, patients were diagnosed as no PE, and other patients needed to undergo CTPA.
RGS + D-D	If Geneva score ≤3 and D-D ≤ 2.5 mg/L, patients were diagnosed as no PE, and other patients needed to undergo CTPA.
WS + PERC	If Wells score ≤4 and PERC = 0, patients were diagnosed as no PE, and other patients needed to undergo CTPA.
RGS + PERC	If Geneva score ≤3 and PERC = 0, patients were diagnosed as no PE, and other patients needed to undergo CTPA.

Abbreviations: CTPA, computed tomography pulmonary angiography; D-D, D-Dimer; PERC, Pulmonary embolism rule-out criteria; RGS, revised Geneva score; WS, Wells score.

standard deviation if normally distributed, or median (interquartile range) if normality was violated. Categorical characteristics were presented using frequency (percentage) per category. Statistical tests were performed using Student's *t* test or analysis of variance for continuous variables (Wilcoxon rank sum test for skewed distribution), and Chi-square or Fisher's exact test for categorical variables. Failure rate, which was defined as the percentage of PE cases failed to be ruled out in each PE severity grade, of each screening approach was calculated and compared. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) were also calculated and compared among different screening approaches. Predictive accuracy of each screening approach was calculated as the area under the receiver operating characteristics (ROC) area under the curve (AUC), and significance of difference was tested by DeLong test.[15] Concordance between the clinical probability of PE was described by Kappa coefficient.[16] 1 − Cohen's W effect sizes was calculated to further measure the difference among the scoring approaches. To detect the failure rate differences between different scoring strategies, following formulation was applied to calculate *Z* statistics. $|Z| > 1.96$ means significant difference at 95% confidence.

$$Z = \frac{FR_1 - FR_2}{\sqrt{\frac{FR_1(1-FR_1)+FR_2(1-FR_2)}{n}}}.$$

3 | RESULTS

3.1 | Patient characteristics

A total of 3973 patients presenting to the emergency department of Zhongshan Hospital from July 2018 to October 2022 met the inclusion criteria and were included in the study. Among them, 17 were younger than 18 years old, 519 could not be confirmed of PE status by CTPA and were thus excluded from the following analyses. Supporting Information: Figure S1 provided a flow diagram for detailed patient inclusion process. Baseline characteristics for the remaining 3437 patients included in the final analysis were described in Table 2. Overall, patients with and without PE were significantly different in terms of most characteristics, especially those related to respiration, such as respiratory distress, heart rate, and oxygen saturation. Age, sex, syncope, palpitation, and temperature were found to be distributed similarly between PE and non-PE patients.

TABLE 2 Baseline characteristics among patients with and without pulmonary embolism (PE).

	With PE (*n* = 698)	Without PE (*n* = 2739)	*p*-Value
Age (IQR), in years	68 (60–76)	68 (59–77)	0.8169
Male	331 (47.4%)	1226 (44.8%)	0.2233
Symptoms			
Respiratory distress	159 (22.8%)	398 (14.5%)	<0.001
Unilateral edema	118 (16.9%)	252 (9.2%)	<0.001
Chest distress	197 (28.2%)	645 (23.5%)	0.0119
Chest pain	107 (15.3%)	772 (28.2%)	<0.001
Hemoptysis	17 (2.4%)	35 (1.3%)	0.0391
Syncope	19 (2.7%)	98 (3.6%)	0.3191
Cough	28 (4.0%)	67 (2.4%)	0.0338
Palpitation	58 (8.3%)	240 (8.8%)	0.7610
Signs			
Temperature, °C	36.5 (36.2–36.8)	36.5 (36.2–36.8)	0.5573
Heart rate, beats/min	93 (80–109)	86 (76–101)	<0.001
Systolic blood pressure, mm Hg	136 (119–152)	140 (122–158)	<0.001
Diastolic blood pressure, mm Hg	77 (69–86)	76 (66–86)	0.0094
Oxygen saturation, %	95 (92–97)	97 (95–98)	<0.001
Medical history			
Cancer	102 (14.6%)	158 (5.8%)	<0.001
Pulmonary embolism	47 (6.7%)	64 (2.3%)	<0.001

Note: Data are presented as *n* (%) for categorical characteristics, and median (IQR) or mean (standard deviation) for continuous characteristics. *p*-Values were obtained using Student's *t* test for normally distributed variables, Wilcoxon rank sum test for skewed distributed variables and Chi-square test for categorical variables.

