ELSEVIER

Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Clinical Microbiology and Infection 27 (2021) 1037.e1–1037.e8

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Original article

Community-acquired pneumonia severity assessment tools in patients hospitalized with COVID-19: a validation and clinical applicability study

Felippe Lazar Neto [1, *, †], Lucas Oliveira Marino [1, †], Antoni Torres [2], Catia Cilloniz [2, 3], Julio Flavio Meirelles Marchini [1], Julio Cesar Garcia de Alencar [1], Andrea Palomeque [2, 3], Núria Albacar [2, 3], Rodrigo Antônio Brandão Neto [1], Heraldo Possolo Souza [1], Otavio T. Ranzani [4, 5], COVID Registry Team[§],

[1) Emergency Medicine Department, Faculdade de Medicina da Universidade de São Paulo (FMUSP), São Paulo, Brazil]
[2) Department of Pneumology, Hospital Clinic of Barcelona, August Pi i Sunyer Biomedical Research Institute - IDIBAPS, University of Barcelona, Biomedical Research Networking Centers in Respiratory Diseases (CIBERES, CIBERESUCICOVID), Barcelona, Spain]
[3) Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain]
[4) Barcelona Institute for Global Health, ISGlobal, Barcelona, Spain]
[5) Pulmonary Division, Heart Institute (InCor), Hospital Das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil]

ARTICLE INFO

Article history:
Received 16 January 2021
Received in revised form
16 March 2021
Accepted 20 March 2021
Available online 2 April 2021

Editor: A. Kalil

Keywords:
Coronavirus
Coronavirus disease 2019
Mortality
Pneumonia
Prediction
Prognosis
Severity
Validation

ABSTRACT

Objective: To externally validate community-acquired pneumonia (CAP) tools on patients hospitalized with coronavirus disease 2019 (COVID-19) pneumonia from two distinct countries, and compare their performance with recently developed COVID-19 mortality risk stratification tools.

Methods: We evaluated 11 risk stratification scores in a binational retrospective cohort of patients hospitalized with COVID-19 pneumonia in São Paulo and Barcelona: Pneumonia Severity Index (PSI), CURB, CURB-65, qSOFA, Infectious Disease Society of America and American Thoracic Society Minor Criteria, REA-ICU, SCAP, SMART-COP, CALL, COVID GRAM and 4C. The primary and secondary outcomes were 30-day in-hospital mortality and 7-day intensive care unit (ICU) admission, respectively. We compared their predictive performance using the area under the receiver operating characteristics curve (AUC), sensitivity, specificity, likelihood ratios, calibration plots and decision curve analysis.

Results: Of 1363 patients, the mean (SD) age was 61 (16) years. The 30-day in-hospital mortality rate was 24.6% (228/925) in São Paulo and 21.0% (92/438) in Barcelona. For in-hospital mortality, we found higher AUCs for PSI (0.79, 95% CI 0.77–0.82), 4C (0.78, 95% CI 0.75–0.81), COVID GRAM (0.77, 95% CI 0.75–0.80) and CURB-65 (0.74, 95% CI 0.72–0.77). Results were similar for both countries. For the 1%–20% threshold range in decision curve analysis, PSI would avoid a higher number of unnecessary interventions, followed by the 4C score. All scores had poor performance (AUC <0.65) for 7-day ICU admission.

Conclusions: Recent clinical COVID-19 assessment scores had comparable performance to standard pneumonia prognostic tools. Because it is expected that new scores outperform older ones during development, external validation studies are needed before recommending their use. **Felippe Lazar Neto, Clin Microbiol Infect 2021;27:1037.e1–1037.e8**

© 2021 The Author(s). Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

---

[*] **Corresponding author:** F. Lazar Neto, Emergency Medicine Department, Faculdade de Medicina da Universidade de São Paulo (FMUSP), Avenida Dr Enéas Carvalho de Aguiar, 255, São Paulo, Brazil.

E-mail address: felippe.neto@alumni.usp.br (F. Lazar Neto).

[†] Felippe Lazar Neto and Lucas Oliveira Marino contributed equally.

[§] The members of COVID Registry Team are listed at Acknowledgements section

https://doi.org/10.1016/j.cmi.2021.03.002

1198-743X/© 2021 The Author(s). Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has infected more than 110 million people and killed nearly 2.5 million worldwide [1]. Although most patients have mild limited symptoms, 15% complain of dyspnoea and 5% present with hypoxaemic respiratory failure, shock or multiorgan dysfunction [2]. Identifying patients who will need advanced support or who are at high risk of poor outcomes challenges physicians. To help decision-making, researchers developed several risk assessment tools specifically for coronavirus disease 2019 (COVID-19); however, most scores had important limitations during development: poor report, over-optimism and high risk of bias [3,4]. In addition, external validation is needed before implementation in routine clinical practice.

Community-acquired pneumonia (CAP) is a common infection and a leading cause of mortality [5,6]. Over the past decades, risk stratification tools have improved CAP clinical management [7]. Unlike COVID-19 prediction rules, CAP scores were extensively validated [8] with some already evaluated on COVID-19 with promising results [9–11]. We evaluated CAP and COVID-19 risk assessment scores on a binational cohort of hospitalized patients with COVID-19 pneumonia in São Paulo and Barcelona during the initial pandemic surge. We hypothesized that CAP prediction rules would have similar performance to the recently developed COVID-19 scores.

Methods

Study design and population

We retrospectively analysed patients with COVID-19 pneumonia admitted to the emergency department of two university hospitals: Hospital das Clínicas (from 14 March to 14 June 2020) and Hospital Clinic (from 28 February to 5 May 2020). Both hospitals were designated to be the tertiary reference for COVID-19 suspected cases in their respective cities: São Paulo (Brazil) and Barcelona (Spain). Both ethics committees approved the studies protocols (CAAE 30417520.0.0000.0068 and Register HCB/2020/0273).

