


RESEARCH ARTICLE



Updating cut-off values of severity scoring systems for community-acquired pneumonia to orchestrate more predictive accuracy

Qi Guo^{a,b} , Hai-yan Li^c, Wei-dong Song^a, Ming Li^b, Xiao-ke Chen^b, Hui Liu^b, Hong-lin Peng^b, Hai-qiong Yu^b, Nian Liu^b, Yan-hong Li^a, Zhong-dong Lü^a, Li-hua Liang^d, Qing-zhou Zhao^d and Mei Jiang^e

^aDepartment of Pulmonary and Critical Care Medicine, Shenzhen Hospital, Peking University, Shenzhen, Guangdong, P.R. China; ^bDepartment of Pulmonary and Critical Care Medicine, The Eighth Affiliated Hospital, Sun Yat-sen University, Shenzhen, Guangdong, P.R. China; ^cDepartment of General Medicine, The Eighth Affiliated Hospital, Sun Yat-sen University, Shenzhen, Guangdong, P.R. China; ^dDepartment of Radiology, The Eighth Affiliated Hospital, Sun Yat-sen University, Shenzhen, Guangdong, P.R. China; ^eState Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital, Guangzhou Medical University, Guangzhou, Guangdong, P.R. China

ABSTRACT

Background: The assessment of severity is crucial in the management of community-acquired pneumonia (CAP). It remains unknown whether updating cut-off values of severity scoring systems orchestrate improvement in predictive accuracy.

Methods: 3,212 patients with CAP were recruited to two observational prospective cohort studies. Three bettered scoring systems were derived from the corresponding well-established and extensively used pneumonia-specific severity scoring systems, i.e. pneumonia severity index, minor criteria and CURB-65 (confusion, urea >7 mmol/L, respiratory rate \geq 30/min, low blood pressure, and age \geq 65 years) score, with the updating cut-off values for tachypnea and low blood pressure. Cronbach α was employed to determine construct validity. Discrimination was valued by calculating the area under the receiver operating characteristic curve (AUROC) and net reclassification improvement (NRI).

Results: Respiratory rate \geq 22/min and systolic blood pressure \leq 100 mm Hg were performed better than respiratory rate \geq 30/min and hypotension for predicting mortality in the derivation cohort, respectively (AUROC, 0.823 vs 0.519, 0.688 vs 0.622; NRI, 0.61, 0.13). Bettered scoring systems orchestrated higher convergences, indicated by greater Cronbach α and more decrease in Cronbach α if the updating cut-off values were deleted. The six scoring systems agreed well with one another. Bettered- pneumonia severity index, minor criteria and CURB-65 score showed higher associations with severity and mortality rates and demonstrated greater predictive accuracies for mortality compared with the corresponding original systems (AUROC, 0.939 vs 0.883, 0.909 vs 0.871, 0.913 vs 0.859; NRI, 0.113, 0.076, 0.108; respectively). The validation cohort confirmed a similar pattern.

Conclusions: Updating cut-off values of severity scoring systems for CAP orchestrate improvement in predictive accuracy, suggesting that it may facilitate the rationalization of clinical triage decision-making and further reduce mortality. The current studies provide the first known prospective evidence of potential benefit of the updating cut-off values of severity scoring systems for CAP in predictive accuracy.

KEY MESSAGES





- Updating cut-off values were performed better for predicting mortality.
- Bettered scoring systems orchestrated higher convergences.
- Bettered scoring systems demonstrated greater predictive accuracies for mortality.


ARTICLE HISTORY

Received 13 January 2023
Revised 17 March 2023
Accepted 9 April 2023

KEYWORDS

Community-acquired pneumonia;
Severity scoring system;
Cut-off value updated;
Mortality

CONTACT Qi Guo  qiguoo07@sina.com  Department of Pulmonary and Critical Care Medicine, Shenzhen Hospital, Peking University, Lianhua road No. 1120, Shenzhen, Guangdong, 518036 P.R. China; Hai-yan Li  haiyanli74@sina.com  Department of General Medicine, The Eighth Affiliated Hospital, Sun Yat-sen University, Shennan middle road No. 3025, Shenzhen, Guangdong 518033, P.R. China

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/07853890.2023.2202414>.

The abstract (NO. 5) has been accepted for Oral Presentation and awarded first prize of excellent papers at Respiratory Branch of Guangdong Medical Association Guangzhou 2022 Annual Congress (21-23 October).

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Introduction

Community-acquired pneumonia (CAP) causes great mortality and morbidity and high costs worldwide [1]. The mortality of CAP remains unacceptably high, in spite of substantial advances in treatment and the emergence of well validated pneumonia severity scoring systems [2–6]. Which mechanisms might be envisaged to interpret the puzzle and paradox? The assessment of severity is crucial in the management of CAP. Therefore, risk prediction model might be one of the culprits. Owing to emerging insights into less accurate mortality prediction, many recalibrated, simplified, or modified pneumonia-specific scores were introduced [7–12].

A new clinical score termed quick sequential [sepsis-related] organ failure assessment (qSOFA), which incorporates respiratory rate of 22/min or greater, altered mentation, and systolic blood pressure (SBP) of 100 mm Hg or less, can rapidly identify adult patients with suspected infection who are likely to have poor outcomes [13,14]. The criteria thresholds of tachypnea and low blood pressure were stricter for pneumonia-specific scores, e.g. pneumonia severity index (PSI) [5], the Infectious Disease Society of America and the American Thoracic Society (IDSA/ATS) minor criteria [3] and CURB-65 (confusion, urea >7 mmol/L, respiratory rate \geq 30/min, low blood pressure, and age \geq 65 years) score [6] than for qSOFA [13,14]. However, it remains unknown which threshold is more accurate.