3.2 | Comparison of screening performances among four scoring approaches

Failure rate decreased from 9.5% and 11.9% in WS and RGS, to 4.0% and 4.0% for WS + D-D and RGS + D-D, and to 1.6% and 0.8% for WS + PERC and RGS + PERC among patients with main pulmonary artery PE. Similar trends were observed among patients with multiple, branch, arteriole, and any PE (Table 3). WS + PERC and RGS + PERC had the highest sensitivity and NPV, while WS had the highest specificity and PPV. Transitioning from the WS to the WS + D-D, and subsequently to the WS + PERC, the scoring approaches become increasingly conservative, with a corresponding decrease in the failure rate. Figure 1 showed the ROC curve and AUC for each of the screening approach. WS had the highest AUC (0.7999, 95% confidence interval [0.7831–0.8169]), followed by RGS (0.7555, [0.7351–0.7760]), while there was no statistically significant difference regarding the AUCs in WS + D-D (0.6974, [0.6819–0.7127]) and RGS + D-D (0.6873, [0.6971–0.7036]), as well as AUCs in WS + PERC (0.5391, [0.5333–0.5451]) and RGS + PERC (0.5392, [0.5334–0.5451]).

The 3437 patients were divided into low and high clinical probability categories according to the corresponding risk scores in each screening approach. Concordance of the clinical probability categories between WS and RGS, WS + D-D and RGS + D-D, as well as WS + PERC and RGS + PERC approaches were provided in Table 4. Overall, there were 2484 patients with concordant clinical probability estimated using WS and RGS, including 955 patients with low clinical probabilities and 1529 patients with high clinical probabilities. The incorporation of D-D has diminished the disparity in the screening performance between the WS and the RGS, number of patients with concordant clinical probability increased to 2910, including 583 patients with low clinical probabilities and 2327 with high clinical probabilities. However, PERC further equalized this disparity that number of patients with concordant clinical probability further increased to 3713, including 199 patients with low clinical probabilities and 3204 with high clinical probabilities. Kappa coefficients between WS and RGS, WS + D-D, and RGS + D-D, as well as WS + PERC and RGS + PERC were 0.4580, 0.5964, and 0.9901, respectively, indicating moderate to almost perfect concordance. Negligible difference existed between WS + PERC and RGS + PERC (1 − Cohen's $W = 0.084$), and medium difference existed between WS + D-D and RGS + D-D (0.364), as well as between WS and RGS (0.490) (Supporting Information: Table S4). Supporting Information: Table S5 showed the Z statistics of failure rate in each PE severity grade.

3.3 | Comparison of screening performances between patients with and without respiratory distress

According to Table 2, patients with PE (22.8%) had significantly higher probability of respiratory distress symptoms than patients without PE (14.5%). Therefore, we conducted subgroup analysis regarding the performances of screening approaches in this subpopulation. 557 out of the 3437 patients were found to be presented with respiratory distress after reviewing their electronic medical records. For WS and RGS, failure rates increased from 7.0% and 8.5% for patients with respiratory distress, to 12.7% and 16.4% for patients without respiratory distress among patients with main

TABLE 3 Comparison of screening performances in four scoring strategies.

