ORIGINAL CLINICAL REPORT

OPEN

External Validation of Prognostic Models in Critical Care: A Cautionary Tale From COVID-19 Pneumonitis

Sebastian Bate, MMath[1,2]
Victoria Stokes, MBChB[3]
Hannah Greenlee, BMBCh[3]
Kwee Yen Goh, MBChB[3]
Graham Whiting, MBChB[3]
Gareth Kitchen, MBChB, PhD[3,4]
Glen P. Martin, PhD[5]
Alexander J. Parker, MBBS[3]
Anthony Wilson, MBBS[3,5]
on behalf of the MRI Critical Care Data Group

OBJECTIVES (BACKGROUND): To externally validate clinical prediction models that aim to predict progression to invasive ventilation or death on the ICU in patients admitted with confirmed COVID-19 pneumonitis.

DESIGN: Single-center retrospective external validation study.

DATA SOURCES: Routinely collected healthcare data in the ICU electronic patient record. Curated data recorded for each ICU admission for the purposes of the U.K. Intensive Care National Audit and Research Centre (ICNARC).

SETTING: The ICU at Manchester Royal Infirmary, Manchester, United Kingdom.

PATIENTS: Three hundred forty-nine patients admitted to ICU with confirmed COVID-19 Pneumonitis, older than 18 years, from March 1, 2020, to February 28, 2022. Three hundred two met the inclusion criteria for at least one model. Fifty-five of the 349 patients were admitted before the widespread adoption of dexamethasone for the treatment of severe COVID-19 (pre-dexamethasone patients).

OUTCOMES: Ability to be externally validated, discriminate, and calibrate.

METHODS: Articles meeting the inclusion criteria were identified, and those that gave sufficient details on predictors used and methods to generate predictions were tested in our cohort of patients, which matched the original publications' inclusion/exclusion criteria and endpoint.

RESULTS: Thirteen clinical prediction articles were identified. There was insufficient information available to validate models in five of the articles; a further three contained predictors that were not routinely measured in our ICU cohort and were not validated; three had performance that was substantially lower than previously published (range *C*-statistic = 0.483–0.605 in pre-dexamethasone patients and *C* = 0.494–0.564 among all patients). One model retained its discriminative ability in our cohort compared with previously published results (*C* = 0.672 and 0.686), and one retained performance among pre-dexamethasone patients but was poor in all patients (*C* = 0.793 and 0.596). One model could be calibrated but with poor performance.

CONCLUSIONS: Our findings, albeit from a single center, suggest that the published performance of COVID-19 prediction models may not be replicated when translated to other institutions. In light of this, we would encourage bedside intensivists to reflect on the role of clinical prediction models in their own clinical decision-making.

KEYWORDS: acute respiratory distress syndrome; clinical prediction modeling; COVID-19 pneumonitis; external validation; intensive care

The COVID-19 pandemic has created an unprecedented amount of research with over 300,000 articles published (1). Prognostic modeling of COVID-19 outcomes has received much attention but systematic review has found that many models suffer from poor methodology and are

DOI: 10.1097/CCE.0000000000001067

KEY POINTS

Question: Can we externally validate prognostic models for COVID-19 pneumonitis in ICU?

Findings: Incomplete reporting and use of predictors that are not routinely collected meant multiple models could not undergo validation. Those that did have sufficient information had a mixed performance in validation.

Meaning: Significant improvements in methodology and reporting are needed for intensive care prognosis models. Large-scale collaboration is recommended if models are to be reliable and widely adopted.

at high risk of bias (2). An individual participant data meta-analysis by de Jong et al (3) examining models for 30-day in-hospital mortality observed that there was substantial variation in the prognostic value of the models, even when models with a high risk of bias had been excluded.

For a prediction model to be adopted into practice, its predictions must be accurate, reliable, and explainable and derived from data that is routinely recorded at most institutions. However, many models are not externally validated or perform poorly when validated (3–5). As a consequence, they have limited clinical utility. We have previously externally validated a model for predicting continuous positive airway pressure failure in COVID-19 patients developed early in the pandemic and demonstrated how its performance fell markedly in patients in whom new standards of care had been adopted. This prompted further examination of available models (6).

The quality of the reporting of prognostic models is often poor despite the common requirement from journals to use the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines (2, 7–9) in model reporting. Key aspects of COVID-19 models are often not reported, including the parameters used and the coefficients of the model (2). A recent systematic review found 32 of 43 mortality prediction models in critical care did not provide sufficient detail to be replicated (10).