There is increased recognition of substantial progress in mechanical ventilation. Moreover, higher prevalence of systolic hypertension and higher systolic arterial pressure are undoubted in recent years. As a result, patients with CAP might breathe less rapidly, and systolic arterial pressures of many patients might not drop to less than 90 mm Hg. Therefore, underestimation of mortality might be inevitable were the original cut-off values of severity scoring systems still applied. Consequently, it seems worthwhile to update the criteria thresholds of tachypnea and low blood pressure for pneumonia-specific scores that were proposed years ago, which might orchestrate improvements in predicting mortality and severity in patients with CAP and have implications for more accurate clinical triage decisions.

Therefore, two observational prospective cohort studies were conducted to determine the intriguing hypothesis *via* the derivations and validations of three well established and extensively used pneumonia-specific severity scoring systems with the updating cut-off values for tachypnea and low blood pressure.

Material and methods

Design and setting

A prospective derivation cohort study of 1598 adult patients with CAP was conducted at the Departments of Pulmonary and Critical Care Medicine in two Chinese tertiary hospitals [Shenzhen Hospital, Peking University, and The Eighth Affiliated Hospital (Shenzhen Futian), Sun Yat-sen University] of two universities (Peking University and Sun Yat-sen University) from 1 July 2016, through December 31, 2018. We then performed a prospective validation cohort study of 1659 adult patients with CAP who presented to the hospitals between 1 January 2019 and 30 June 2021. The current database partly overlapped the databases of the articles published [15,16].

Three bettered severity scoring systems were derived from the corresponding pneumonia-specific severity scoring systems, i.e. PSI, IDSA/ATS minor criteria and CURB-65 score, with the updating cut-off values for tachypnea and low blood pressure. Respiratory rate \geq 22/min was substituted for respiratory rate \geq 30/min, and hypotension (SBP <90 mm Hg or diastolic blood pressure \leq 60 mm Hg) was replaced by SBP \leq 100 mm Hg.

Criteria for enrollment

CAP was defined as an acute infection of the pulmonary parenchyma manifested an acute infiltrate on the chest radiograph accompanying with two or more symptoms, such as fever ($>38^{\circ}\text{C}$), hypothermia ($<36^{\circ}\text{C}$), rigors, sweats, new cough or change in color of respiratory secretions, chest discomfort or dyspnoea [17]. CAP excluded any patient who was resided in a nursing home or long-term care facility [3]. Patients younger than 18 years, hospitalized during the 28 days before the study, having severe immunosuppression (human immunodeficiency virus infection, hematologic malignancy, chemotherapy-induced neutropenia, organ transplantation, and so on), active tuberculosis or end-stage diseases, having a written 'do not resuscitate' order, with COVID-19, or unconscious before pneumonia onset were excluded.

Clinical management

These studies were performed according to the principles of human experimentation guidelines of the United States Department of Health and Human Services. Our report was based on the Strengthening the Reporting of Observational Studies in Epidemiology

(STROBE) guidelines. Patients suffering from CAP were cared for by respiratory physicians and intensivists in accordance with the IDSA/ATS guidelines [3] and the Surviving Sepsis Campaign guidelines [18,19]. PSI class \geq IV, IDSA/ATS minor criteria \geq 3, or CURB-65 score \geq 3 was warranted a transfer to respiratory intensive care unit. Antibiotic regimens for the empirical treatment were prescribed based on the guidelines, and then adjusted in the light of subsequently cultured pathogens. All patients clinically stable and afebrile were discharged home [3].

Approval of study design

These studies were endorsed by our Institutional Review Boards (Review Board of Sun Yat-sen University and Review Board of Peking University, No. 20160913 and No. 20161020, respectively). All procedures conducted in these studies involving human participants were based on the 1964 Helsinki Declaration and its later amendments. Ethical approval from the regulation committee was permitted for the study protocol. Written informed consent (except that from unconscious patients) was obtained from all individual participants included in these studies before enrollment. A total of 345 consents was obtained from the family members of the unconscious patients.

Sample size calculation

Unit-level design prevalence, cluster-level design prevalence, test sensitivity, target cluster sensitivity, and target system sensitivity were set to 12%, 1%, 0.9, 0.5, and 0.95, respectively. The total sampled numbers of clusters were 598, and the maximum number of samples was 4,186.

Data collection

A total of 1598 patients were enrolled consecutively and 21 cases were excluded from the derivation cohort due to exclusion criteria, including 6 patients with written 'do not resuscitate' orders on or after admission. 24 cases were excluded from 1659 consecutive patients in the validation cohort, including 8 patients who showed written 'do not resuscitate' orders during admission. All the patients took chest radiographs and/or computer tomography scans. The frontal and lateral chest radiographic findings and computer tomography scan images were classified independently by two senior radiologists (LH Liang and QZ Zhao). We collected clinical and diagnostic data, and

radiological features. Pneumonia-specific severity scores and sequential organ failure assessment (SOFA) scores on admission were calculated. Measurements of laboratory variables were made by the hospital clinical laboratories. The statistician was blinded to these studies.

Outcomes

The main outcome measures incorporated the predictive abilities of the cut-off values of tachypnea and low blood pressure and of the six severity scoring systems for mortality in CAP. Secondary outcome was 30-day mortality.