We defined COVID-19 pneumonia as a new infection-compatible infiltrate on lung CT or chest X-ray associated with acute inferior respiratory tract infection symptoms. All patients were admitted and treated according to the institutional protocol. A real-time quantitative RT-PCR (RT-qPCR) test of samples from the upper (nasopharyngeal or oropharyngeal) or lower (endotracheal) respiratory tract was collected to confirm SARS-CoV-2 infection. A standardized form was used for data collection, which included questions on demographics, past medical history, clinical examination and vital signs. The clinical information was retrieved from the first medical assessment in the emergency department and laboratory tests were taken from the first available result up to 48 h after admission. Collected variables had similar definitions for both cohorts and harmonization between cohorts was elementary. The Barcelona cohort included only positive RT-qPCR results, whereas the São Paulo cohort included RT-qPCR-positive cases and patients with a clinical–epidemiological diagnosis (see Supplementary material, Appendix S1) as RT-qPCR was not widely available. Sensitivity analysis on only RT-qPCR-positive patients was performed and is presented in the Supplementary material.

Scores selections and definitions

We applied the following risk assessment scores according to admission variables: Pneumonia Severity Index (PSI) [12], CURB [13], CURB-65 [13], IDSA/ATS Minor Criteria [14], quick Sepsis Related Organ Failure Assessment (qSOFA) [15,16], Severe Community Acquired Pneumonia (SCAP) [17], SMART-COP [18], The Risk of Early Admission to ICU index (REA-ICU) [19], COVID-GRAM [20], CALL [21] and 4C [22]. We used their original descriptions (see Supplementary material, Appendix S2). The cut-off values for each score were chosen based on the development report if available or the standard use. A 10% risk threshold was selected for COVID-GRAM based on similar risk prediction tools [18,19].

We considered the need for supplemental oxygen therapy or peripheral oxygen saturation <92% equivalent to documented laboratory hypoxaemia ($\mathrm{po}_2 < 60$ mmHg) when deriving scores that included hypoxaemia. Variables that were not in the database and consequently could not be imputed were assigned zero for risk calculation and are specified in the Supplementary material (Appendix S2).

Outcomes

Our primary outcome was in-hospital mortality at 30 days. Patients still hospitalized at 30 days were considered alive. The secondary outcome was admission to intensive care unit (ICU) until the 7th day (excluding those individuals who were on mechanical ventilation or vasoactive drugs before hospital admission).

Statistical analysis

Mean, standard deviation (SD), median and interquartile range (IQR) were used for descriptive statistics according to variable distribution.

We defined *a priori* the statistical analysis plan. We expected a great proportion of missing values due to the large number of risk scores tested and the wide range of different variables considered by each score. We performed a single imputation procedure with chained equations, assuming a missing-at-random pattern, in which missing values are conditional on measured variables. We used predictive mean matching because of its flexibility for imputation of different types of variables [23]. Outcome and country were included as predictors during the imputation process. The Table S1 (see Supplementary material) provides the missing percentage descriptive statistics.

Model predictive performance was assessed with the area under the receiver operating characteristics curve (AUC) and the Brier score. The Brier score is an overall model fit metric, combining both discrimination and calibration aspects. The Brier score is better when the values are closer to 0 ('perfect model'). Calibration was evaluated using calibration plots sub-divided in quintiles of predicted probabilities. Clinical utility was analysed using sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio and negative likelihood ratio. Confidence intervals (95%) were calculated after 1000 bootstrap re-samples.

To incorporate the clinical decision reasoning in model evaluation, we used the decision curve analysis framework [24] where predictive models can be compared with common strategies of treating all or none of the patients. To accomplish this, we calculated the net-benefit for each strategy by subtracting the proportion of false positives from the true positives, weighted by the relative harm of a false-positive and a false-negative result. In short, we take into account how much the physician is willing to treat more false-positive patients to avoid not treating true-negative patients [24,25]. The net-benefit of treated patients is the result of subtracting the net-benefit of the evaluated model from the net-benefit of the treating all strategy. This number is then used for computing the number of avoidable interventions per 100 patients. For this study, intervention would be optimization of hospital resources (intensity of care), which is the ultimate decision goal when applying mortality risk stratification tools at the emergency department. We restricted probability thresholds between 0.01 and 0.2 (99:1 and 4:1 false-positive/false-negative weights, respectively) as is commonly done for infectious diseases, including pneumonia. The decision curve analysis and calibration were restricted to in-hospital mortality.

We followed the Transparent Reporting of a Multivariable Prediction for Individual Prognosis or Diagnosis (TRIPOD) framework [26]. All statistical analyses were performed in R version 3.6.2.

Results

Patient characteristics

Table 1 shows baseline admission data. Of 1363 patients, 77.2% (1053/1363) had positive RT-qPCR (São Paulo 66.4%, 615/925 and Barcelona 100%, 438/438). The mean age was 61 years and most were male (59.2%, 807/1363). The Barcelona cohort had older patients compared with São Paulo. The most common co-morbidity was hypertension (51.8%, 706/1363), followed by diabetes (31.1%, 424/1363), cancer (10.1%, 138/1363) and congestive heart failure (7.7%, 105/1363). We observed comparable proportions between cohorts for hypertension, congestive heart failure, cancer, asthma and chronic obstructive pulmonary disease, but not for diabetes, hepatic dysfunctions and autoimmune diseases. Overall, patients had decreased circulating lymphocytes and elevated D-dimer, C-reactive protein and lactic dehydrogenase. The 30-day in-hospital mortality was 23.5% (320/1363) (São Paulo 24.6%, 228/925 and Barcelona 21.0%, 92/438), overall ICU admission 47.4% (646/1363) (São Paulo 52.6%, 487/925 and Barcelona 36.3%, 159/438) and 7-day ICU admission 36% (410/1137) (São Paulo 37.8%, 272/719 and Barcelona 33.0%, 138/418).