Many COVID-19 prognostic models were developed in the early stages of the pandemic (2). Since then, there have been significant advances in management of the disease, with an increasing role for immunomodulatory and anti-viral therapies and changing attitudes among physicians concerning the role of noninvasive ventilation in the management of COVID-19 pneumonitis (11–13). In addition, there have been changes in the immune status of the general population both through virus exposure and from widespread uptake of effective vaccination (14). We anticipated these changes may have impacted the performance of models that aim to predict survival for patients with COVID-19, particularly in an ICU setting. There have been external validations of COVID-19 prediction models, but there is yet to be one that specifically focuses on models developed in ICUs rather than general wards (3, 15–18).

This article addresses the following questions:

1] Are published ICU multivariable prediction models sufficiently described to allow new predictions to be generated, and performance compared with the original publication?
2] Do these models use parameters that are routinely recorded in most ICUs?
3] Is the performance (by discrimination and/or calibration) of these models replicable in patients from a similar period to when the model was developed, either pre- or post-the introduction of dexamethasone?
4] Is the performance (by discrimination and/or calibration) of these models replicable using patients for the duration of the pandemic?

METHODS

This work is reported in line with TRIPOD (8, 9) focusing on the methods and results components of the TRIPOD checklist. Prospective models for inclusion were identified by three methods (**Fig. 1**): articles identified from the "living systematic review" (fourth update) by Wynants et al (2); PubMed search was conducted; and further articles by reverse citation.

We included models where the study population used to develop the original model was adults (over 16 yr old) admitted to the ICU with a diagnosis of COVID pneumonitis confirmed by polymerase chain reaction test or confirmed clinically by a senior ICU doctor using a combination of radiographic results and symptoms; models that required at least two predictor variables, and which were measured within 24 hours of,

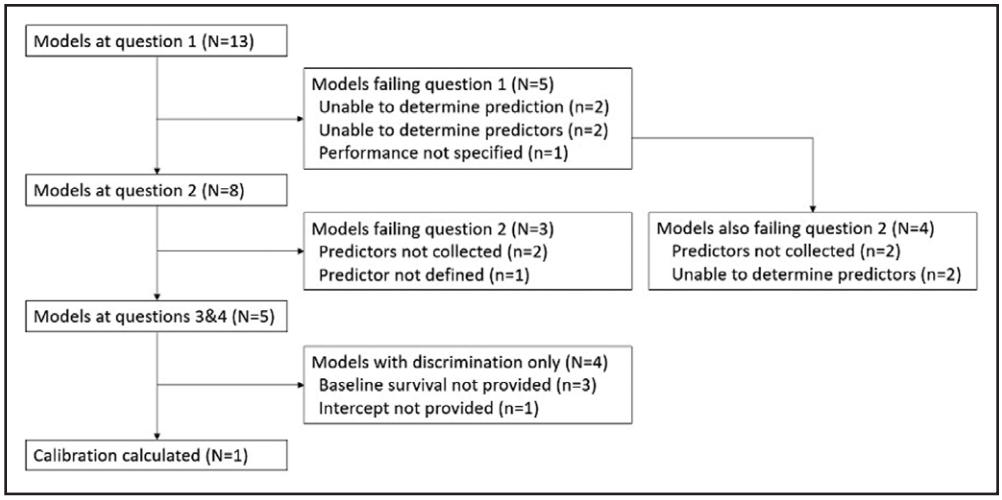

Figure 1. Studies included.

or before, ICU admission or 24 hours before onset of invasive mechanical ventilation (IMV), depending on the model, or demographic data. We included models where the endpoint was death at any point, need for IMV, or length of stay in ICU.

Articles underwent title and abstract screening and if required full-text screening with included articles agreed by consensus with the study team authors. The whole team extracted the model parameters as published. The clinical team reviewed the predictors listed within the articles and determined whether they were collected at our institution. Disagreements were solved by consensus. Models failing to meet the criteria for question 1 or question 2 (Fig. 1) were not evaluated for questions 3 and 4. The criteria for passing question 1 included models specifying all predictors and a method for generating predictions (or at least ranked predictions). The criteria for passing question 2 were that all predictors were collected on our ICU, or could be calculated post hoc, and sufficiently well-defined.