Statistical analysis

All statistical analyses were conducted with Statistical Package for the Social Science for Windows version 28.0 (IBM SPSS, Chicago, IL, USA) and MedCalc version 20.013 (Mariakerke, Belgium). Categorical variables and continuous variables with normal distributions were described as the percentages and the mean \pm standard deviation (SD), respectively. Chi-Square test, Spearman rank correlation, univariate logistic regression, and one-way ANOVA were adopted. Odds ratio (OR) for mortality were computed. Construct validity was assessed according to the agreement between different measures analogous to the multitrait-multimethod matrix approach of Campbell and Fiske, employing the Cronbach α to measure agreement or commonality [20,21]. Cronbach α values of 0.7 to 0.8 are considered satisfactory. The receiver operating characteristic (ROC) curves were constructed and the corresponding areas under the ROC curves (AUROCs) with the 95% confidence interval (CI) were calculated to assess the mortality prediction of scores. AUROCs were considered poor at 0.6 to 0.7, adequate at 0.7 to 0.8, good at 0.8 to 0.9, and excellent at 0.9 or higher [22]. Predictive capabilities of scores was also evaluated *via* computing net reclassification improvement (NRI) [23–25]. The NRI is regarded as the relative prediction improvement or deterioration when comparing two prediction models. The sensitivities, specificities, positive likelihood ratios (PLRs), negative likelihood ratios (NLRs), positive predictive values (PPVs), negative predictive values (NPVs), and Youden's indices were also counted to estimate robustness of the variables. All testings were two-sided. *P*-values less than 0.05 were regarded as statistically significant.

Results

Patient characteristics

The baseline characteristics of patients were summarized in Table 1. Demographic characteristics, comorbidities, alcohol abuse, smoking, numbers of criteria present, and outcomes were similar between the two cohorts. Intriguingly, most of the non-survivals breathed less than 30/min and did not show hypotension, indicating that underestimation of severity might be inevitable without updating the

Table 1. Baseline characteristics of study cohorts.

Characteristic	Derivation cohort (n = 1577)	Validation cohort (n = 1635)
Age (years)		
Median (95% CI) (IQR)	48 (46–52) (30–76)	48 (47–52) (31–78)
Mean	52.2	52.5
Age ≥65 (No.) (%)	618 (39.2)	637 (39.0)
Sex, (No.) (%)		
Men	694 (44.0)	663 (40.6)
Women	883 (55.0)	972 (59.4)
Comorbidities, No. (%)		
Hypertension	473 (30.0)	659 (28.1)
Coronary heart disease	139 (8.8)	148 (9.1)
Heart failure	50 (3.2)	59 (3.6)
NYHA class IV	27 (1.7)	25 (1.5)
COPD	103 (6.5)	110 (6.7)
GOLD 3 and 4	82 (3.9)	82 (3.9)
Diabetes mellitus	117 (7.4)	119 (7.3)
Chronic renal insufficiency	66 (4.2)	72 (4.4)
Dialysis	36 (2.3)	38 (2.3)
Liver disease	88 (5.6)	98 (6.0)
Nervous system disease	68 (4.3)	64 (3.9)
Tumour	109 (6.9)	119 (7.3)
Alcohol abuse, No. (%)	57 (3.6)	62 (3.8)
Smoking, No. (%)	295 (18.7)	312 (19.1)
Respiratory rate ≥30/min (No.) (%)	235 (14.9)	280 (17.1)
Respiratory rate ≥22/min (No.) (%)	601 (38.1)	617 (37.7)
PaO ₂ /FiO ₂ ≤ 250 mm Hg (No.) (%)	358 (22.7)	406 (24.8)
Multilobar infiltrates (No.) (%)	782 (49.6)	816 (49.9)
Confusion (No.) (%)	192 (12.2)	153 (9.4)
Uremia (No.) (%)	405 (25.7)	393 (24.0)
Leukopenia (No.) (%)	203 (12.9)	228 (13.9)
Thrombocytopenia (No.) (%)	140 (8.9)	114 (7.0)
Hypothermia (No.) (%)	127 (8.1)	131 (8.0)
Hypotension ^a (No.) (%)	423 (26.8)	389 (23.8)
SBP ≤100 mm Hg (No.) (%)	303 (19.2)	296 (17.1)
Outcomes, No. (%)		
Ventilated patients	106 (6.7)	114 (7.0)
Patients received catecholamines	153 (9.7)	165 (10.1)
Hospital length of stay (days)	12.1 ± 6.9	12.4 ± 8.1
30-day mortality	109 (6.9)	115 (7.0)

CI: Confidence interval; IQR: Interquartile range.; NYHA: New York heart association; COPD: Chronic obstructive pulmonary disease; GOLD: Global initiative for chronic obstructive lung disease; PaO₂/FiO₂: Arterial oxygen pressure/fraction inspired oxygen; SBP: Systolic blood pressure.

^aHypotension was defined as an SBP <90 mm Hg or diastolic blood pressure ≤60 mm Hg.

All *p*-values > 0.05.

corresponding cut-off values. The etiology of pneumonia was not detected in each patient. The data were shown in Table 2.

