Intensive Care Med (2021) 47:1472–1474
https://doi.org/10.1007/s00134-021-06539-3

EDITORIAL

Assess COVID-19 prognosis … but be aware of your instrument's accuracy!

Maurizia Capuzzo[1*], Andre Carlos Kajdacsy-Balla Amaral[2] and Vincent X. Liu[3]

The World Health Organization declared the novel coronavirus disease 2019 (COVID-19) outbreak a global pandemic on March 11, 2020. The large number of patients needing hospital care led to shortages of beds in intensive care units (ICUs), requiring an increase in ward and ICU capacity, which continues to stress healthcare systems [1]. The imbalance between supply and demand for medical resources has raised concerns about how scarce medical resources can be optimally allocated during the COVID-19 pandemic. Rationing ICU beds is a potential strategy to cope with limited resources [2], with prognostic scores potentially helping in triage decisions, such as identifying the optimal ward for patients (e.g., step-up or step-down units), considering transfer to other hospitals with available capacity or, in some cases, discussing limitations of care [3]. These scores can also be useful in stratifying patients by mortality risk for clinical research aimed at identifying effective treatments and protocols.

Nevertheless, physicians need to be confident that the prognostic instrument they use is accurate: that the value it predicts is close to the true value of the relevant outcome. Independent external validation of previously developed models, which evaluates score performance in a similar population drawn from a different cohort, is an essential way to guarantee the accuracy and the generalizability of prognostic instruments [4, 5]. Uncertain confidence in prognostic models may be particularly acute for COVID-19, as a recent systematic review of 31 prediction models for COVID-19 concluded that most published models have been poorly reported and were at high risk of bias [6]. Similar uncertainty regarding model performance when tested in external or independent samples has been described for other prediction models recently [7]. A cursory PubMed search using "prediction" and some ICU relevant conditions over a period of 20 months shows a plethora of publications (Fig. 1). Even if we consider some overlapping between different conditions, the large number of articles on prognostic models and COVID-19 ($n=1231$) highlights the challenges that clinicians have faced in finding useful information on existing predictive models in this pandemic.

In this issue of Intensive Care Medicine, Lombardi et al. [8] evaluate the performance of existing prognostic scores to predict in-hospital mortality and the composite outcome of in-hospital mortality or ICU transfer in SARS-CoV-2-infected patients. The external validation was performed in a large database (14,343 patients) containing data collected in 39 hospitals spread across Paris and its surrounding region. The authors selected 32 prognostic scores, which met pre-specified criteria for "higher quality" and were computable with their data. When assessing the performance of each score to predict the outcome most similar to the one used in the original study, only the 4C Mortality score [9] had an Area Under the receiver operating curve (AUC) significantly higher than that previously published. When assessing the performance of each score to predict 30-day in-hospital mortality, seven scores showed at least very good performance (AUC > 0.75); the 4C Mortality [9] (AUC 0.793, 95% CI 0.783–0.803) and the ABCS [10] (0.79, 95% CI 0.78–0.801) scores exhibited the highest AUCs. Using the 4C Mortality score at a low-risk cutpoint of 3 to predict in-hospital mortality, the sensitivity (true positive rate) was 99% and specificity (true negative rate) was 8%. At a high-risk cutpoint of 15, sensitivity was 21%, and specificity was 96% (Table S8). The CORONATION-TR score [11] had the highest AUC (0.724, 95% CI 0.714–0.733) to predict the composite outcome.

*Correspondence: cpm@unife.it
[1] Department of Translational Medicine, Intensive Care Section, University of Ferrara, Ferrara, Italy
Full author information is available at the end of the article

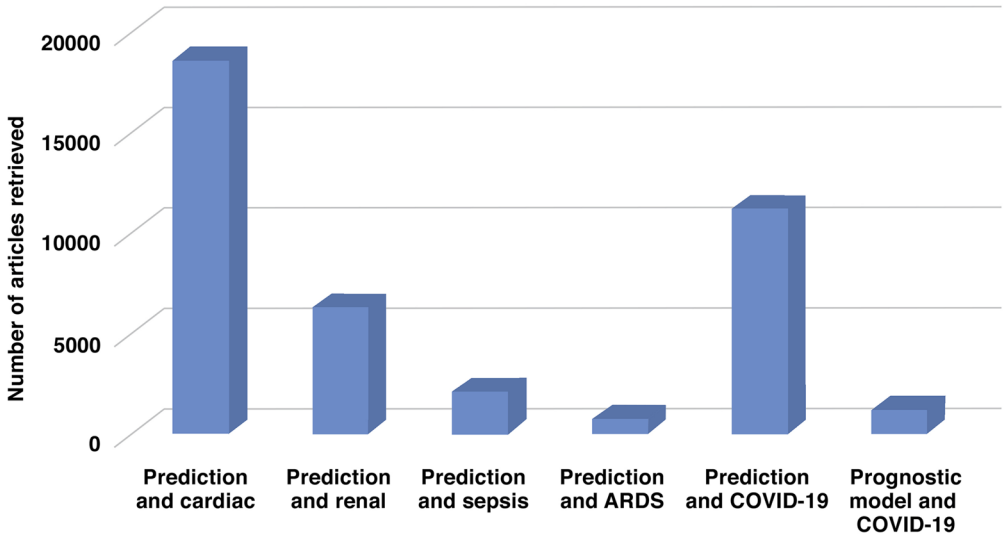

Fig. 1 Number of articles published from January 2020 to August 2021 retrieved by search in PubMed, according to the key words used