Data Source

Patient data were extracted for all patients admitted to the ICU at Manchester Royal Infirmary, Manchester, United Kingdom, from March 1, 2020, to February 28, 2022, with confirmed or strongly suspected COVID pneumonitis using the patient electronic record systems (EPRs) (Ethical approval 21/HRA/3518). Extraction techniques and general characteristics of COVID-19 patients from our ICU have been previously described in detail (19, 20). Patients, who in the opinion of the treating physician, had an incidental COVID-19 finding or who had expressed a wish to not take part in research via the National Health Service National Data Opt-Out were excluded. The sample size was pragmatic based on all patients available at our center.

The data for each model validation only included patients meeting the inclusion/exclusion criteria for the model and with a complete set of parameters. The characteristics of the patients who contributed data to each validation cohort are summarized in **Supplementary Tables 1** and **2** (http://links.lww.com/CCX/B325). For composite parameters not calculated within the EPR but for which all components were collected, values were calculated post hoc. Data were not imputed.

For models that used the Sequential Organ Failure Assessment (SOFA) score, this was determined using the worst value for each component in the first 24 hours of ICU stay unless specified otherwise (21). Where relevant, patient notes were consulted. A conversion factor of 0.1 was used to convert C-reactive protein (CRP) to high-sensitivity CRP (22). The "pre-dexamethasone era" is defined as patients admitted before June 16, 2020, and "post-dexamethasone era" otherwise. This date was chosen because it coincided with when the results from the RECOVERY trial into dexamethasone were published and practice changed almost overnight (23). For question 3, models trained on patients in the pre-dexamethasone era were tested using all patients in the dataset admitted in the pre-dexamethasone era and further matching for the study's inclusion and exclusion criteria (Supplementary Table 2, http://links.lww.com/CCX/B325). No blinding was performed as outcomes were retrospectively extracted.

Statistical Analysis

Where articles passed questions 1 and 2 and were able to be validated, predictions were generated for each patient in our dataset and the *C*-statistic (also known

as concordance or area under the curve of the receiver operator characteristic) calculated for the population (or for survival models, Harrell's *C*). Calibration curves were calculated where sufficient detail was provided. Where only ranked predictions could be determined (e.g., the intercept was missing but relative weights given), discrimination was calculated without calibration.

RESULTS

In total, 13 articles and 22 models were included (Fig. 1). **Table 1** shows the articles with their characteristics and whether they met the requirements of questions 1 and 2.

Among the five models that failed to pass question 1 (ability to be validated), Vaid et al (29) and Vaid et al (30) did not list the predictors used in their model; Gerotziafas et al (24) and Pan et al (27) did not report model coefficients.

Of the models that did not report their model coefficients, all four articles used machine learning algorithms to develop their models (Gerotziafas et al [24] used the t-distributed stochastic neighbor embedding algorithm, and the remainder used [XGBoost] [37, 38]). The models that used XGBoost had a very large number of candidate predictors, which were then shrunk as part of the model fitting. The articles using machine learning did report the variable importance but not the exact specification of the model, which prevents the articles from being validated.

The Instrumental Activities of Daily Living Scale 8 and biomarkers interleukin-6, soluble E-selectin, soluble platelet selectin, angiopoietin-2, soluble intercellular adhesion molecule-1, and von Willebrand factor are not routinely measured in our institution and chest tightness was neither well defined by the publication nor routinely measured in our institution so Falandry et al (33), Popadic et al (28), Vassiliou et al (31), and Wang et al (32) were not able to be validated.

Table 2 shows the characteristics of the patients included in our validation cohort. Three hundred forty-nine patients were included in the full dataset of which 47 did not meet the inclusion criteria for any of the four studies.

Table 3 shows the concordance of each of the models that were validated against the originally published outcome. Three of the models performed poorly relative to their original publication; one (Arina et al [25]) performed poorly after the introduction of dexamethasone, and one (Moisa et al [36]) remained similar but had lower discriminative ability at publication. Calibration was performed for Arina et al (25) and the plot is in **Supplementary Figure 1** (http://links.lww.com/CCX/B325). Cao et al (26) did not publish an intercept and Leoni et al (34), Leoni et al (35), and Moisa et al (36) did not publish baseline survival so calibration could not be performed.

Only Cao et al (26) provided a cutoff for high- and low-risk patients. In the pre-dexamethasone era, this model yielded sensitivity, specificity, positive predictive value, and negative predictive values of 0.70, 0.50, 0.54, and 0.67, respectively. For all patients in our cohort, the values were 0.8, 0.36, 0.50, and 0.71, respectively.