	WS	RGS	WS + D-D	RGS + D-D	WS + PERC	RGS + PERC
Failure rate						
Failure rate in main pulmonary artery PE	9.5%	11.9%	4.0%	4.0%	1.6%	0.8%
Failure rate in multiple PE	8.0%	7.3%	3.0%	2.3%	0.0%	0.0%
Failure rate in branch PE	15.6%	17.4%	3.6%	6.0%	0.6%	0.6%
Failure rate in arteriole PE	15.7%	21.6%	4.9%	12.7%	1.0%	1.0%
Failure rate in any PE	11.5%	12.8%	3.6%	5.0%	0.6%	0.4%
Other metrics						
Sensitivity	88.5%	87.2%	96.3%	94.8%	99.4%	99.6%
Specificity	63.5%	34.8%	38.3%	21.3%	8.3%	7.2%
NPV	95.6%	91.5%	97.6%	94.2%	98.3%	98.5%
PPV	38.2%	25.4%	28.4%	23.5%	21.7%	21.5%
Negative LR	0.1811	0.3678	0.0974	0.2423	0.0698	0.0556
Positive LR	2.4247	1.3374	1.5594	1.2049	1.0840	1.0732

Abbreviations: D-D, D-Dimer; LR, likelihood ratio; NPV, negative predictive value; PE, pulmonary embolism; PERC, pulmonary embolism rule-out criteria; PPV, positive predictive value; RGS, Revised Geneva score; WS, Wells score.