Predictive accuracies and construct validities of the cut-off values for tachypnea and low blood pressure

Respiratory rate ≥22/min and SBP ≤100 mm Hg exhibited higher ORs and more reliabilities and were performed better than respiratory rate ≥30/min and hypotension for the prediction of mortality in the derivation cohort (AUROC, 0.823 vs 0.519, 0.688 vs 0.622; 95% CI, 0.803–0.841 vs 0.494–0.543, 0.664–0.710 vs 0.598–0.646; NRI, 0.61, 0.13; respectively. Tables 3 and 4, and Figure 1). The validation cohort confirmed a similar pattern (Supplementary Figure S1). Bettered-minor criteria and CURB-65 score orchestrated higher convergences in the two cohorts, indicated by greater Cronbach α data compared with the corresponding original systems. Interestingly, Cronbach α decreased sharply if respiratory rate ≥22/min deleted, decreased to a small extent in most severity scoring systems if respiratory rate ≥30/min deleted, decreased slightly if SBP ≤100 mm Hg deleted, and somewhat increased if hypotension deleted. The data corroborated respiratory rate ≥22/min and SBP ≤100 mm Hg demonstrated higher convergences compared with respiratory rate ≥30/min and hypotension, respectively. Pearson correlation coefficients between respiratory rate ≥22/min and the other variables of bettered- minor criteria and CURB-65 score were all more than 0.3, as were SBP ≤100 mm Hg, while the indices for respiratory rate ≥30/min and hypotension were all less than 0.3. Therefore, respiratory rate ≥22/min was associated more closely with acute respiratory failure identified as PaO₂/FiO₂ ≤ 250 mm Hg compared with respiratory rate ≥30/min. Bettered PSI agreed well with PSI (α = 0.987; 95% lower confidence limit [LCL], 0.986) and the other scoring systems (all α data were more than 0.8), as did bettered- minor criteria and CURB-65 score (Supplementary Table S1).

Table 2. Most common etiologies of CAP.

Etiology	Derivation cohort (n = 1577)	Validation cohort (n = 1635)
<i>Streptococcus pneumoniae</i>	407	482
<i>Mycoplasma pneumoniae</i>	284	314
<i>Haemophilus influenzae</i>	155	183
Respiratory viruses	93	107
<i>Staphylococcus aureus</i>	52	76
<i>Legionella</i> species	43	52
Gram-negative bacilli	31	35

CAP: Community-acquired pneumonia.

Table 3. Association of the cut-off values for severe CAP with mortality and the performance for the prediction of mortality among patients with CAP.

Variable	OR (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	PLR (95% CI)	NLR (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Youden's index	AUROC (95% CI)
Derivation cohort (n = 1577)									
Respiratory rate ≥ 30 /min	1.310 (0.790–2.173)	18.3 (11.6–26.9)	85.4 (83.4–87.1)	1.253 (0.827–1.897)	0.957 (0.873–1.048)	8.5 (5.8–12.3)	93.4 (92.8–93.9)	0.04	0.519 (0.494–0.543)
Respiratory rate ≥ 22 /min	105.484 (25.934–429.045)	98.2 (93.5–99.8)	66.3 (63.9–68.8)	2.917 (2.703–3.148)	0.028 (0.007–0.109)	17.8 (16.7–18.9)	99.8 (99.2–99.9)	0.65	0.823 (0.803–0.841)
Hypotension ^a	2.924 (1.973–4.334)	49.5 (39.8–59.3)	74.9 (72.6–77.1)	1.971 (1.599–2.429)	0.674 (0.558–0.814)	12.8 (10.6–15.3)	95.2 (94.3–96.0)	0.24	0.622 (0.598–0.646)
SBP ≤ 100 mm Hg	5.919 (3.964–8.840)	54.1 (44.3–63.7)	83.4 (81.4–85.2)	3.257 (2.647–4.007)	0.550 (0.448–0.675)	19.5 (16.4–22.9)	96.1 (95.2–96.8)	0.38	0.688 (0.664–0.710)
Validation cohort (n = 1635)									
Respiratory rate ≥ 30 /min	1.456 (0.922–2.300)	30.6 (21.0–41.5)	83.3 (81.3–85.1)	1.830 (1.304–2.570)	0.833 (0.722–0.961)	9.3 (6.8–12.6)	95.5 (94.9–96.1)	0.14	0.569 (0.545–0.594)
Respiratory rate ≥ 22 /min	113.897 (28.025–462.897)	98.3 (93.9–99.8)	66.8 (64.4–69.2)	2.963 (2.748–3.196)	0.026 (0.007–0.103)	18.3 (17.2–19.5)	99.8 (99.2–99.9)	0.65	0.826 (0.806–0.844)
Hypotension	1.715 (1.145–2.570)	33.9 (25.3–43.3)	77.0 (74.8–79.1)	1.473 (1.123–1.932)	0.859 (0.751–0.981)	10.0 (7.8–12.8)	93.9 (93.1–94.6)	0.11	0.554 (0.530–0.579)
SBP ≤ 100 mm Hg	5.481 (3.708–8.102)	50.4 (41.0–59.9)	84.3 (82.4–86.1)	3.221 (2.597–3.996)	0.588 (0.488–0.708)	19.6 (16.4–23.2)	95.7 (94.9–96.4)	0.35	0.674 (0.651–0.697)

CAP: Community-acquired pneumonia; OR: Odds ratio; CI: Confidence interval; PLR: Positive likelihood ratio; NLR: Negative likelihood ratio; PPV: Positive predictive value; NPV: Negative predictive value; AUROC: The area under the receiver operating characteristic curve; SBP: Systolic blood pressure.

^a Hypotension was defined as a SBP < 90 mm Hg or diastolic blood pressure ≤ 60 mm Hg.

Table 4. Reliability of the scoring systems.