DISCUSSION

In this article, we reviewed 13 COVID-19 prediction articles intended for use in ICU settings, which incidentally were developed during the first wave of the pandemic. We were able to externally validate five of these models using routinely collected data from a large ICU in the United Kingdom and found that only one model showed acceptable reproducibility in a pre-dexamethasone cohort and that all the models performed poorly when using the entire cohort comprising patients from both pre- and post-dexamethasone eras.

van Royen et al (39) described the "leaky prognostic model pipeline" and the reasons why prognostic models are not adopted into clinical practice. In this study, we examined aspects of this "pipeline" and identified articles that failed, in the terminology by van Royen et al (39), either as a result of "incomplete reporting of prediction model," "expensive, unavailable predictors," "predictions not trusted," or "predictions outdated."

Our first finding is the difficulty in externally validating published models due to incomplete reporting of the model. The models included in this study represent a broad scope of models in the literature and for the majority (8/13 articles), there were insufficient details to allow external validation. This is not the first study to identify deficiencies in reporting quality and that the adoption of the TRIPOD guidelines by authors to clearly report their models is imperative (2, 40–42).

Secondly ambiguities in how each predictor was handled and when the data were collected were common. In critical care, timing is key, and there can be rapid changes

TABLE 1.

Summary of the 13 Articles Identified in This Review and Whether They Met the Requirements of Questions 1 and 2 As Outlined in the Introduction

References	Endpoint	Predictors Used	Further Specific Inclusion/ Exclusion Criteria	Development Cohort Size	Model Type	Q1	Q2
Gerotziafas et al (24)	IMV	sP-selectin, D-dimer, tissue factor pathway inhibitor, tissue factor activity, factor XII, IL-6		118	T-distributed stochastic neighbor embedding	✗	✗
Arina et al (25)	IMV or death	C-reactive protein, N-terminal pro B-type natriuretic peptide	Continuous positive airway pressure use for at least 6 hr in a 24-hr period	93	Logistic regression	✓	✓
Cao et al (26)	Death	Urea, hs-CRP	IMV	77	Logistic regression	✓	✓
Pan et al (27)	Death	Lactate dehydrogenase, prothrombin time, creatinine, lymphocyte percentage, total bilirubin, albumin level, neutrophil percentage, eosinophil percentage		98	XGBoost	✗	✓
Popadic et al (28)	Death	Albumin, IL-6, D-dimer		160	Logistic regression	✓	✗
Vaid et al (29)	Death	Not reported	IMV < 48 hr after ICU admission	4029	Multiple	✗	✗
Vaid et al (30)	Death	Not reported	IMV < 48 hr after ICU admission	4098	XGBoost	✗	✗
Vassiliou et al (31)	Death	Soluble E-selectin, sP-selectin, angiopoietin-2, soluble intercellular adhesion molecule-1, von Willebrand factor		38	Unclear	✗	✗
Wang et al (32)	Death	Age, aspartate transaminase, urea, chest tightness		104	Logistic regression	✓	✗
Falandry et al (33)	Death within 30 d	Age, Instrumental Activities of Daily Living Scale 8	Age ≥ 60	231	Logistic regression	✓	✗
Leoni et al (34)	Death censored at 28 d	Age, obesity, procalcitonin, SOFA, PaO_2/FIO_2	IMV or $FIO_2 > 60\%$	229	Cox regression	✓	✓
Leoni et al (35)	Death censored 28 d	Modified NUTrition Risk In the Critically ill, hs-CRP, neutrophils		98	Cox regression	✓	✓
Moisa et al (36)	Death censored at 28 d	Age, neutrophil-to-lymphocyte ratio, SOFA		425	Cox regression	✓	✓

hs-CRP = high-sensitivity C-reactive protein, IL-6 = interleukin-6, IMV = invasive mechanical ventilation, SOFA = Sequential Organ Failure Assessment, sP-selectin = soluble platelet selectin, XGBoost = eXtreme Gradient Boosting.