Table 1

Characteristics of patients at admission and outcomes for the imputed dataset by country

	All ($N = 1363$)	Brazil ($n = 925$)	Spain ($n = 438$)
Age (years), mean (SD)	61.05 (16.39)	59.25 (15.5)	64.83 (17.54)
Sex			
Male	807 (59.21%)	522 (56.43%)	285 (65.07%)
Life habits			
Active smokers	95 (6.97%)	56 (6.05%)	39 (8.9%)
Ex-smokers	429 (31.47%)	345 (37.3%)	84 (19.18%)
Obesity	180 (13.21%)	155 (16.76%)	25 (5.71%)
Co-morbidities			
Hypertension	706 (51.8%)	487 (52.65%)	219 (50%)
Diabetes	424 (31.11%)	334 (36.11%)	90 (20.55%)
Asthma	42 (3.08%)	28 (3.03%)	14 (3.2%)
Chronic obstructive pulmonary disease	51 (3.74%)	35 (3.78%)	16 (3.65%)
Congestive heart failure	105 (7.7%)	78 (8.43%)	27 (6.16%)
Stroke	53 (3.89%)	35 (3.78%)	18 (4.11%)
Chronic kidney disease	30 (2.2%)	25 (2.7%)	5 (1.14%)
Cirrhosis	40 (2.93%)	9 (0.97%)	31 (7.08%)
Cancer	138 (10.12%)	91 (9.84%)	47 (10.73%)
HIV	14 (1.03%)	13 (1.41%)	1 (0.23%)
Autoimmune disorders	35 (2.57%)	10 (1.08%)	25 (5.71%)
Other cardiovascular diseases	186 (13.65%)	95 (10.27%)	91 (20.78%)
Other respiratory diseases	68 (4.99%)	26 (2.81%)	42 (9.59%)
Days of symptoms, median (IQR)	7 (5–10)	7 (5–10)	6 (4–8)
Vital signs[a]			
Systolic pressure (mmHg), mean (SD)	125.77 (23.05)	125.31 (23.29)	126.73 (22.53)
Diastolic pressure (mmHg), mean (SD)	74.81 (14.12)	75.43 (15)	73.5 (11.97)
Respiratory rate, mean (SD)	24.5 (6.93)	25.5 (7.05)	22.39 (6.16)
Heart rate, mean (SD)	90.32 (17.21)	90.16 (16.77)	90.67 (18.12)
Temperature (°C), mean (SD)	36.83 (1.2)	36.63 (1.22)	37.25 (1.03)
Oxygen saturation (%), mean (SD)	93.57 (4.51)	92.99 (4.61)	94.8 (4.04)
Laboratory results[a]			
Creatinine (mg/dL), median (IQR)	0.94 (0.73–1.4)	0.94 (0.71–1.42)	0.95 (0.78–1.35)
Urea (mg/dL), median (IQR)	36 (24–61)	38 (25–64)	32 (22–55.75)
Haematocrit (g/dL), median (IQR)	38 (33.9–42)	37.1 (32.7–40.6)	40.55 (36.85–44)
Leucocytes (× 1000/mm^3), median (IQR)	7.15 (5.26–10.31)	7.88 (5.68–10.85)	6.1 (4.42–8.7)
Lymphocytes (× 1000/mm^3), median (IQR)	0.88 (0.6–1.21)	0.96 (0.66–1.32)	0.7 (0.5–1)
Arterial pH, median (IQR)	7.43 (7.38–7.46)	7.42 (7.37–7.46)	7.45 (7.42–7.47)
Arterial po$_2$ (mmHg), median (IQR)	67.1 (58.6–82.9)	67.3 (58.8–83.3)	67.05 (57.45–81.8)
Arterial pCo$_2$ (mmHg), median (IQR)	36.4 (32.7–41.8)	37.4 (32.9–43.3)	35 (32–39.68)
Albumin (g/dL), median (IQR)	3.3 (2.9–3.6)	3.2 (2.8–3.5)	3.4 (3–3.7)
D-Dimer (ng/dL), median (IQR)	1100 (600–2470)	1278 (676–3314)	900 (500–1900)
C-reactive protein (mg/L), median (IQR)	115.3 (61.6–211.4)	136.8 (71.5–225.9)	84.35 (40.3–167.62)
Lactic dehydrogenase (U/L), median (IQR)	365 (283–485)	378 (293–509)	338 (268.25–437.25)
RT-qPCR confirmed for SARS-CoV-2	1053 (77.25%)	615 (66.48%)	438 (100%)
Outcomes			
In-hospital mortality 30 days	320 (23.48%)	228 (24.65%)	92 (21%)
ICU admission	646 (47.40%)	487 (52.65%)	159 (36.3%)
ICU 7-day admission[b]	410/1137 (36.06%)	272/719 (37.83%)	138/418 (33.01%)

Abbreviations: HIV, human immunodeficiency virus; ICU, intensive care unit; IQR, interquartile range; RT-qPCR, real-time quantitative RT-PCR; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

[a] Measured on admission to Emergency Department.

[b] Excluding those already on mechanical ventilation or vasoactive drugs before hospital admission.

Scores distribution

Scores distributions are shown in the Supplementary material (Fig. S1 and Table S2). Few patients crossed the threshold for qSOFA ≥2 (19.0%, 259/1363), whereas a large proportion had a 4C ≥ 4 (93.1%, 1269/1363), CALL ≥6 (92.7%, 1264/1363), CURB ≥1 (77.8%, 1061/1363), SCAP ≥10 (77.1%, 1051/1363) and SMART-COP ≥3 (72.5%, 989/1363). Intermediate proportions were observed for PSI ≥4 (59.9%, 817/1363), COVID-GRAM ≥0.1 (69.1%, 943/1363), REA-ICU ≥7 (62.0%, 845/1363), CURB-65 ≥ 2 (56.0%, 764/1363) and IDSA/ATS minor ≥3 (55.3%, 754/1363). The São Paulo cohort had more patients above the cut-offs for all scores. For most prediction assessment tools, a point increase was followed by an increase in the observed mortality rate (Fig. 1).