Cronbach α (Pearson correlation coefficient)	IDSA/ATS minor criteria		Bettered minor criteria		CURB-65 score		Bettered CURB-65 score	
	0.613	0.645	0.695	0.724	0.483	0.472	0.637	0.627
Confusion	0.558 (0.420)		0.659 (0.442)		0.361 (0.394)		0.574 (0.435)	
	0.613 (0.360)		0.704 (0.387)		0.395 (0.317)		0.588 (0.373)	
Uremia	0.535 (0.461)		0.618 (0.592)		0.339 (0.379)		0.470 (0.603)	
	0.595 (0.409)		0.670 (0.544)		0.338 (0.356)		0.452 (0.604)	
Respiratory rate ≥ 30 /min	0.599 (0.239)		..		0.496 (0.140)		..	
	0.625 (0.290)		..		0.430 (0.232)		..	
Respiratory rate ≥ 22 /min	..		0.614 (0.594)		..		0.522 (0.501)	
	..		0.645 (0.641)		..		0.463(0.563)	
Hypotension ^a	0.627 (0.156)		..		0.524 (0.121)		..	
	0.670 (0.115)		..		0.562 (0.035)		..	
SBP ≤ 100 mm Hg	..		0.653 (0.350)		..		0.605 (0.314)	
	..		0.691 (0.305)		..		0.595 (0.325)	
PaO ₂ /FiO ₂ ≤ 250 mm Hg	0.514 (0.535)		0.628 (0.557)		
	0.547 (0.574)		0.656 (0.607)		
Multilobar infiltrates	0.535 (0.452)		0.663 (0.409)		
	0.512 (0.646)		0.654 (0.601)		
Leukopenia	0.602 (0.220)		0.685 (0.280)		
	0.619 (0.318)		0.701 (0.394)		
Thrombocytopenia	0.604 (0.208)		0.686 (0.275)		
	0.648 (0.153)		0.728 (0.187)		
Hypothermia	0.649 (–0.058)		0.725 (–0.034)		
	0.673 (–0.007)		0.746 (0.036)		
Age ≥ 65 years		0.386 (0.316)		0.594 (0.375)	
		0.308 (0.380)		0.532 (0.456)	

IDSA/ATS: The Infectious Disease Society of America and the American Thoracic Society; CURB-65: Confusion; urea > 7 mmol/L, respiratory rate ≥ 30 /min, low blood pressure, and age ≥ 65 years. SBP: Systolic blood pressure; PaO₂/FiO₂: Arterial oxygen pressure/fraction inspired oxygen.

Cronbach α data in blue-shaded shorter cells derived from the derivation cohort, while the right side of the cells was the data from the validation cohort. The data in blue-shaded longer cells derived from the derivation cohort that included Cronbach α data if item deleted plus corrected Item-Total correlations encapsulated in parentheses, while below the cells were the data from the validation cohort.

^aHypotension was defined as a SBP < 90 mm Hg or diastolic blood pressure ≤ 60 mm Hg.

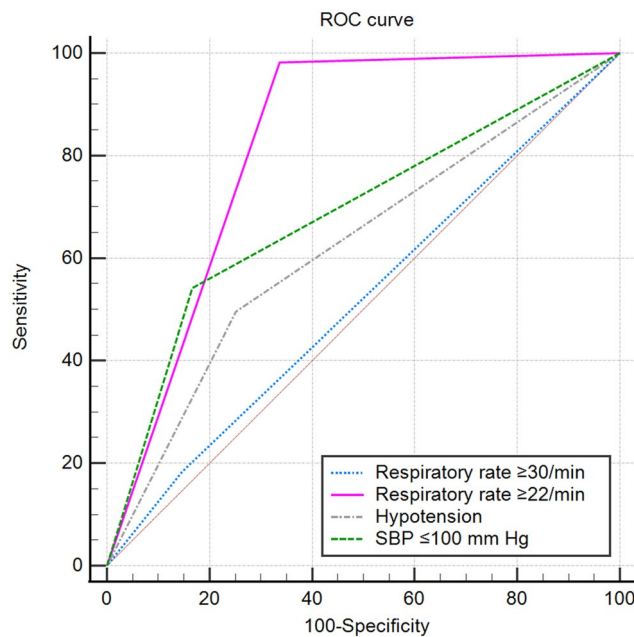


Figure 1. ROC curves for mortality prediction by the cut-off values for tachypnea and low blood pressure in the derivation cohort. ROC: The receiver operating characteristic; SBP: Systolic blood pressure; Hypotension was defined as a SBP <90 mm Hg or diastolic blood pressure ≤60 mm Hg.

SOFA scores according to the six scoring systems

SOFA scores increased significantly with the numbers present or scores of six scoring systems in the two cohorts (Supplementary Tables S2, S3 and S4), all the differences between the classes of PSI and bettered PSI were significant ($p < 0.001$), as were most of the differences between the ranks of other four scoring systems. Most importantly, the rank correlations improved while updating the cut-off values.

Relationships between the ranks of six scoring systems and risk of mortality

The mortalities in the derivation cohort demonstrated nearly stepwise increases with the categories of six scoring systems ($p < 0.001$ for all. Supplementary Table S5). The validation cohort confirmed a similar paradigm. Interestingly, the relationships also improved while updating the cut-off values, indicated by higher r_s .

Comparisons of the scoring systems for predicting mortality

The sensitivities, specificities, and predictive values of the ranks of six scoring systems for predicting mortality were shown in (Supplementary Tables S6, S7 and S8. Youden's index of and AUROC for PSI class ≥5

improved significantly while updating the cut-off values for tachypnea and low blood pressure in the two cohorts, whereas there were almost no improvements in the other PSI classes. IDSA/ATS minor criteria ≥4 and CURB-65 score ≥3 confirmed similar paradigms with additional improvements in PLR and PPV. Furthermore, the predictive abilities became bad if the cut-off values moved up or down. Therefore, bettered PSI class ≥5, the presence of 4 or more bettered minor criteria, and bettered CURB-65 score ≥3 might be more valuable cut-off values of the corresponding scoring systems for severe CAP, while the thresholds of three original scoring systems for severe CAP were not changed, except for CURB-65 score in the validation cohort (the score ≥2 seemed more valuable).