TABLE 2.
The Characteristics of the Patients Included in Our Validation Cohort

Characteristic	All Patients, *n* = 302	Pre-Dexamethasone, *n* = 55
Age (yr)	57.5 (49.8–66.2)	57.0 (49.5–70.0)
Male	207 (67.2%)	40 (72.7%)
Ethnicity		
White	123 (39.9%)	28 (50.9%)
Asian	88 (28.6%)	14 (25.5%)
Black	54 (17.5%)	9 (16.4%)
Other	6 (1.9%)	2 (3.6%)
Not stated	37 (12.0%)	2 (3.6%)
BMI (kg/m^2)	29.4 (26.2–35.2)	27.8 (25.1–32.5)
Obese (BMI > 30)	147 (47.7%)	20 (36.4%)
Sequential Organ Failure Assessment score	6.0 (3.0–8.0)	5.5 (3.0–8.8)
Modified NUTrition Risk In the Critically ill score	4.0 (2.0–5.0)	4.0 (3.0–5.0)
C-reactive protein (mg/L)	121.0 (71.2–189.8)	156.0 (102.0–239.0)
Lymphocytes (×10^9/L)	0.70 (0.50–0.90)	0.70 (0.50–1.00)
Neutrophils (×10^9/L)	7.9 (5.6–10.5)	7.4 (6.0–10.0)
Neutrophil-to-lymphocyte ratio	12.4 (7.6–19.4)	12.4 (7.6–19.4)
Procalcitonin (ng/mL)	0.30 (0.14–0.92)	0.60 (0.22–1.30)
Urea (mmol/L)	6.8 (4.9–10.4)	6.2 (3.9–12.1)
Invasive mechanical ventilation	143 (46.4%)	23 (41.8%)
Died	132 (42.9%)	28 (50.9%)

BMI = body mass index.
n is all patients included in one of the four models being externally validated. See Supplementary Tables 1 and 2 (http://links.lww.com/CCX/B325) for demographics of patients included in each validation. Data reported as *n* (%) or median (interquartile range). Pre-dexamethasone era defined as before June 16, 2020.

TABLE 3.
The Concordance of Each of the Models That Were Validated Using Patients in the Pre-Dexamethasone Era and the Entire Validation Cohort

	Original Publication		Pre-Dexamethasone Era		All Patients	
References	*n*	Reported Concordance Statistic	*n*	Concordance Statistic	*n*	Concordance Statistic
Arina et al (25)	93	0.804 (0.728– 0.880)	27	0.793 (0.618–0.968)	103	0.596 (0.482–0.710)
Cao et al (26)	77	0.857 (0.77–0.94)	22	0.567 (0.321–0.813)	141	0.558 (0.460–0.655)
Leoni et al (34)	229	0.821 (0.766–0.876)	33	0.605 (0.471–0.739)	204	0.564 (0.506–0.622)
Leoni et al (35)	98	0.720 (0.67–0.79)	34	0.560 (0.428–0.692)	201	0.546 (0.484–0.608)
Moisa et al (36)	425	0.697 (0.755–0.833)	53	0.672 (0.574–0.770)	291	0.686 (0.640–0.732)

The originally published outcomes are displayed for reference.

in patients' health meaning that over a 24-hour period, two very different measurements could be taken. The most complex parameter to handle was the SOFA score, which we calculated retrospectively. It is well recognized that when and how to calculate a SOFA score can be open to interpretation (21, 43). It serves as a useful illustration of the importance of ensuring that predictor variables are specified in detail when models are reported.

Attributing reasons why the models tested were not able to reproduce the same level of discrimination is complex. Each study used was developed in a single or small number of centers without large, multinational cohorts. Many of the models were generated during an evolving pandemic, in which the treatment pathways and the COVID-19 virus itself were changing (44, 45). There are potentially intrinsic differences in healthcare systems, the patient populations and host response to the COVID-19 virus (46, 47). Differences in care pathways may also affect which patients were included in each dataset. For example, Cao et al (26) used patients who received IMV but the criteria for initiation of IMV may be different in their center compared with others. Similarly, the criteria for ICU admission may vary between centers.

Our findings are in keeping with those of Meijs et al (18) who used data from a regional clinical network to validate COVID-19 models in ICU, although these were developed on general wards and tested in ICUs. Even in 2020, the low quality, high volume of published COVID-19 models was recognized as a significant problem with calls for better data sharing and better reporting (48). In ICU, where the pressure on beds is high and the decisions around such resource use so important, clinical prediction models are particularly appealing and yet demonstrably unhelpful at present. On the basis of our findings, we would encourage bedside intensivists to reflect on the role of clinical prediction models in their clinical decision-making.