30-Day in-hospital mortality performance and clinical utility

Overall performance is shown in Table 2. PSI had the best AUC (0.79, 95% CI 0.77–0.82) followed closely by 4C (0.78, 95% CI 0.75–0.81), COVID-GRAM (0.77, 95% CI 0.75–0.80) and CURB-65 (0.74 95% CI 0.72–0.77). The 4C score had the lowest Brier score (0.146) followed by COVID GRAM (0.147) and PSI (0.148). PSI, 4C and COVID GRAM had both the best AUCs and Brier scores for our sample. We observed small departures in the predictive performance when analysing only RT-qPCR-confirmed cases ($n = 1053$, see Supplementary material, Tables S3 and S4). Analysis by country (see Supplementary material, Tables S5 and S6) shows PSI with the highest AUC in both; however, most scores had better performance in Barcelona compared with São Paulo. Overall, calibration was

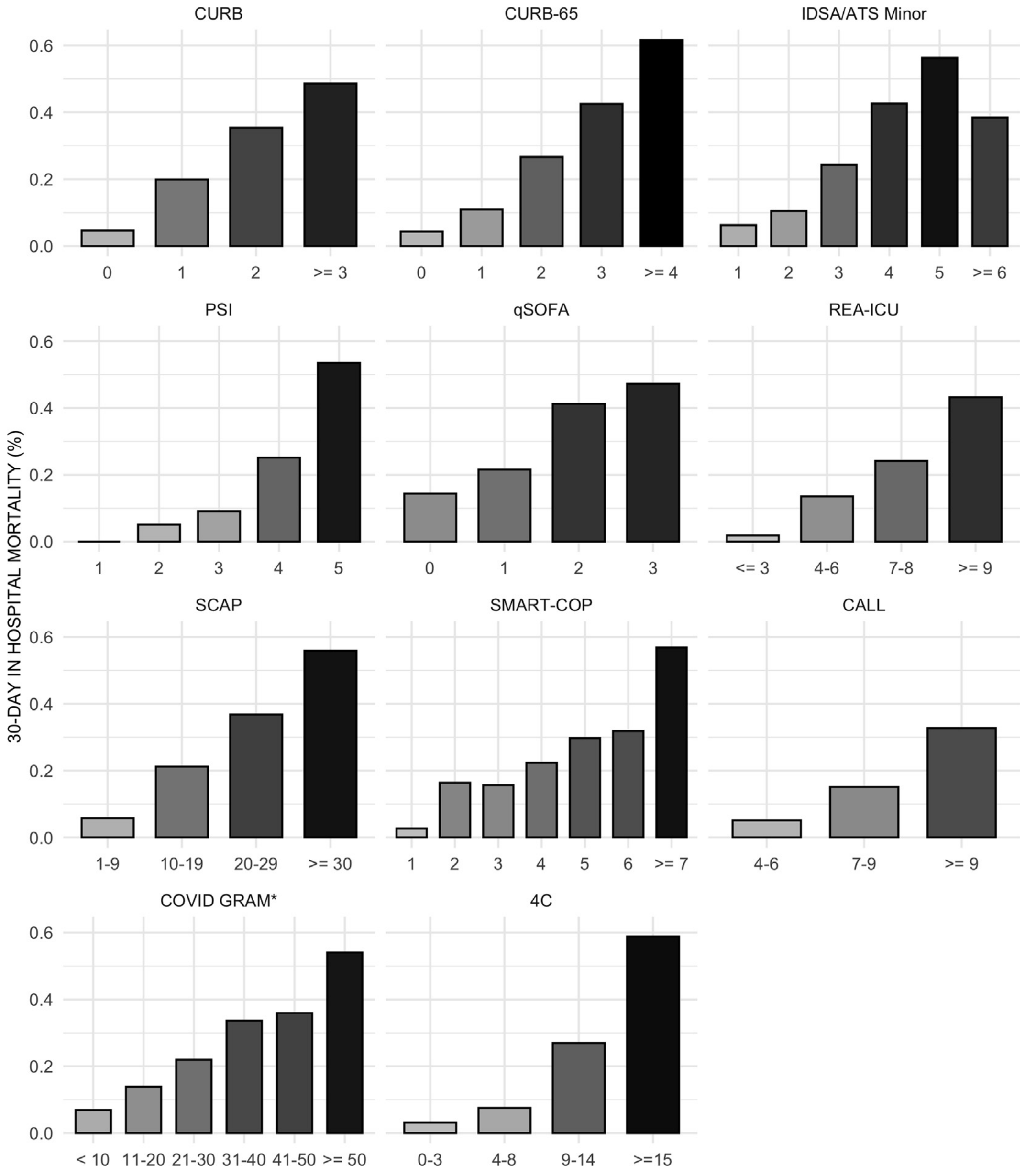

Fig. 1. Mortality rates by pneumonia risk assessment tools stratification at admission. Asterisk marks values in the x-axis described as percentages.