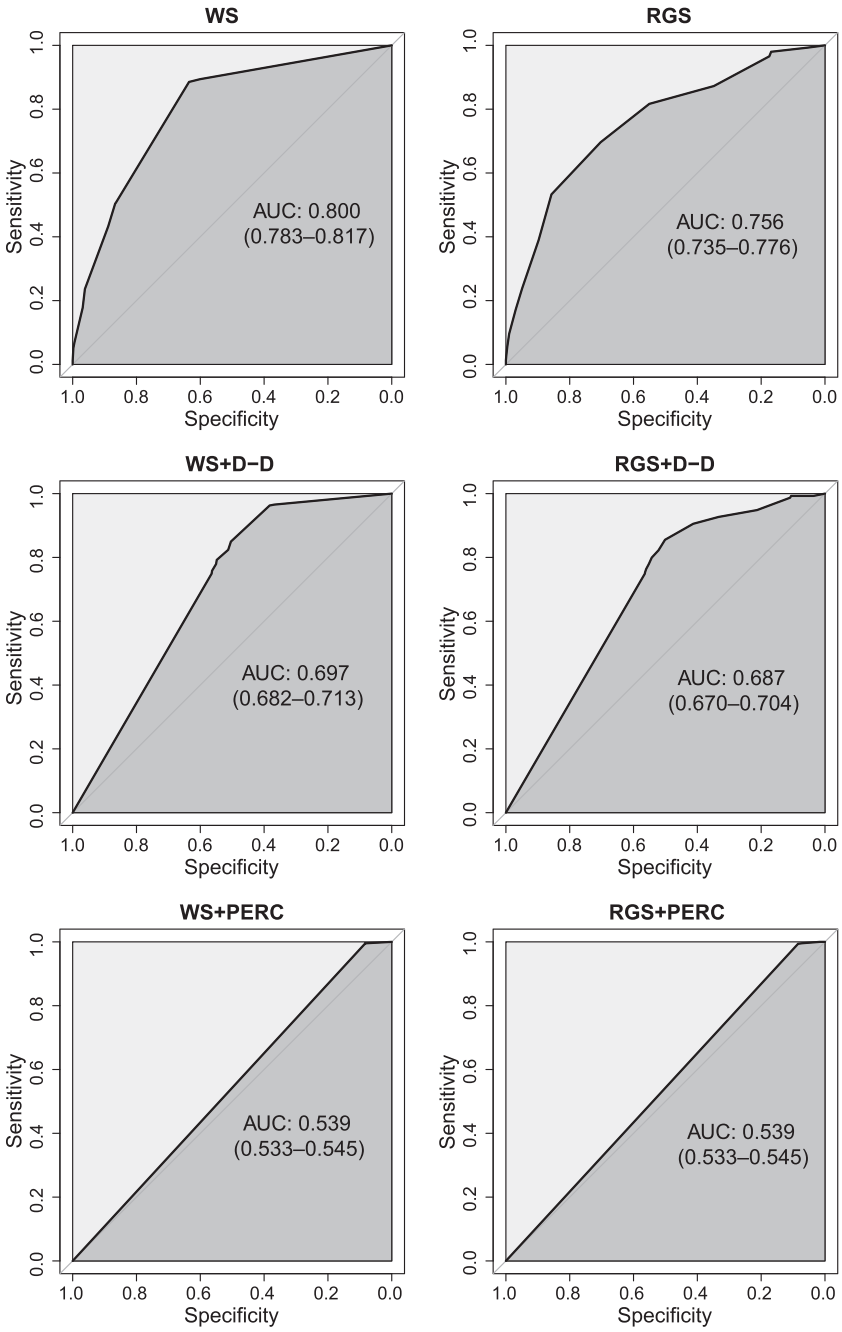

FIGURE 1 ROC curve and AUC (95% confidence interval) in each of the six screening approaches. AUC, area under the curve; D-D, D-Dimer; PERC, pulmonary embolism rule-out criteria; RGS, revised Geneva score; ROC, receiver operating characteristics; WS, Wells score.

pulmonary artery PE. Similar trends were observed among those with other PE severity grades (Table 5). In other words, screening performances of WS and RGS increased among patients with respiratory distress, relative to those without. We also observed the similar results in RGS + D-D. For WS + D-D, we observed inverse results in patients with main pulmonary artery PE and multiple PE. However, the benefit among patients with respiratory distress was not significant for WS + PERC or RGS + PERC due to low failure rates. In addition, sensitivity and PPV decreased, while specificity and NPV increased among patients with respiratory distress. Overall, we found that all six screening approaches achieved higher accuracy but lower efficiency among patients with respiratory distress.

4 | DISCUSSION

In this article, we comprehensively compared the screening performances of six scoring systems commonly used to aid the clinical diagnosis of PE. Failure rates across all PE severity grades decreased significantly from WS alone and RGS alone to WS + D-D and RGS + D-D, to WS + PERC and RGS + PERC, respectively. The percentages of PE cases failed to be diagnosed by the corresponding screening strategies were minimized to less than 1% in WS + PERC and RGS + PERC. These two integrative screening strategies (WS + PERC and RGS + PERC) achieved near perfect performances in terms of sensitivity, however, their specificity remained much lower than the simple screening approaches (WS alone and RGS alone).

WS + D-D and RGS + D-D were in balanced compared with other four scoring approaches. As the two most commonly used screening rules of PE, Wells score and RGS only achieved moderate screening concordance in dividing suspected PE patients into low and high-risk categories, while integrating D-D or PERC with WS or RGS could significantly increase the screening concordance between them. In addition, all six screening strategies achieved lower failure rate among patients with respiratory distress compared to those without across all PE severity grades, except for the most severe case (main pulmonary artery PE), where none of the main pulmonary artery PE cases without respiratory distress were misdiagnosed using RGS + PERC and 1.4% main pulmonary artery PE with respiratory distress failed to be diagnosed. This suggests that RGS + PERC could serve as a potentially better diagnostic tool for suspected PE patients without respiratory distress.

TABLE 4 Concordance of clinical probability categories between WS and RGS, between WS + PERC and RGS + PERC, and between WS + D-D and RGS + D-D.

		RGS			
		Low	High	Total	Kappa coefficient
WS	Low	955	865	1820	0.4580
	High	88	1529	1617	
	Total	1043	2394	3437	
		RGS + D-D			
		Low	High	Total	Kappa coefficient
WS + D-D	Low	583	491	1074	0.5965
	High	36	2327	2363	
	Total	619	2818	3437	
		RGS + PERC			
		Low	High	Total	Kappa coefficient
WS + PERC	Low	199	33	232	0.9901
	High	1	3204	3205	
	Total	200	3237	3437	

Abbreviations: D-D, D-Dimer; PERC, pulmonary embolism rule-out criteria; RGS, Revised Geneva score; WS, Wells score.