A positive outcome from the COVID-19 pandemic has been the agility of the clinicians and academics to investigate and implement new treatments for COVID-19. However, it has proved difficult to match this performance with clinical prediction models on account of the challenges described above. Arguably, these challenges may be overcome by improved specification and reporting of prediction models and by collaboration between institutions to create larger training datasets, which may yield more generalizable results. We concede that numerous data security and governance barriers make such collaboration complex and perhaps more so in the accelerated pandemic timescales. It is encouraging to see the development of multicenter, collaborative critical care datasets such as the Critical Care Health Informatics Collaborative in the United Kingdom or the electronic ICU dataset in the United States (49, 50). These have the potential to overcome many of the above challenges in the future.

We acknowledge several limitations. Our validation data are drawn from a single U.K. center and may not be representative of all ICUs. However, this does not prevent our study from highlighting the need for caution when applying predictive models in new contexts. A number of laboratory parameters and subjective assessments are not routinely collected at our institution, which prevented validation of some models. This is likely to be true for many institutions and highlights the importance of building prediction models utilizing routinely collected and widely available prediction parameters. The sample size of the pre-dexamethasone cohort was small and led to large CIs. Finally, we recognize our data were extracted retrospectively and that patients were cared for with the standard of care at the time of admission, which changed rapidly throughout the pandemic. There were several significant changes in the standard of care, and it was the consensus of our study team to only examine across the first major change in care; the introduction of dexamethasone. This decision reflected the fact that the majority of the models assessed were developed before the RECOVERY trial's results (12, 23). The introduction of other treatments for COVID-19 may, therefore, have influenced our findings but this only serves to underline the importance of context when applying a predictive model in new circumstances.

CONCLUSIONS

This study highlights the caution required when interpreting COVID-19 prediction models in ICU and when translating results into clinical practice. Researchers should ensure models are fully reported and contain routinely collected, widely available predictors that are unambiguously defined. Where possible, models should be developed and tested using multicenter research datasets, which are increasingly available. Clinicians should exercise caution and ensure that any model they use has been externally validated and clinically tested.

ACKNOWLEDGMENTS

We thank the U.K. Intensive Care National Audit and Research Centre data contributors and the Clinical Data Science Unit at Manchester University National Health Service Foundation Trust for their help.

1 *Research & Innovation, Manchester Academic Health Science Centre, Manchester University NHS Foundation Trust, Manchester, United Kingdom.*

2 *Centre for Biostatistics, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom.*

3 *Adult Critical Care, Manchester Royal Infirmary, Manchester University NHS Foundation Trust, Manchester, United Kingdom.*

4 *Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom.*

5 *Division of Informatics, Imaging & Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom.*

The MRI Critical Care Data Group are as follows: Dr. Andrew Martin, Dr. Daniel Haley, Dr. Henry Morriss, and Dr. John Moore.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (http://journals.lww.com/ccejournal).

All authors conceived the study. Drs. Parker and Wilson performed the data extraction. Mr. Bate performed the analysis. Mr. Bate, Dr. Stokes, Dr. Greenlee, Dr. Goh, and Dr. Whiting reviewed studies. All authors resolved data ambiguities. Mr. Bate, Dr. Stokes, Dr. Greenlee, Dr. Parker, and Dr. Wilson produced the first draft. All authors critically reviewed and edited the article.

The authors have disclosed that they do not have any potential conflicts of interest.

Mr. Bate and Dr. Stokes contributed equally.