Table 2
Performance and clinical utility measurements for 30-day in-hospital mortality and 7-day intensive care unit admission

Score	AUC (95% CI)	Brier score	Threshold (≥)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	LR+ (95% CI)	NPV (95% CI)	LR−(95% CI)
30-day in-hospital mortality									
CURB	0.71 (0.68–0.74)	0.162	1	0.96 (0.93–0.98)	0.28 (0.25–0.3)	0.29 (0.26–0.32)	1.32 (1.27–1.39)	0.95 (0.93–0.98)	0.16 (0.08–0.25)
CURB65	0.74 (0.72–0.77)	0.154	2	0.84 (0.8–0.88)	0.53 (0.5–0.56)	0.35 (0.32–0.39)	1.78 (1.65–1.93)	0.92 (0.89–0.94)	0.3 (0.22–0.38)
QSOFA	0.63 (0.6–0.66)	0.171	2	0.34 (0.29–0.39)	0.86 (0.83–0.88)	0.42 (0.36–0.48)	2.37 (1.91–2.96)	0.81 (0.78–0.83)	0.77 (0.7–0.83)
PSI	0.79 (0.77–0.82)	0.148	4	0.9 (0.86–0.93)	0.49 (0.46–0.52)	0.35 (0.32–0.38)	1.76 (1.65–1.89)	0.94 (0.92–0.96)	0.21 (0.14–0.28)
SMART COP	0.71 (0.68–0.74)	0.161	3	0.89 (0.85–0.92)	0.32 (0.3–0.35)	0.29 (0.26–0.32)	1.31 (1.24–1.39)	0.9 (0.87–0.93)	0.35 (0.24–0.46)
IDSA/ATS Minor	0.73 (0.7–0.76)	0.157	3	0.82 (0.78–0.86)	0.53 (0.5–0.56)	0.35 (0.31–0.38)	1.76 (1.62–1.9)	0.91 (0.88–0.93)	0.33 (0.25–0.41)
REA-ICU	0.69 (0.65–0.72)	0.166	7	0.8 (0.76–0.85)	0.44 (0.4–0.47)	0.3 (0.27–0.33)	1.42 (1.31–1.53)	0.88 (0.85–0.9)	0.46 (0.36–0.57)
SCAP	0.74 (0.71–0.77)	0.159	10	0.94 (0.92–0.97)	0.28 (0.25–0.31)	0.29 (0.26–0.31)	1.31 (1.26–1.38)	0.94 (0.91–0.97)	0.2 (0.11–0.31)
COVID GRAM	0.77 (0.75–0.8)	0.147	0.1	0.91 (0.88–0.94)	0.37 (0.35–0.4)	0.31 (0.28–0.34)	1.45 (1.37–1.55)	0.93 (0.91–0.95)	0.24 (0.16–0.34)
CALL	0.71 (0.68–0.74)	0.162	6	0.99 (0.98–1)	0.09 (0.08–0.11)	0.25 (0.23–0.27)	1.09 (1.07–1.12)	0.97 (0.93–1)	0.1 (0–0.23)
4C	0.78 (0.75–0.81)	0.146	4	0.99 (0.98–1)	0.09 (0.07–0.1)	0.25 (0.23–0.27)	1.09 (1.06–1.11)	0.97 (0.93–1)	0.11 (0–0.27)
7-day intensive care unit admission									
CURB	0.59 (0.55–0.62)	0.226	1	0.8 (0.76–0.84)	0.29 (0.26–0.32)	0.39 (0.36–0.42)	1.13 (1.05–1.2)	0.72 (0.67–0.78)	0.69 (0.54–0.86)
CURB65	0.54 (0.51–0.58)	0.229	2	0.55 (0.49–0.6)	0.52 (0.48–0.55)	0.39 (0.35–0.43)	1.13 (1–1.27)	0.67 (0.63–0.71)	0.88 (0.77–1)
QSOFA	0.59 (0.56–0.62)	0.225	2	0.13 (0.1–0.16)	0.9 (0.88–0.92)	0.43 (0.34–0.52)	1.32 (0.92–1.83)	0.65 (0.62–0.68)	0.96 (0.92–1.01)
PSI	0.52 (0.49–0.56)	0.230	4	0.58 (0.53–0.62)	0.47 (0.44–0.51)	0.38 (0.34–0.42)	1.09 (0.98–1.21)	0.66 (0.62–0.7)	0.9 (0.77–1.02)
SMART COP	0.64 (0.61–0.67)	0.218	3	0.8 (0.77–0.84)	0.39 (0.36–0.43)	0.43 (0.39–0.46)	1.33 (1.23–1.44)	0.78 (0.74–0.83)	0.5 (0.39–0.61)
IDSA/ATS Minor	0.6 (0.57–0.64)	0.224	3	0.6 (0.55–0.65)	0.57 (0.54–0.61)	0.44 (0.4–0.48)	1.41 (1.25–1.57)	0.72 (0.68–0.75)	0.7 (0.6–0.8)
REA-ICU	0.6 (0.56–0.63)	0.224	7	0.65 (0.6–0.7)	0.49 (0.45–0.52)	0.42 (0.38–0.46)	1.27 (1.14–1.39)	0.71 (0.67–0.75)	0.72 (0.62–0.84)
SCAP	0.6 (0.57–0.63)	0.225	10	0.81 (0.77–0.85)	0.32 (0.28–0.35)	0.4 (0.37–0.44)	1.19 (1.11–1.27)	0.75 (0.7–0.8)	0.59 (0.46–0.73)
COVID GRAM	0.52 (0.48–0.55)	0.231	0.1	0.66 (0.62–0.71)	0.36 (0.33–0.4)	0.37 (0.34–0.4)	1.04 (0.95–1.14)	0.66 (0.61–0.7)	0.93 (0.78–1.1)
CALL	0.52 (0.49–0.56)	0.230	6	0.95 (0.92–0.97)	0.09 (0.07–0.11)	0.37 (0.34–0.4)	1.04 (1.01–1.08)	0.75 (0.65–0.83)	0.6 (0.36–0.91)
4C	0.55 (0.52–0.59)	0.229	4	0.95 (0.93–0.97)	0.1 (0.08–0.12)	0.37 (0.34–0.4)	1.05 (1.02–1.09)	0.77 (0.69–0.86)	0.52 (0.31–0.78)

Abbreviations: AUC, area under the curve; LR+, positive likelihood ratio; LR−, negative likelihood ratio; NA, not applicable; NPV, negative predictive value; PPV, positive predictive value.

good (see Supplementary material, Figs S2 and S3). Higher sensitivities were found for 4C, CALL and CURB; and higher specificities for qSOFA and CURB-65.