Failure rate, which is defined as the number of PE cases out of the total number of suspected PE cases in the corresponding severity grade, is of particular clinical importance for the early and correct diagnosis of PE.[17] Since PERC is designed to safely rule out PE among suspected PE cases with lower clinical risk, it is expected to decrease the severity specific and overall failure rates when integrated with WS or RGS, which are designed to stratify patients with suspected PE into different risk categories.[18] Therefore, we observed significantly better screening performances in terms of sensitivity and failure rate using WS + PERC and RGS + PERC compared with WS and RGS alone. In other words, when combined with PERC, WS and RGS would be much more sensitive at diagnosing more suspected PE cases, which is desired for severe disorders such as PE with high mortality and morbidity.

Although integrating PERC would increase the sensitivity for diagnosing PE cases using WS or RGS, it comes with a compromise of decreasing the overall specificity. This would be expected since for any screening strategy, an increase of sensitivity would be accompanied with a decrease of specificity, that is, any screening tool would

TABLE 5 Comparison of screening performances among patients with and without respiratory distress.

	WS		RGS		WS + D-D		RGS + D-D		WS + PERC		RGS + PERC	
	RD	No RD	RD	No RD	RD	No RD	RD	No RD	RD	No RD	RD	No RD
Failure rate												
Failure rate in main pulmonary artery PE	7.0%	12.7%	8.5%	16.4%	2.8%	1.4%	5.5%	7.3%	1.4%	1.8%	1.4%	0.0%
Failure rate in multiple PE	7.0%	10.3%	5.6%	11.5%	1.4%	0.5%	6.9%	6.9%	0.0%	0.0%	0.0%	0.0%
Failure rate in branch PE	11.1%	20.8%	14.4%	20.8%	3.3%	5.5%	3.9%	6.5%	0.0%	1.3%	0.0%	1.3%
Failure rate in arteriole PE	5.7%	26.5%	15.1%	28.6%	0.0%	9.4%	10.2%	16.3%	0.0%	2.0%	0.0%	2.0%
Failure rate in any PE	8.1%	16.8%	9.3%	18.3%	1.9%	2.8%	6.3%	8.6%	0.2%	1.1%	0.2%	0.7%
Other metrics												
Sensitivity	91.9%	83.2%	90.7%	81.7%	97.9%	97.0%	93.7%	91.4%	99.8%	98.9%	99.8%	99.3%
Specificity	53.0%	69.3%	27.1%	39.1%	29.2%	15.0%	43.3%	24.8%	5.4%	9.9%	4.8%	8.5%
NPV	93.6%	96.5%	86.8%	93.4%	96.9%	91.8%	97.8%	95.0%	98.1%	98.3%	97.9%	98.7%
PPV	46.3%	29.2%	35.5%	16.9%	37.9%	33.5%	20.0%	15.6%	31.8%	14.3%	31.6%	14.1%
Negative LR	0.1528	0.2424	0.3432	0.4680	0.0718	0.2017	0.1465	0.3466	0.0370	0.1111	0.0417	0.0824
Positive LR	1.9553	2.7101	1.2442	1.3415	1.3820	1.1408	1.6514	1.2150	1.0550	1.0977	1.0483	1.0852

Abbreviations: D-D, D-Dimer; LR, likelihood ratio; NPV, negative predictive value; PE, pulmonary embolism; PERC, pulmonary embolism rule-out criteria; PPV, positive predictive value; RD, respiratory distress; RGS, Revised Geneva score; WS, Wells score.

be increasing the sensitivity by diagnosing more suspected cases at the cost of false positives.[19] Therefore, we observed much lower specificity when WS or RGS is combined with PERC. This could also explain the relatively lower AUCs in these two integrative screening strategies. Since AUC, which is calculated as the area under the ROC curve, is indicative of the discriminative power of the corresponding screening approach with sensitivity plotted against specificity when probability is thresholded sequentially between 0 and 1.[20] The poor discriminative performances of the integrative approaches are likely to be due to their relatively lower specificity. However, under disease screening scenarios in community setting or disease diagnosis scenarios in emergency medical setting, true disease status remains unknown and medical professionals would largely rely on the positive/negative predictive values, which predict the probability that a suspected PE patient would be truly PE given positive screening result (PPV) or that a suspected PE patient could be ruled out of PE given negative screening result (NPV).[21] The high NPVs and low PPVs of the integrative approaches suggest that, it is relatively safe to rule out of PE patients given the negative screening result, however, inadequate to confirm the PE diagnosis given the positive screening result.