For information regarding this article, E-mail: anthony.wilson@mft.nhs.uk

REFERENCES

1. COVID-19 Open Access Project: Living Evidence on COVID-19. 2020. Available at: https://ispmbern.github.io/covid-19/living-review/. Accessed March 16, 2023
2. Wynants L, Van Calster B, Collins GS, et al: Prediction models for diagnosis and prognosis of Covid-19: Systematic review and critical appraisal. *BMJ* 2020; 369:m1328
3. de Jong VM, Rousset RZ, Antonio-Villa NE, et al: Clinical prediction models for mortality in patients with Covid-19: External validation and individual participant data meta-analysis. *BMJ* 2022; 378:e069881
4. Bleeker SE, Moll HA, Steyerberg EW, et al: External validation is necessary in prediction research: A clinical example. *J Clin Epidemiol* 2003; 56:826–832
5. Siontis GCM, Tzoulaki I, Castaldi PJ, et al: External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol* 2015; 68:25–34
6. Stokes V, Goh KY, Whiting G, et al: External validation of a prediction model for CPAP failure in COVID-19 patients with severe pneumonitis. *Crit Care* 2022; 26:293
7. Dhiman P, Ma J, Andaur Navarro CL, et al: Methodological conduct of prognostic prediction models developed using machine learning in oncology: A systematic review. *BMC Med Res Methodol* 2022; 22:1–16
8. Collins GS, Reitsma JB, Altman DG, et al: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD statement. *Br J Surg* 2015; 102:148–158
9. Moons KG, Altman DG, Reitsma JB, et al: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and elaboration. *Ann Intern Med* 2015; 162:W1–73
10. Cox EGM, Wiersema R, Eck RJ, et al: External validation of mortality prediction models for critical illness reveals preserved discrimination but poor calibration. *Crit Care Med* 2023; 51:80–90
11. Weerakkody S, Arina P, Glenister J, et al: Non-invasive respiratory support in the management of acute COVID-19 pneumonia: Considerations for clinical practice and priorities for research. *Lancet Respir Med* 2022; 10:199–213
12. The RECOVERY Collaborative Group: Dexamethasone in hospitalized patients with Covid-19. *N Engl J Med* 2021; 384:693–704
13. The RECOVERY Collaborative Group: Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): A randomised, controlled, open-label, platform trial. *Lancet* 2021; 397:1637–1645
14. Amit B, Steven J, Gabriel M, et al: Vaccination reduces need for emergency care in breakthrough COVID-19 infections: A multicenter cohort study. *Lancet Reg Health Am* 2021; 4:100065
15. Gupta RK, Marks M, Samuels THA, et al; UCLH COVID-19 Reporting Group: Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: An observational cohort study. *Eur Respir J* 2020; 56:2003498
16. Paul MELD, Noortje Z, Sander MJK, et al: Performance of prediction models for short-term outcome in COVID-19 patients in the emergency department: A retrospective study. *Ann Med* 2021; 53:402–409
17. Wickstrøm KE, Vitelli V, Carr E, et al: Regional performance variation in external validation of four prediction models for severity of COVID-19 at hospital admission: An observational multi-centre cohort study. *PLoS One* 2021; 16:e0255748
18. Meijs DAM, van Kuijk SMJ, Wynants L, et al; CoDaP Investigators: Predicting COVID-19 prognosis in the ICU remained challenging: External validation in a multinational regional cohort. *J Clin Epidemiol* 2022; 152:257–268
19. Parker AJ, Mishra M, Tiwary P, et al: A tale of two waves: Changes in the use of noninvasive ventilation and prone positioning in critical care management of coronavirus disease 2019. *Crit Care Explor* 2021; 3:e0587
20. Parker AJ, Humbir A, Tiwary P, et al: Recovery after critical illness in COVID-19 ICU survivors. *Br J Anaesth* 2021; 126:e217–e219
21. Lambden S, Laterre PF, Levy MM, et al: The SOFA score-development, utility and challenges of accurate assessment in clinical trials. *Crit Care* 2019; 23:374

