Performance of prognostic scoring systems in patients with COVID-19 interstitial pneumonia in Acute Medical Unit

Silvia Accordino¹, Massimo Cazzaniga², Fabiola Sozzi¹, Roberto Assandri³, Giuseppe Lauria⁴, and Ciro Canetta¹

¹IRCCS Foundation Maggiore Policlinico Hospital
²ASST di Lecco
³Maggiore Hospital Crema
⁴Fondazione IRCCS Istituto Neurologico Carlo Besta

June 25, 2021

Abstract

Background: The early detection of COVID-19 patients with interstitial pneumonia at high risk of dismal outcome is necessary to deliver proper care and optimize management of limited resources. Objective: The aim of this study was to analyse the performance of pre-existing scores in predicting in-hospital mortality and ICU transfer at admission in an Acute Medical Unit. Methods: 106 consecutive patients with acute respiratory failure due to COVID-19 interstitial pneumonia admitted to Acute Medical Unit were enrolled. The performances of NEWS, SIRS, RAPS, REMS, qSOFA, APACHE II, CURB-65 and PSI were analysed by the Area Under the Receiver Operator Characteristic (AUROCs) and standard indices of accuracy. Results: Considering in-hospital mortality PSI and APACHE II had the higher AUROCs, 0.83 (95% CI 0.75-0.91) and 0.80 (95% CI 0.71-0.86), followed by REMS, 0.77 (95% CI 0.67-0.86), and CURB-65, 0.73 (95% CI 0.63-0.82), whereas the AUROCs of the other scores were < 0.7. PSI and APACHE II had good sensitivity (0.92 and 0.97), negative predictive value (0.96 and 0.97) and negative likelihood ratio (0.1 and 0.1), accurately identifying patients at low risk to die. However, the low specificity (0.70 and 0.47) and positive likelihood ratio (3.1 and 1.8) could limit their usefulness in predicting in-hospital mortality. Considering ICU admissions all the scores, except NEWS, SIRS and qSOFA, showed a worse performance. Conclusions: PSI and APACHE II showed good prognostic results in predicting in-hospital mortality but no pre-existing score validated for acute care settings was totally satisfactory to predict adverse outcomes in COVID-19 interstitial pneumonia after admission to Acute Medical Unit. The application setting and selected outcome criteria should always be considered to evaluate and compare scoring systems' performance analysis.

Title

Performance of prognostic scoring systems in patients with COVID-19 interstitial pneumonia in Acute Medical Unit

Running Title

Prognostic scores in COVID-19 pneumonia

Keywords: COVID-19 interstitial pneumonia; acute medical unit; prognostic scoring systems; in-hospital mortality

Introduction

COVID-19 is a global ongoing pandemic that is overwhelming several national healthcare systems. Approximately 15% of patients develop severe interstitial pneumonia requiring oxygen support, and 5% have critical
disease [1]. During this last year, hospital overcrowding and increase of ICU’s bed occupancy rate up to saturation explain the great need for reorganization of hospitalized patients’ clinical governance. The close monitoring for worsening of respiratory status is critical to allow rapid initiation of optimized treatments and referral to a designated destination in the more adequate setting [2].

Different scoring systems have been previously specifically designed and validated as predictors of ICU admission or death, even if in COVID-19 patients it might be difficult to determine which one is most suitable for different settings in the acute care chain.

The aim of this study was to analyse and compare the performance of pre-existing scoring systems in predicting the risk of dismal outcome in a cohort of hospitalized COVID-19 patients in Acute Medical Unit (AMU), assuming in-hospital mortality as primary outcome while ICU transfer and a composite criterion of in-hospital mortality and ICU transfer were considered as secondary outcomes.

Methods

We investigated all 106 consecutive patients with acute respiratory failure due to COVID-19 interstitial pneumonia admitted, between 2020 Feb 21st and Apr 21st, to the AMU of the Maggiore Hospital of Crema (Lombardy, Italy). The local Institutional Ethics Committee approved the study and waived the requirement for informed consent from the study subjects due to the study design. The study complied with the Declaration of Helsinki. Accessed data were aggregated and anonymized.