7-Day ICU admission performance and clinical utility

All scores had poor AUC (Table 2) with SMART COP (0.64, 95% CI 0.61–0.67), REA-ICU (0.60, 95% CI 0.56–0.63) and SCAP (0.60, 95% CI 0.57–0.63) having the highest values. Although there were performance differences between cohorts, SMART-COP had the best AUC in both (see Supplementary material, Tables S7 and S8). 4C and CALL had higher sensitivities, and qSOFA had the highest specificity.

Net benefit

Fig. 2 shows the decision curve analysis for in-hospital mortality. PSI had the best net benefit for most tested thresholds (1%–20%), followed by the 4C score. At a probability threshold of 5% (number-willing-to-treat of 20), PSI is the best strategy as it would avoid 6.2 interventions per 100 screened patients (Table 3). As the probability threshold increases, the best strategies change: at a 10% threshold (number-willing-to-treat 10), PSI, 4C and CURB would avoid 15.8, 14.6 and 11.9 interventions (per 100 patients) respectively; at a 20% threshold, SCAP would avoid 28.3, 4C 28.0 and PSI 27.9 interventions.

Discussion

We observed that the predictive performance of classical CAP severity scores was comparable with that of recently developed COVID-19 scores in 1363 hospitalized patients with COVID-19 pneumonia in Brazil and Spain. PSI and the recent 4C score had comparable performances in all evaluations. Among the tested scores, results were consistent for both cohorts, which are expected to have significant unmeasured differences regarding treatment and risk factors to poor outcomes because of socio-economic discrepancies in the underlying population (upper-middle-income country versus a high-income country).

PSI had the highest performance compared with other prediction rules regardless of the country origin and our results are comparable with those found in similar pandemic scenarios [9,10]. A possible explanation is that PSI heavily weights on comorbidities and age, which are known to be strong independent mortality risk factors for COVID-19 [27]. The same reasoning applies to the 4C score. On both, a 71-year old man is classified as intermediate risk based solely on age and gender regardless of any other information. The methodological rigour during PSI development, which included a large sample size, helped to build a robust model that was vastly validated in the literature [8,28,29].

The use of risk stratification scores in clinical practice requires analysis of the decision curve. PSI would be the best strategy in our cohort for threshold probabilities ≤5%. However, such a low threshold would only be reasonable in a scenario with little risk of overcrowding; a context that does not apply to many countries during this pandemic. Moreover, because there is currently no specific treatment for COVID-19, higher intensity of care in patients not at high risk of death may increase nosocomial infections and other related-complications without necessarily decreasing mortality. For thresholds between 6% and 20%, PSI and 4C had the highest net-benefit throughout the range. Although hospitalization is often unavoidable (e.g. need for oxygen therapy) even with low predicted mortality risk, these instruments can help manage the limited resources of

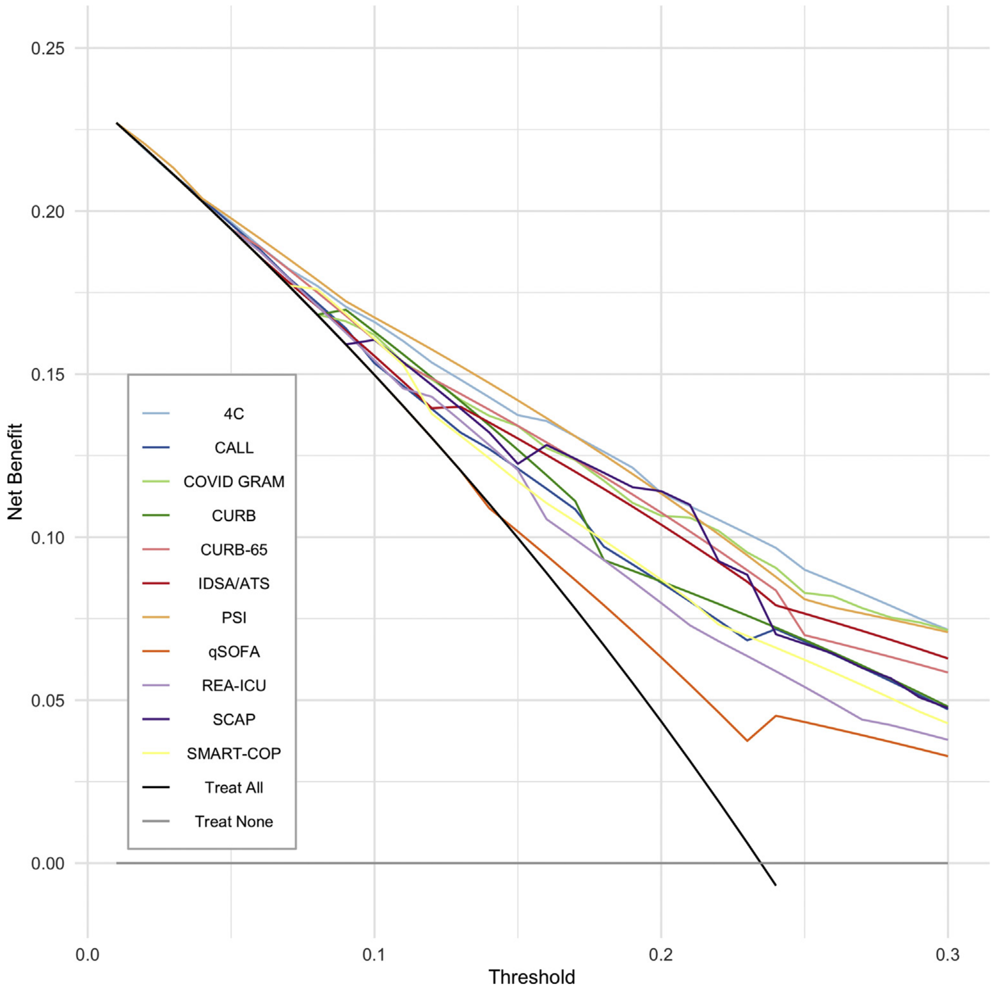

Fig. 2. Decision curve analysis.