22. Milone MTBA, Kamath AFMD, Israelite CLMD: Converting between high- and low-sensitivity C-reactive protein in the assessment of periprosthetic joint infection. *J Arthroplasty* 2014; 29:685–689
23. University of Oxford: Low-Cost Dexamethasone Reduces Death by Up to One Third in Hospitalised Patients With Severe Respiratory Complications of COVID-19. 2020. Available at: https://www.ox.ac.uk/news/2020-06-16-low-cost-dexamethasone-reduces-death-one-third-hospitalised-patients-severe. Accessed September 11, 2022
24. Gerotziafas GT, Van Dreden P, Fraser DD, et al: The COMPASS-COVID-19-ICU study: Identification of factors to predict the risk of intubation and mortality in patients with severe COVID-19. *Hemato* 2022; 3:204–218
25. Arina P, Baso B, Moro V, et al; UCL Critical Care COVID-19 Research Group: Discriminating between CPAP success and failure in COVID-19 patients with severe respiratory failure. *Intensive Care Med* 2021; 47:237–239
26. Cao L, Zhang S, Wang E, et al: The CB index predicts prognosis of critically ill COVID-19 patients. *Ann Transl Med* 2020; 8:1654
27. Pan P, Li Y, Xiao Y, et al: Prognostic assessment of COVID-19 in the intensive care unit by machine learning methods: Model development and validation. *J Med Internet Res* 2020; 22:e23128
28. Popadic V, Klasnja S, Milic N, et al: Predictors of mortality in critically ill COVID-19 patients demanding high oxygen flow: A thin line between inflammation, cytokine storm, and coagulopathy. *Oxid Med Cell Longevity* 2021; 2021:6648199
29. Vaid A, Jaladanki SK, Xu J, et al: Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: Machine learning approach. *JMIR Med Inform* 2021; 9:e24207
30. Vaid A, Somani S, Russak AJ, et al: Machine learning to predict mortality and critical events in a cohort of patients with COVID-19 in New York City: Model development and validation. *J Med Internet Res* 2020; 22:e24018
31. Vassiliou AG, Keskinidou C, Jahaj E, et al: ICU admission levels of endothelial biomarkers as predictors of mortality in critically ill COVID-19 patients. *Cells* 2021; 10:186–113
32. Wang B, Zhong F, Zhang H, et al: Risk factors analysis and nomogram construction of non-survivors in critical patients with COVID-19. *Jpn J Infect Dis* 2020; 73:452–458
33. Falandry C, Bitker L, Abraham P, et al: Senior-COVID-rea cohort study: A geriatric prediction model of 30-day mortality in patients aged over 60 years in ICU for severe COVID-19. *Aging Dis* 2022; 13:614–623
34. Leoni MLG, Lombardelli L, Colombi D, et al: Prediction of 28-day mortality in critically ill patients with COVID-19: Development and internal validation of a clinical prediction model. *PLoS One* 2021; 16:e0254550
35. Leoni MLG, Moschini E, Beretta M, et al: The modified NUTRIC score (mNUTRIC) is associated with increased 28-day mortality in critically ill COVID-19 patients: Internal validation of a prediction model. *Clin Nutr ESPEN* 2022; 48:202–209
36. Moisa E, Corneci D, Negutu MI, et al: Development and internal validation of a new prognostic model powered to predict 28-day all-cause mortality in ICU COVID-19 patients—the COVID-SOFA score. *J Clin Med* 2022; 11:4160
37. Van der Maaten L, Hinton G: Visualizing data using t-SNE. *J Mach Learn Res* 2008; 9:2579–2605
38. Chen T, Guestrin C (Eds): *XGBoost: A Scalable Tree Boosting System*. Ithaca, NY, Cornell University Library, 2016
39. van Royen FS, Moons KG, Geersing G-J, et al: Developing, validating, updating and judging the impact of prognostic models for respiratory diseases. *Eur Respir J* 2022; 60:2200250
40. Canturk TC, Czikk D, Wai EK, et al: A scoping review of complication prediction models in spinal surgery: An analysis of model development, validation and impact. *N Am Spine Soc J* 2022; 11:100142
41. Carr BL, Jahangirifar M, Nicholson AE, et al: Predicting postpartum haemorrhage: A systematic review of prognostic models. *Aust N Z J Obstet Gynaecol* 2022; 62:813–825
42. Dhiman P, Ma J, Andaur Navarro CL, et al: Risk of bias of prognostic models developed using machine learning: A systematic review in oncology. *Diagn Progn Res* 2022; 6:13
43. Vincent JL, Moreno R, Takala J, et al: The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med* 1996; 22:707–710
44. Rolland S, Lebeaux D, Tattevin P, et al: Evolution of practices regarding COVID-19 treatment in France during the first wave: Results from three cross-sectional surveys (March to June 2020). *J Antimicrob Chemother* 2021; 76:1372–1374
45. Centers for Disease Control and Prevention: SARS-CoV-2 Variant Classifications and Definitions. 2021. Available at: https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html. Accessed March 16, 2023
46. Public Health England: Disparities in the Risk and Outcomes of COVID-19. 2020
47. Mesotten D, Meijs DAM, van Bussel BCT, et al; COVID Data Platform (CoDaP) Investigators: Differences and similarities among COVID-19 patients treated in seven ICUs in three countries within one region: An observational cohort study. *Crit Care Med* 2022; 50:595–606
48. Sperrin M, Grant SW, Peek N: Prediction models for diagnosis and prognosis in Covid-19. *BMJ (Clin Res Ed)* 2020; 369:m1464
49. Harris S, Shi S, Brealey D, et al: Critical Care Health Informatics Collaborative (CCHIC): Data, tools and methods for reproducible research: A multi-centre UK intensive care database. *Int J Med Inform* 2018; 112:82–89
50. Pollard TJ, Johnson AEW, Raffa JD, et al: The eICU collaborative research database, a freely available multi-center database for critical care research. *Sci Data* 2018; 5:180178