Further, patients diagnosed with PE had significantly higher prevalence of respiratory distress compared to those without among the samples included in our study, which confirms that respiratory distress is among one of the most common symptoms occurring to PE. This can also be reflected in our subgroup analyses where almost all screening strategies achieved better performances among suspected PE patients with respiratory distress compared with those without. However, one should note that among patients with main pulmonary artery PE and without respiratory distress, the RGS + PERC approach achieved zero failure rate, a big improve than among those with main pulmonary artery PE but without respiratory distress. This suggests that RGS + PERC could potentially benefit the clinical diagnosis of this most difficult subgroup, which represent PE patients with untypical clinical symptoms and severe PE consequences.

There are several strengths associated with our study. First, we are the first to stratify PE patients into different PE severity grades using the gold standard diagnostic CTPA and evaluated the screening performances for each PE severity grade. The grade specific failure rates help to differentiate the clinical significance of each screening tool from the most severe PE cases to the mildest cases. Second, the screening performances of all four strategies are compared comprehensively using rule out criteria (such as failure rate), sensitivity/specificity, as well as a compromise between them (AUC), which provides diverse angles toward different clinical usages. For example, medical professionals focusing on the clinical diagnosis might emphasize the performance on failure rate, while those working in the community screening of associated risk factors might consider specificity as more important. Thirdly, the large sample size and relatively high prevalence of PE cases among the suspected PE population make our results reliable and generally applicable to patients with moderate to high PE risks of east Asian ancestry.

Nevertheless, there are several limitations related to our study. First, our study is retrospective in nature and could not avoid bias due to potentially unmeasured confounding. Second, all patients analyzed in our study are of East Asian ancestry, and thus the results and conclusions might not be generalizable to other population. Finally, our study is a single center investigation and results need to be validated using multicenter prospective cohorts.

5 | CONCLUSION

In this study, we comprehensively compared the screening performances of WS, RGS, WS + D-D, RGS + D-D, WS + PERC, and RGS + PERC among suspected PE patients across various PE severity grades. Results suggest WS + PERC and RGS + PERC as outstanding PE rule out method compared to other scoring approaches, with near perfect sensitivity and negative predictive values. And WS + D-D and RGS + D-D balanced the sensitivity and specificity well. Finally, the performances of all approaches were improved among suspected PE patients with respiratory distress compared to those without respiratory distress, and RGS + PERC could potentially serve as a better screening tool for main pulmonary artery PE without respiratory distress.

AUTHOR CONTRIBUTIONS

Luojia Tang: Conceptualization; data curation; funding acquisition; investigation; methodology; validation; writing—original draft. **Yundi Hu**: Data curation; formal analysis; methodology; software; validation; visualization; writing—original draft. **Min Min**: Data curation; formal analysis; investigation. **Jianyong Gu**: Data curation; formal analysis; investigation. **Dong Pan**: Data curation. **Xiaolei Lin**: Formal analysis; methodology; methodology; supervision; supervision; validation; writing—review and editing; writing—review and editing. **Chaoyang Tong**: Conceptualization; data curation; investigation; supervision; writing—review and editing.

ACKNOWLEDGMENTS

The authors appreciate help from colleagues from Fudan University and Zhongshan Hospital with this project. This work was supported by National Key Research and Development Program of China (2021YFC2501800) and Science and Technology of Shanghai Committee (Grant No. 21MC1930400). The funding agency had no role in the study design, data collection, data analysis, result interpretation, writing and decision to submit the report for publication of the manuscript associated with this study.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ETHICS STATEMENT

The study was registered (No. CJ0647) and approved by Human Genetic Resources of China in April 2022. Ethical approval was received from the Medical Ethics Committee of Zhongshan Hospital (NO. B2021-839R).