The predictive accuracy of bettered PSI was excellent for mortality (AUROC = 0.939; 95% CI, 0.926–0.951) and was statistically greater than PSI (AUROC = 0.883; 95% CI, 0.866–0.899; $p < 0.0001$. NRI, 0.113. (Supplementary Table S9, and Figure 2), as were bettered- minor criteria and CURB-65 score (AUROC, 0.909 vs 0.871, 0.913 vs 0.859; 95% CI, 0.894–0.923 vs 0.853–0.887, 0.898–0.927 vs 0.841–0.876; $P < 0.0001$, < 0.0001 ; NRI, 0.076, 0.108; respectively). The validation cohort confirmed a similar pattern (AUROC, 0.910 vs 0.862, 0.837 vs 0.760, 0.889 vs 0.832; 95% CI, 0.895–0.923 vs 0.844–0.878, 0.818–0.855 vs 0.739–0.781, 0.873–0.904 vs 0.813–0.850; $P < 0.0001$, < 0.0001 , < 0.0001 ; NRI, 0.097, 0.154, 0.115; respectively (Supplementary Table S10 and Figure S2).

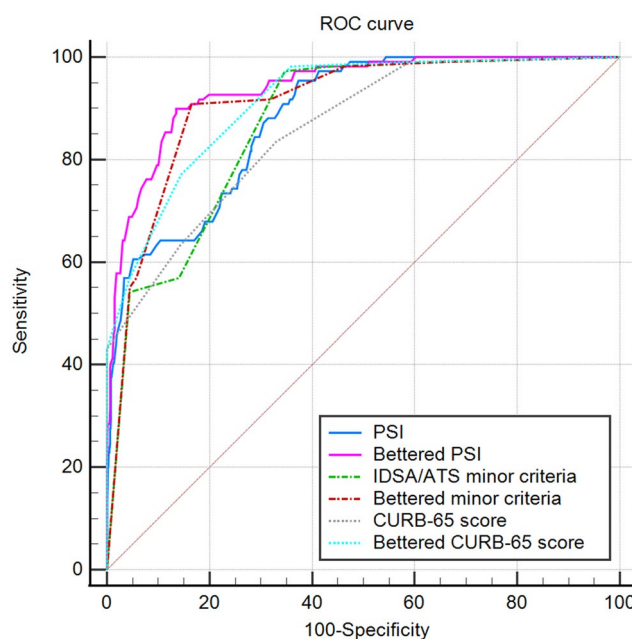


Figure 2. ROC curves for mortality prediction by the six scoring systems in the derivation cohort. ROC: The receiver operating characteristic. PSI: Pneumonia severity index. IDSA/ATS: The Infectious Disease Society of America and the American Thoracic Society. CURB-65: Confusion, urea >7 mmol/L, respiratory rate ≥ 30 /min, low blood pressure, and age ≥ 65 years.

Discussion

The main findings of the current studies comprise the following: Respiratory rate ≥ 22 /min and SBP ≤ 100 mm Hg exhibited higher ORs and performed better than respiratory rate ≥ 30 /min and hypotension for the prediction of mortality, respectively. Bettered-minor criteria and CURB-65 score orchestrated higher convergences, indicated by greater Cronbach α and more decrease in Cronbach α if the updating cut-off values deleted. The six scoring systems agreed well with one another. Bettered scoring systems showed higher associations with SOFA scores and mortality rates and orchestrated significant improvements in predicting mortality in the grades of PSI class ≥ 5 , IDSA/ATS minor criteria ≥ 4 and CURB-65 score ≥ 3 , which might be more valuable cut-off values of the corresponding scoring systems for severe CAP. Bettered-PSI, minor criteria and CURB-65 score demonstrated greater predictive accuracies for mortality compared with the corresponding original systems.

The prognosis for patients with CAP ranges from rapid recovery to death [5,26]. The great variability seen in rates of hospital admission and lengths of stay for pneumonia in part reflects uncertainty among physicians in assessing the severity of this illness and the perceived benefits of hospital care [5]. Emerging prediction rules were designed to reduce such uncertainty and to foster more appropriate use of hospitals in the management of this illness [3,5,6]. The situations have

been changing, were the cut-off values for variables of the scoring systems warranted to be updated for better clinical practice?

The current study shed light on the questions about the criteria thresholds. Only 20 of 109 non-survivals breathed more than 30/min and 54 non-survivals demonstrated hypotension in the derivation cohort, while 107 non-survivals breathed 22/min or greater and 59 non-survivals showed SBP of 100 mm Hg or less. Underestimation of mortality might be tremendous were the original cut-off values of severity scoring systems still employed. The validation cohort confirmed a similar paradigm. Therefore, respiratory rate ≥ 22 /min and SBP ≤ 100 mm Hg better performed in predicting mortality, as manifested by that significant improvement of AUROC and NRI were orchestrated. As CAP is often complicated by sepsis [27], the success of parsimonious qSOFA in predicting mortality in sepsis seems to corroborate our findings to some extent [13,14,28,29].