We applied standard definitions to calculate eight scoring systems: National Early Warning Score (NEWS), Systemic Inflammatory Response Syndrome (SIRS), Rapid Physiology Score (RAPS), Rapid Emergency Medicine Score (REMS), quick Sequential Organ Failure Assessment (qSOFA), Acute Physiology and Chronic Health Evaluation II (APACHE II), Confusion, Urea nitrogen, Respiratory rate, Blood pressure and age ≥65 (CURB-65) and Pneumonia Severity Index (PSI). [Table 1]

Patients were divided in survivors and non-survivors and we analysed the accuracy of all the previous scores performing the Area Under the Receiver Operator Characteristic (AUROC). The optimal cut-off values identified by maximizing Youden’s index for each scoring system were determined and sensitivity, specificity, predictive values and likelihood ratios were assessed for the primary outcome.

Statistical Analysis

Descriptive variables are reported as mean ± standard deviation (min-max) for continuous variables and absolute numbers and percentages for categorical data. We compared data between survivors and non-survivors with non-parametric tests such as Mann–Whitney, chi-square test or Fisher exact test as appropriated. The performance of each score was estimated using the Area Under the Receiver Operator Characteristic (AUROC) analysis for in-hospital mortality and for secondary outcomes (ICU transfers and composite outcome of deaths and ICU transfers). The optimal cut-off values, identified by maximizing Youden’s index for each scoring system, were assessed by the sensitivity, specificity, predictive values (positive – PPV – and negative – NPV) and likelihood ratios (LH). Statistical analyses were performed with an SPSS package (SPSS 26 for Mac, SPSS Inc., Chicago, IL, USA).

Results

The overall population’s characteristics and comparison between survivors and non-survivors are shown in Table 2. At admission in AMU, 97 patients (91.5%) had a PaO₂/FiO₂ ratio < 300, 34 patients (32.1%) had a PaO₂/FiO₂ ratio < 100. Seventy-five patients (75.5%) needed to undergo non-invasive ventilatory support (Continuous Positive Airway Pressure - CPAP and/or NIPPV), eight patients (7.5%) were transferred to ICU for mechanical ventilation.
according to defined criteria (PaO$_2$/FiO$_2$ < 100 and Respiratory Rate > 30 breaths/min in non-responders to CPAP/NIPPV and without do-not-resuscitate order). The overall in-hospital mortality was 28.3%.

The AUROC analysis demonstrated the overall discrimination power of prognostic scores for primary and secondary outcomes [Table 3].

Only PSI and APACHE II showed good discrimination power in predicting in-hospital mortality, 0.83 (95% CI 0.75-0.91) and 0.80 (95% CI 0.71-0.88) respectively, followed by REMS, 0.77 (95% CI 0.67-0.86), and CURB-65, 0.73 (95% CI 0.63-0.82), while the other scores did not reach an acceptable value (0.70).

All the scores, except NEWS, SIRS and qSOFA, showed worse AUROCs considering ICU admissions, while all AUROCs did not significantly differ from in-hospital mortality ones considering the composite outcome.

To predict in-hospital mortality PSI and APACHE II showed good sensitivity (0.92 and 0.97), negative predictive value (0.96 and 0.97) and negative likelihood ratio (0.1 and 0.1) suggesting that they accurately identified COVID-19 patients at low risk to die. However, the low specificity and positive likelihood ratio (0.7, 3.1 for PSI and 0.47, 1.8 for APACHE II) limited their usefulness in predicting those at high in-hospital mortality risk. [Table 4]. Considering APACHE II and PSI a LH- of 0.1 in both scoring systems and assuming the pre-test probability of 28%, the post-test probability is reduced to about 3% according to Massachusetts’ Nomograms for probabilities and likelihood ratios.

**Discussion**

In COVID-19 patients multiple severity assessment tools, pneumonia-specific, generic sepsis or early warning scores have been used for risk stratification, however there is no evidence to suggest which tool should be recommended in designated settings to identify the high-risk ones.