Table 3
Number of avoidable interventions for different probability thresholds and scores

Score	Probability threshold		
	≥5%	≥10%	≥20%
CURB	0.00	11.89	17.17
CURB-65	0.00	9.76	25.61
qSOFA	0.00	0.00	7.85
PSI	6.24	15.85	27.95
SMART-COP	0.00	9.90	17.39
IDSA/ATS Minor	0.00	5.21	24.14
REA-ICU	0.07	4.04	14.53
SCAP	0.00	9.68	28.25
COVID GRAM	0.00	11.08	25.24
CALL	2.86	3.23	17.02
4C	4.04	14.60	28.03

The number of avoidable interventions (per 100 patients) for each score and probability threshold is shown.

hospitals by suggesting referral to higher- or lower-complexity facilities.

Deciding which assessment tool to apply involves not only the decision curve, but previous validations, generalizability, test availability and estimation complexity. The higher number of required variables for PSI calculation can make it time-consuming and unrealistic in under-resourced or overwhelmed scenarios. By contrast, qSOFA, a simpler tool that relies only on three clinical variables and consequently is widely applicable, had poor overall performance and an unexpected low sensitivity and high specificity for a screening tool—findings in line with similar studies in CAP [16]. Nevertheless, qSOFA had the highest positive predictive value, which placed it as a risk-stratification tool to be further evaluated. Alternatives with reasonable performance, lower number of required tests and easy estimation potentially applicable to low-resource settings are the CURB-65 (mixed clinical variables and urea) and the 4C score (mixed clinical variables, urea and C-reactive protein). CURB-65 has some advantages over 4C as it was extensively validated in different scenarios [8] and is already part of routine risk assessment for CAP in many emergency departments.

Remarkably, none of the evaluated scores performed well for 7-day ICU admission. Scores that were developed aimed at this particular outcome such as SMART-COP, SCAP and REA-ICU presented better overall performance in our cohort. SCAP and SMART-COP had the highest sensitivities among the three (81% and 80%, respectively) and better ability to exclude the outcome when negative (negative predictive value ≥ 75%). However, most patients were over the threshold at admission and therefore still on reasonable risk of ICU admission: 73 and 77 out of 100 admitted patients had SMART-COP ≥3 and SCAP ≥10, respectively. Although both 4C ≥ 4 and CALL ≥6 had high sensitivities and negative predictive value, they included over 92% of admitted patients, making them less useful. One possible explanation for the underperformance of CAP scores is that they rely on image findings (unilateral, multi-lobar or bilateral) that are known to affect prognosis in CAP [30] but are still unknown in COVID-19. Overall, these results coincide with those found for mortality in that new scores for COVID-19 had similar performance compared with CAP scores.

Our study has limitations. First, our results may not apply to secondary or primary settings as both medical centres were tertiary. Second, because we did not evaluate outpatients, the current external validation cannot support these scores for COVID-19 triage and so future studies are needed to clarify their clinical applicability in this setting. Additionally, our study comprises the initial pandemic surge in both countries, subject to the learning curve of COVID-19 treatment and the high demand for the health system, limiting our conclusions to similar scenarios. Third, we included patients with clinical COVID-19 diagnosis in Brazil because of RT-qPCR shortage during the early pandemic. However, our sensitivity analysis including only RT-qPCR showed comparable predictive performance. Finally, it is a challenge to apply risk stratification tools in tertiary referenced settings as previous treatments may lead to underscoring at admission (e.g. use of anti-pyretic medications and temperature at admission). Despite these limitations, the present study provides a validation of several scores already described for CAP that could help physicians to address patient safety and manage hospital resources. Among its strengths, our study has shown consistent validation results for cohorts from two countries with distinct socio-economic, ethnic and demographic backgrounds.

In summary, the performance of standard CAP risk assessment scores was comparable to three recently developed COVID-19 mortality risk stratification tools. It is expected that new scores will outperform older ones during development because they are often trained and tested in similar data sets. Therefore, more external validation studies are needed to ensure generalizability before recommending their use.

Availability of data and materials

The data sets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Funding

Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (#2016-14566-4 and #2020/04738-8); OTR is funded by a Sara Borrell grant from the Instituto de Salud Carlos III (CD19/00110). OTR acknowledges support from the Spanish Ministry of Science and Innovation through the Centro de Excelencia Severo Ochoa 2019–2023 Programme and from the Generalitat de Catalunya through the CERCA Programme.

Transparency declaration

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgements

COVID Registry Team:

Medical Students: Alexandre Lemos Bortolotto, Alicia Dudy Müller Veiga, Arthur Petrillo Bellintani, Beatriz Larios Fantinatti, Bianca Ruiz Nicolao, Bruna Tolentino Caldeira, Carlos Eduardo Umehara Juck, Cauê Gasparatto Bueno, Diego Juniti Takamune, Diogo Visconti Guidotte, Edwin Albert D'Souza, Emily Cristine Oliveira Silva, Erika Thiemy Brito Miyaguchi, Ester Minã Gomes da Silva, Everton Luis Santos Moreira, Fernanda Máximo Fonseca e Silva, Gabriel de Paula Maroni Escudeiro, Gabriel Travessini, Giovanna Babikian Costa, Henrique Tibucheski dos Santos, Isabela Harumi Omori, João Martelleto Baptista, João Pedro Afonso Nascimento, Laura de Góes Campos, Ligia Trombetta Lima, Luiza Boscolo, Manuela Cristina Adsuara Pandolfi, Marcelo de Oliveira Silva, Marcelo Petrof Sanches, Maria Clara Saad Menezes, Mariana Mendes Gonçalves Cimatti De Calasans, Matheus Finardi Lima de Faria, Nilo Arthur Bezerra Martins, Patricia Albuquerque de Moura, Pedro Antonio Araújo Simões, Rafael Berenguer Luna, Renata Kan Nishiaka, Rodrigo Cezar Miléo, Rodrigo de Souza Abreu, Rodrigo Werner Toccoli, Tales Cabral Monsalvarga, Vitor Macedo Brito Medeiros, Yasmine Souza Filippo Fernandes.