TRANSPARENCY STATEMENT

The lead author Xiaolei Lin, Chaoyang Tong affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

ORCID

Yundi Hu http://orcid.org/0000-0003-1166-2803
Chaoyang Tong http://orcid.org/0000-0002-2727-2853

REFERENCES

1. Farmakis IT, Valerio L, Bikdeli B, et al. Annual mortality related to pulmonary embolism in the US before and during the COVID-19 pandemic. *J Am Coll Cardiol*. 2022;80(16):1579-1581.
2. Huisman MV, Barco S, Cannegieter SC, et al. Pulmonary embolism. *Nat Rev Dis Primers*. 2018;4:18028.
3. Keller K, Tesche C, Gerhold-Ay A, et al. Quality of life and functional limitations after pulmonary embolism and its prognostic relevance. *J Thromb Haemostasis*. 2019;17(11):1923-1934.
4. Belzile D, Jacquet S, Bertoletti L, et al. Outcomes following a negative computed tomography pulmonary angiography according to pulmonary embolism prevalence: a meta-analysis of the management outcome studies. *J Thromb Haemostasis*. 2018;16(6):1107-1120.
5. Hunsaker AR, Lu MT, Goldhaber SZ, Rybicki FJ. Imaging in acute pulmonary embolism with special clinical scenarios. *Circ Cardiovasc Imaging*. 2010;3(4):491-500.
6. Wells P, Hirsh J, Anderson D, et al. Accuracy of clinical assessment of deep-vein thrombosis. *The Lancet*. 1995;345(8961):1326-1330.
7. Kline JA, Mitchell AM, Kabrhel C, Richman PB, Courtney DM. Clinical criteria to prevent unnecessary diagnostic testing in emergency department patients with suspected pulmonary embolism. *J Thromb Haemostasis*. 2004;2(8):1247-1255.
8. Le Gal G, Righini M, Roy P-M, et al. Prediction of pulmonary embolism in the emergency department: the revised Geneva score. *Ann Intern Med*. 2006;144(3):165-171.
9. Wells P, Anderson D, 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(03):416-420.
10. Wolf SJ, McCubbin TR, Nordenholz KE, Naviaux NW, Haukoos JS. Assessment of the pulmonary embolism rule-out criteria rule for evaluation of suspected pulmonary embolism in the emergency department. *Am J Emerg Med*. 2008;26(2):181-185.
11. Hendriksen JM, Geersing G-J, Lucassen WA, et al. Diagnostic prediction models for suspected pulmonary embolism: systematic review and independent external validation in primary care. *BMJ*. 2015;8(351):h4438.
12. 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(5):503-510.
13. Kohn MA, Klok FA, van Es N. D-dimer interval likelihood ratios for pulmonary embolism. *Acad Emerg Med*. 2017;24(7):832-837.
14. Linkins LA, Bates SM, Ginsberg JS, Kearon C. Use of different D-dimer levels to exclude venous thromboembolism depending on clinical pretest probability. *J Thromb Haemostasis*. 2004;2(8):1256-1260.
15. 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.
16. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. 2012;22(3):276-282.
17. Stals MAM, Takada T, Kraaijpoel N, et al. Safety and efficiency of diagnostic strategies for ruling out pulmonary embolism in clinically relevant patient subgroups: a systematic review and individual-patient data meta-analysis. *Ann Intern Med*. 2022;175(2):244-255.
18. Florkowski CM. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin Biochem Rev*. 2008;29(suppl 1):S83.
19. Naeger DM, Kohi MP, Webb EM, Phelps A, Ordovas KG, Newman TB. Correctly using sensitivity, specificity, and predictive values in clinical practice: how to avoid three common pitfalls. *Am J Roentgenol*. 2013;200(6):W566-W570.
20. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med*. 2013;4(2):627-635.
21. Sundbøll J, Adelborg K, Munch T, et al. Positive predictive value of cardiovascular diagnoses in the Danish National Patient Registry: a validation study. *BMJ Open*. 2016;6(11):e012832.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Tang L, Hu Y, Min M, et al. Comparisons of clinical scoring systems among suspected pulmonary embolism patients presenting to emergency department. *Health Sci Rep*. 2024;7:e70003. doi:10.1002/hsr2.70003