Different studies have been conducted to analyse the performance of pre-existing scores in predicting dismal outcomes, as NEWS, NEWS 2, MEWS, REMS, qSOFA, SOFA and CURB-65 [3, 4, 5, 6, 7, 8, 9, 10], some of them showing good results while others enhanced that the pre-existing scoring systems underestimated the risks compared with pre-COVID-19 cohorts.

The principal outcomes considered were mortality and ICU transfers, however different ICU admission criteria, different definitions of respiratory deterioration, different settings of application (EDs, general inpatients wards, acute medical wards or ICUs) and different patients’ characteristics could explain this controversial. Moreover, the expansion of the selected outcome would apparently improve the prognostic power of any score, as well as the change of the cut-off of each score would increase the usability as a screening tool at cost of specificity. Caution must be taken in applying the results in patients admitted in critical care areas. The accuracy of systems has strong implications, both in terms of identifying more cases (ie, sensitivity) to minimize adverse events, and in preventing excessive alarms (ie, specificity) to minimize resource expenditure.

In time of limited healthcare resources, all standard indices of accuracy for each scoring system should be considered
to limit risk either of under-triage, with low rate of false negative predictions to avoid the admission of high-risk patients in non-critical areas, or of over-triage, with low rate of true positive predictions to avoid the admission of low-risk patients in critical areas.

The ideal score should be both simple and accurate, with few chances of calculation errors; it should have a specific defined outcome and setting of application in the acute care chain. Simple bedside systems as the Early Warning Scores (EWSs) are appealing due to their ease of use, although simplicity and accuracy can be hardly combined.

Thereby the identification of individual susceptibility to become critical upon infection is crucial to plan the escalation of treatment and guide decision-making and management. No prognostic score could faultlessly detect all patients at risk of adverse outcomes. The more helpful parameters in COVID-19 patients with interstitial pneumonia were physiological variables that define the severity of respiratory failure, associated with the identification of a phenotype of high-risk patients (i.e. older age, history of cardiovascular disease).

The development of life-threatening conditions is often difficult to be identified, since many COVID-19 patients do not present initial respiratory effort and impairment of vital signs normally included in EWSs, which are simple but maybe not sufficiently accurate in COVID-19 patients.

In AMUs patients are selected according to clinical instability and admitted after a prior medical assessment; the aim can’t be exclusively the interception of imminent deterioration (as in EDs) nor long term results (30-day mortality).

In our analysis the scores that appeared more accurate in predicting in-hospital mortality were the ones including physiological data as well as the age: REMS, CURB-65, APACHE II and PSI. PSI and APACHE II which consider also comorbidities reached the higher AUROCs for the primary outcome, but they are complex to calculate and difficult to use in busy hospital settings. ICU admission, instead, was not proper for all patients: the comprehensive evaluation of the age, comorbidities and functional status and not only of the higher scores’ values may explain the relative low rate of ICU transfers and may be a confounder in its use as secondary outcome.

The same considerations based on a global evaluation of each patient can be done analysing the in-hospital length of stay, longer in survivors (median 14.5, IQR 10-18, range 2-34) than in non-survivors (median 6.5, IQR 2.5-11.8, range 1-30): the large ranges were conditioned not only by critical conditions or clinical outcome, but also by familiar environment, frailty and social factors that have affected dischargeability.

The comparison of other current COVID-19 studies should consider many several selective biases in the cohorts of patients analysed, depending on different clinical characteristics, different levels of care wards and different health care systems reality in different countries.

These results are not intended to be conclusive, they include a population from a single centre in Italy with a modest number of patients during the first dramatic phase of the pandemic, but they can provide tips that can be useful in clinical decision-making, considering clinical risk. During second and third wave of the pandemic, the early recognition and escalation of treatment of deteriorating patients have allowed the improvement of the hospital macro-organization, to guarantee a better patients’ management, in the right place and time.

Conclusions

In Acute Medical Unit, among available severity and prognostic scoring systems, PSI and APACHE II showed good prognostic performance to predict in-hospital mortality, although none can be totally satisfactory considering overall adverse events. A specific scoring system tailored on COVID-19 patients characteristics is warranted to
optimize and standardize patients’ care and management of available resources. The application setting and selected outcome criteria should always be considered to evaluate and compare scoring systems’ performance analysis.

References


Hosted file