When items are used to form a scale they need to have internal consistency. The items should all measure the same thing, so they should be correlated with one another [20]. Higher agreement or commonality of the variables plays a pivotal role in the greater predictive accuracy of a predicting system. The updating cut-off values orchestrated higher convergence, indicated by greater Cronbach α and more decrease in Cronbach α if the updating cut-off values deleted. Cronbach α

data with variables of the four scoring systems were almost less than 0.7 even if an item deleted, but the six scoring systems agreed well with one another, indicating that the three updating scoring systems might be reasonable and clinically practicable. The generalizability of the results needs to be ascertained.

It was intriguing that the updating cut-off values only orchestrated significant improvements in predicting mortality in the grades of PSI class ≥ 5 , IDSA/ATS minor criteria ≥ 4 and CURB-65 score ≥ 3 . Which mechanisms might be envisaged to interpret the phenomena? 100 (91.7%) non-survivals breathed between 30/min and 22/min in the derivation cohort, while 95 (82.6%) in the validation cohort. Systolic arterial pressures of 59 (54.1%) non-survivals dropped to less than 100 mm Hg, but still ≥ 90 mm Hg in the derivation cohort, and 49 (42.6%) in the validation cohort. The findings revealed that updating cut-off values might be more valuable to discriminate between survival and non-survival. As a consequence, it might not be robust enough to identify severity during the implementation to severe patients or non-severe patients. This is in consonance with the concept that the criteria threshold possesses the best predictive ability and becomes bad if moving up or down. The current studies authenticate the concept in both the original and updated systems.

Mean ages in the current cohorts were younger than those in the cohorts developing CURB-65 score (64.1 years) and PSI (Age <50 years; 16.7%, 15.5% and 42.7% in three cohorts, respectively) and validating IDSA/ATS minor criteria (65.7 years) [5,6,17]. The dissimilarity in the age of our patients to the original PSI and CURB-65 cohorts and minor criteria validation cohort suggests that our findings might unreasonably be extrapolated to most CAP populations. The comorbidities of our patients were less than those in the original PSI and CURB-65 cohorts and minor criteria validation cohort [5,6,17], in additional consideration of the reduced mean age of the current populations, the impact of comorbidities may have been less influential on mortality rate than the hemodynamic or respiratory complications in our cohorts. Therefore, future prospective multicenter cohort studies are warranted to assess its generalizability.

Space is changing. Chalmers et al. [30] asserted it may be necessary to perform local recalibration of the score based on the different populations. Schuetz et al. [31] discovered that the need for recalibration should be studied in order to avoid substantial underestimation of mortality and misclassification of patients. Furthermore, time is changing. Human physique, physiological functions and pathophysiological indices, environments, and treatments are inevitably changing,

it is scarcely possible to keep the criteria thresholds permanent. Inaccuracy of criteria thresholds (e.g. lung function indices) in diagnosis or prediction is looming. It is gratifying that a call for change is emerging. If current diagnostic criteria remain unchanged, the identified shifts in European values will allow the easier fulfilment of diagnostic criteria for lung diseases such as COPD, but the systematic underestimation of lung disease severity [32]. It is widely accepted that prognostic model development is a three-stage process, comprising derivation, validation and impact analysis (applying the rule and determining if it can improve clinical outcomes for patients) [30,33]. Therefore, future prospective multicenter cohort studies of patients with CAP are needed to assess the impact of any updating pneumonia severity score on clinical outcomes, especially the rationalization of clinical triage decision-making (i.e. resources optimization and identification of subjects requesting admission to intensive care unit) and reduction in mortality.

Limitations

Our findings must be interpreted in light of several limitations. First, the prospective cohorts were derived from two centers, but not multicenter settings. Generalization of these findings to all patients with CAP should be made with caution. Second, there were relatively small samples. Had the numbers been larger, perhaps the results might have been more robust. Third, only more than one hundred patients met the secondary outcome, which may be considered relatively low. Fourth, the median age was relatively younger. Fifth, this study focused on adult patients only and did not evaluate children at risk for severe CAP.

Conclusions

Updating cut-off values of severity scoring systems for CAP orchestrate improvement in predictive accuracy, suggesting that it may facilitate the rationalization of clinical triage decision-making and further reduce mortality. The current studies provide the first known prospective evidence of the potential benefit of the updating cut-off values of severity scoring systems for CAP in predictive accuracy.

Acknowledgments

We are indebted to the nurses, further education physicians, and postgraduates of the Departments of Pulmonary and Critical Care Medicine for making contributions to these studies.

Authors' contributions

Q.G had full access to all the data in these studies and takes responsibility for the integrity of the data and the accuracy of the data analysis. Q.G was in charge of funding acquisition and project administration. Q.G, H-Y.L and W-D.S made substantial contributions to conception and design and were in charge of data collection and curation, and the writing of the manuscript. L-H.L and Q-Z.Z read the chest radiographs and computer tomography scans. M.L, X-K.C, H.L, H-L.P, H-Q.Y, N.L, Y-H.L, Z-D.L, L-H.L and Q-Z.Z made substantial contributions to acquisition of data. M.J was in charge of statistical analysis. Each author has participated in the writing of the manuscript, been involved in the analysis of the data, and seen and approved the submitted version.

Disclosure statement

No potential conflict of interest was reported by the authors.

Data availability

The datasets used and/or analysed during the current studies are available from the corresponding author upon reasonable request.

Funding

This work was supported by the Medical Science and Technology foundation of Guangdong province [grant number A2010553]; the Planned Science and Technology project of Shenzhen municipality [grant number 201102078]; and the non-profit Scientific Research Project of Futian District [grant number FTWS201120].

Role of the Funding Source: The funding sources had no role in the design and conduct of these studies; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication.