Residents: Ademar Lima Simões MD, Andrew Araujo Tavares MD, Clara Carvalho de Alves Pereira MD, Daniel Rodrigues Ribeiro MD, Danilo Dias de Francesco MD, Debora Lopes Emerenciano MD, Eduardo Mariani Pires de Campos MD, Felipe Liger Moreira MD, Felipe Mouzo Bortoleto MD, Gabriel Martinez MD, Geovane Wiebelling da Silva MD, Gustavo Biz Martins MD, Julio Cesar Leite Fortes MD, Lucas Gonçalves Dias Barreto MD, Maria Lorraine Silva de Rosa MD, Mauricio Ursoline do Nascimento MD, Rafael Faria Pisciolaro MD, Rodolfo Affonso Xavier MD, Stefany Franhan Barbosa de Souza MD, Thiago Areas Lisboa Netto MD.

Attending physicians: Sabrina Ribeiro MD, PhD, Carine Faria MD, Hassam Rahhal MD, Eduardo Padrão MD, Fernando Valente MD, Yago Henrique Padovan Chio MD, Luz Marina Gomez Gomez PhD.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cmi.2021.03.002.

References

[1] Dong E, Hongru D, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020;20:533–4. https://doi.org/10.1016/S1473-3099(20)30120-1.

[2] Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese center for disease control and prevention. JAMA 2020;323:1239–42.

[3] Prediction models for diagnosis and prognosis in Covid-19 [editorial] BMJ 2020;369:m1464.

[4] Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020;369:m1328.

[5] Ross JS, Normand SL, Wang Y, Ko DT, Chen J, Drye EE, et al. Hospital volume and 30-day mortality for three common medical conditions. N Engl J Med 2010;362:1110–8.

[6] Waterer GW, Self WH, Courtney DM, Grijalva CG, Balk RA, Girard TD, et al. In-hospital deaths among adults with community-acquired pneumonia. Chest 2018;154:628–35.

[7] Metlay JP, Waterer GW, Long AC, Anzueto A, Brozek J, Crothers K, 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:e45–67.

[8] Chalmers JD, Singanayagam A, Akram AR, Mandal P, Short PM, Choudhury G, et al. Severity assessment tools for predicting mortality in hospitalised patients with community-acquired pneumonia. Systematic review and meta-analysis. Thorax 2010;65:878–83.

[9] Fan G, Tu C, Zhou F, Liu Z, Wang Y, Song B, et al. Comparison of severity scores for COVID-19 patients with pneumonia: a retrospective study [letter]. Eur Respir J 2020;56.

[10] Satici C, Demirkol MA, Sargin Altunok E, Gursoy B, Alkan M, Kamat S, et al. Performance of pneumonia severity index and CURB-65 in predicting 30-day mortality in patients with COVID-19. Int J Infect Dis 2020;98:84–9.

[11] Gupta RK, Marks M, Samuels THA, Luintel A, Rampling T, Chowdhury H, et al. Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: an observational cohort study. Eur Respir J 2020.

[12] Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. N Engl J Med 1997;336:243–50.

[13] Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax 2003;58:377–82.

[14] Phua J, See KC, Chan YH, Widjaja LS, Aung NW, Ngerng WJ, et al. Validation and clinical implications of the IDSA/ATS minor criteria for severe community-acquired pneumonia. Thorax 2009;64:598–603.
[15] Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of clinical criteria for Sepsis: for the third international consensus definitions for Sepsis and septic shock (Sepsis-3). JAMA 2016;315:762–74.
[16] Ranzani OT, Prina E, Menéndez R, Ceccato A, Cilloniz C, Méndez R, et al. New Sepsis definition (Sepsis-3) and community-acquired pneumonia mortality. A validation and clinical decision-making study. Am J Respir Crit Care Med 2017;196:1287–97.
[17] España PP, Capelastegui A, Gorordo I, Esteban C, Oribe M, Ortega M, et al. Development and validation of a clinical prediction rule for severe community-acquired pneumonia. Am J Respir Crit Care Med 2006;174:1249–56.
[18] Charles PG, Wolfe R, Whitby M, Fine MJ, Fuller AJ, Stirling R, et al. SMART-COP: a tool for predicting the need for intensive respiratory or vasopressor support in community-acquired pneumonia. Clin Infect Dis 2008;47:375–84.
[19] Renaud B, Labarère J, Coma E, Santin A, Hayon J, Gurgui M, et al. Risk stratification of early admission to the intensive care unit of patients with no major criteria of severe community-acquired pneumonia: development of an international prediction rule. Crit Care 2009;13:R54.
[20] Liang W, Liang H, Ou L, Chen B, Chen A, Li C, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. JAMA Intern Med 2020;180:1081–9.
[21] Ji D, Zhang D, Xu J, Chen Z, Yang T, Zhao P, et al. Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score. Clin Infect Dis 2020;71.
[22] Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ 2020;370:m3339.
[23] Steyerberg EW. Clinical prediction models. Heidelberg: Springer; 2019. p. 558.
[24] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–74.
[25] Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res 2019;3:18.
[26] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55–63.
[27] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet 2020;395:1054–62.
[28] Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368.
[29] Aujesky D, Fine MJ. The pneumonia severity index: a decade after the initial derivation and validation. Clin Infect Dis 2008;47:S133–9.
[30] Liapikou A, Cillóniz C, Gabarrús A, Amaro R, De la Bellacasa JP, Mensa J, et al. Multilobar bilateral and unilateral chest radiograph involvement: implications for prognosis in hospitalised community-acquired pneumonia. Eur Resp J 2016;48:257–61.