The two scores with the highest AUC to predict 30-day in-hospital mortality (the 8-item 4C Mortality score [9] and the 10-item ABCS score [10]) include different weightings for age, sex, and C reactive protein. Of the remaining variables of 4C Mortality score, four are clinical and one is a laboratory test (urea). Of the remaining variables of ABCS, six are laboratory tests while only one is clinical (chronic obstructive pulmonary disease). Interestingly, while these scores differ in their balance between clinical and laboratory variables, both appeared to demonstrate similar performance in terms of discrimination and calibration. Moreover, both scores appear to be easy to use.

One of the major strengths of the study is that the external validation was performed using a large number of hospitals and patients, which strengthens the generalizability of the results. In addition, Lombardi et al. [8] performed multiple supporting analyses to enhance the utility of their findings. First, they assessed discrimination using data from different waves of the pandemic (no substantial difference) and across age subgroups (lower performance was seen in patients > 65 years old for some scores). Second, they assessed model performance using standard AUC values as well as area under the precision-recall curve (results unchanged) which can account for imbalanced outcomes. Finally, they assessed calibration using the calibration curve by deciles of risk, which allowed them to describe that the models overestimated mortality, especially during the epidemic waves subsequent to the first one.

Weaknesses of the study were that the authors were unable to exactly replicate all of the scores as they were originally designed because of missing data or other cohort differences. Nevertheless, they clearly describe their missing data and no substantial change was found in AUC values between datasets with multiple imputation or complete case analysis (Table 7S). Other limitations are the retrospective study design that could include selection and information bias, as reported in the discussion. In addition, prognostic scores are useful instruments to assess groups of patients, but may have limitations when applied in the clinical practice for individual risk prediction due to their probabilistic nature [12].

In conclusion, we strongly commend Lombardi et al. [8] for their carefully conducted and comprehensive study. Although prior reports suggest that existing COVID-19 predictive models are at high risk for bias, this study helps lend confidence to a key set of prognostic scores by rigorously evaluating their performance in an independent sample. In the future, this work might allow researchers to devote their limited time to verifying the actual utility of COVID-19 prognostic scores in the clinical setting, instead of attempting to create new prognostic scores.

Author details
[1] Department of Translational Medicine, Intensive Care Section, University of Ferrara, Ferrara, Italy. [2] Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, ON, Canada. [3] Kaiser Permanente Division of Research, Oakland, CA, USA.

Declarations

Conflict of interests
The authors declare that they have no conflicts of interest.

Publisher's Note

Received: 15 September 2021 Accepted: 17 September 2021
Published online: 5 October 2021

References

1. Arabi YM, Azoulay E, Al-Dorzi HM et al (2021) How the COVID-19 pandemic will change the future of critical care. Intensive Care Med 47:282–291. https://doi.org/10.1007/s00134-021-06352-y
2. Emanuel EJ, Govind Persad G, Upshur R et al (2020) Fair allocation of scarce medical resources in the time of COVID-19. N Engl J Med 382(21):2049–2055. https://doi.org/10.1056/NEJMsb2005114
3. Williams JM, Greenslade JH, Chu KH, Brown AF, Lipman J (2018) Utility of community-acquired pneumonia severity scores in guiding disposition from the emergency department: intensive care or short-stay unit? Emerg Med Australas 30:538–546. https://doi.org/10.1111/1742-6723.12947
4. Altman DG, Vergouwe Y, Royston P, Moons KGM (2009) Prognosis and prognostic research: validating a prognostic model. BMJ 338:b605. https://doi.org/10.1136/bmj.b605
5. Steyerberg EW, Harrell FE Jr (2016) Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 69:245–247. https://doi.org/10.1016/j.jclinepi.2015.04.005
6. Wynants L, Calster BV, Bonten MMJ et al (2020) Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ. https://doi.org/10.1136/bmj.m1328
7. Wong A, Otles E, Donnelly JP et al (2021) External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med 181:1065–1070. https://doi.org/10.1001/jamainternmed.2021.2626
8. Lombardi Y, Azoyan L, Szychowiak P et al. (2021) External validation of prognostic scores for Covid-19: a multicentre cohort study of patients hospitalized in Greater Paris University Hospitals. Intensive Care Med. https://doi.org/10.1007/s00134-021-06524-w
9. Knight SR, Ho A, Pius R et al (2020) 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 370:m3339. https://doi.org/10.1136/bmj.m3339
10. Jiang M, Li C, Zheng L et al (2021) A biomarker-based age, biomarkers, clinical history, sex (ABCS)-mortality risk score for patients with coronavirus disease 2019. Ann Transl Med 9:230. https://doi.org/10.21037/atm-20-6205
11. Tanboğa IH, Canpolat U, Çetin EHÖ et al (2021) Development and validation of clinical prediction model to estimate the probability of death in hospitalized patients with COVID-19: insights from a nationwide database. J Med Virol 93:3015–3022. https://doi.org/10.1002/jmv.26844
12. Teres D, Lemeshow S (1994) Why severity models should be used with caution. Crit Care Clin 10:93–110