ORCID

Qi Guo  <http://orcid.org/0000-0001-7684-4072>

References

- [1] Prina E, Ranzani OT, Torres A. Community-acquired pneumonia. *Lancet*. 2015;386(9998):1–10.
- [2] Metlay JP, Waterer GW, Long AC, et al. Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American thoracic society and infectious diseases society of America. *Am J Respir Crit Care Med*. 2019;200(7):e45–e67.
- [3] Mandell LA, Wunderink RG, Anzueto A, et al. Infectious diseases society of America/American thoracic society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis*. 2007;44(Supplement_2):S27–S72.
- [4] Guo Q, Li HY, Li YM, et al. Compliance with severe sepsis bundles and its effect on patient outcomes of severe community-acquired pneumonia in a limited resources country. *Arch Med Sci*. 2014;10(5): 970–978.
- [5] Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med*. 1997;336(4):243–250.
- [6] Lim WS, van der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax*. 2003;58(5):377–382.
- [7] Elmoheen A, Abdelhafez I, Salem W, et al. External validation and recalibration of the CURB-65 and PSI for predicting 30-day mortality and critical care intervention in multiethnic patients with COVID-19. *Int J Infect Dis*. 2021;111:108–116.
- [8] Salih W, Schembri S, Chalmers JD. Simplification of the IDSA/ATS criteria for severe community acquired pneumonia using meta-analysis and observational data. *Eur Respir J*. 2014;43(3):842–851.
- [9] Li HY, Guo Q, Song WD, et al. Modified IDSA/ATS minor criteria for severe community-acquired pneumonia best predicted mortality. *Medicine*. 2015;94(36):e1474.
- [10] Ahn JH, Choi EY. Expanded A-DROP score: a new scoring system for the prediction of mortality in hospitalized patients with community-acquired pneumonia. *Sci Rep*. 2018;8(1):14588. .
- [11] Li HY, Guo Q, Song WD, et al. CUR-65 score for community-acquired pneumonia predicted mortality better than CURB-65 score in low-mortality-rate settings. *Am J Med Sci*. 2015;350(3):186–190.
- [12] Pflug MA, Tiutan T, Wesemann T, et al. Short-term mortality of adult inpatients with community-acquired pneumonia: external validation of a modified CURB-65 score. *Postgrad Med J*. 2015;91(1072):77–82.
- [13] Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*. 2016;315(8):801–810.
- [14] Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*. 2016;315(8):762–774.
- [15] Guo Q, Song WD, Li HY, et al. Cold-inducible RNA-binding protein might determine the severity and the presences of major/minor criteria for severe community-acquired pneumonia and best predicted mortality. *Respir Res*. 2020;21(1):192.
- [16] Guo Q, Li HY, Song WD, et al. qSOFA predicted pneumonia mortality better than minor criteria and worse than CURB-65 with robust elements and higher convergence. *Am J Emerg Med*. 2022;52:1–7.
- [17] Phua J, See KC, Chan YH, et al. Validation and clinical implications of the IDSA/ATS minor criteria for severe community-acquired pneumonia. *Thorax*. 2009;64(7): 598–603.
- [18] Dellinger RP, Levy MM, Rhodes A, et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive Care Med*. 2013;39(2):165–228.

- [19] Rhodes A, Evans LE, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Intensive Care Med.* 2017;43(3):304–377.
- [20] Bland JM, Altman DG. Cronbach's alpha. *BMJ.* 1997;314(7080):572.
- [21] Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull.* 1959;56(2):81–105.
- [22] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143(1):29–36.
- [23] Pencina MJ, D'Agostino RB, D'Agostino RB, Jr, et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* 2008;27(2):157–172.
- [24] Pencina MJ, D'Agostino RB, Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med.* 2011;30(1):11–21.
- [25] Leening MJ, Vedder MM, Witteman JC, et al. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. *Ann Intern Med.* 2014;160(2):122–131.
- [26] Fine MJ, Smith MA, Carson CA, et al. Prognosis and outcomes of patients with community-acquired pneumonia. A meta-analysis. *JAMA.* 1996;275(2):134–141.
- [27] Dremsizov T, Clermont G, Kellum JA, et al. Severe sepsis in community-acquired pneumonia: when does it happen, and do systemic inflammatory response syndrome criteria help predict course? *Chest.* 2006;129(4):968–978.
- [28] Freund Y, Lemachatti N, Krastinova E, et al. Prognostic accuracy of sepsis-3 criteria for in-hospital mortality among patients with suspected infection presenting to the emergency department. *JAMA.* 2017;317(3):301–308.
- [29] Rudd KE, Seymour CW, Aluisio AR, et al. Association of the quick sequential (sepsis-related) organ failure assessment (qSOFA) score with excess hospital mortality in adults with suspected infection in low- and middle-income countries. *JAMA.* 2018;319(21):2202–2211.
- [30] Chalmers JD, Singanayagam A, Akram AR, et al. Severity assessment tools for predicting mortality in hospitalised patients with community-acquired pneumonia. Systematic review and meta-analysis. *Thorax.* 2010;65(10):878–883.
- [31] Schuetz P, Koller M, Christ-Crain M, et al. Predicting mortality with pneumonia severity scores: importance of model recalibration to local settings. *Epidemiol Infect.* 2008;136(12):1628–1637.
- [32] Allinson JP, Afzal S, Çolak Y, et al. Changes in lung function in European adults born between 1884 and 1996 and implications for the diagnosis of lung disease: a cross-sectional analysis of ten population-based studies. *Lancet Respir Med.* 2022;10(1):83–94.
- [33] McGinn TG, Guyatt GH, Wyer PC, et al. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-based medicine working group. *JAMA.* 2000;284(1):79–84.