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

Journal of Clinical Epidemiology ■ (2023) ■

Journal of Clinical Epidemiology

REVIEW ARTICLE

Validity of prognostic models of critical COVID-19 is variable. A systematic review with external validation

Gabriela Cárdenas-Fuentes[a,b,c,d,*], Magda Bosch de Basea[a,b,c], Inés Cobo[a,b,c], Isaac Subirana[e,f], Mario Ceresa[g], Ernest Famada[b], Elena Gimeno-Santos[a,b,c,h], Laura Delgado-Ortiz[a,b,c], Rosa Faner[i,j,k], María Molina-Molina[k,l,m], Àlvar Agustí[h,i,j,k], Xavier Muñoz[k,n,o], Oriol Sibila[h,i,k], Joaquim Gea[b,k,p,q], Judith Garcia-Aymerich[a,b,c]

[a]*ISGlobal, Barcelona, Spain*
[b]*Universitat Pompeu Fabra (UPF), Barcelona, Spain*
[c]*CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain*
[d]*School of Health Sciences, Blanquerna-Universitat Ramon Llull, Barcelona, Spain*
[e]*Instituto Hospital del Mar de Investigaciones Médicas (IMIM), Barcelona, Spain*
[f]*CIBER Enfermedades Cardiovasculares (CIBERCV), ISCIII, Spain*
[g]*BCNMedTech, Universitat Pompeu Fabra (UPF), Barcelona, Spain*
[h]*Respiratory Institute, Hospital Clinic, Barcelona, Spain*
[i]*Instituto de Investigaciones Biomédicas August Pi i Sunyer (IDIBAPS), Barcelona, Spain*
[j]*Universitat de Barcelona, Barcelona, Spain*
[k]*CIBER Enfermedades Respiratorias (CIBERES), ISCIII, Spain*
[l]*Servicio de Neumología, Hospital Universitario de Bellvitge, L'Hospitalet de Llobregat, Spain*
[m]*Instituto de Investigación Biomédica de Bellvitge (IDIBELL), L'Hospitalet de Llobregat, Spain*
[n]*Servicio de Neumología, Hospital Universitario Vall d'Hebron, Barcelona, Spain*
[o]*Departamento de Biología celular, fisiología e inmunología, Universitat Autònoma de Barcelona, Barcelona, Spain*
[p]*Servicio de Neumología, Hospital del Mar-IMIM, Barcelona, Spain*
[q]*Fundació Barcelona Respiratory Network (BRN), Barcelona, Spain*

Accepted 25 April 2023; Published online xxxx

Abstract

Objectives: To identify prognostic models which estimate the risk of critical COVID-19 in hospitalized patients and to assess their validation properties.

Competing interests: X.M. has received fees as a speaker, scientific advisor, and as a participant of clinical studies of (in alphabetical order) AstraZeneca, Boehringer Ingelheim, Chiesi, Faes, GlaxoSmithKline, Menarini, Mundipharma, Novartis, Sanofi, and Teva. R.F. has received grants from GlaxoSmithKline LLC, Menarini, AstraZeneca, ISC-III, and the Spanish Health Service; consulting fees from GlaxoSmithKline LLC; and other payments (lectures, presentations, and speaker) from Chiesi. J.G. has received grants from SEPAR and from FIS, ISCIII—EU Program Horizon; fees from GSK and Menarini; and other payments (lectures, presentations, and speaker) from Menarini. M.M.M. has received grants supported by Roche and Boehringer Ing.; other payments (lectures, presentations, and speaker) from Esteve-Teijin, Chiesi; and support for attending meetings/travel from Boehringer Ing. The rest of the authors declared that they have no conflict of interest.

Ethics approval and consent to participate: The study protocol was approved by the Ethics Committee of Parc de Salut Mar (2020/9206/I). The study used only retrospective and routinely collected data, and therefore, informed consent was waived.

Consent for publication: Not applicable.

Availability of data and materials: The data that support the findings of this study are available from HM and Institut Català de la Salut, but restrictions apply to the availability of these data and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of HM and Institut Català de la Salut.

Funding: Supported by AstraZeneca Farmacéutica (ESR-20-20,762). ISGlobal acknowledges support from the Spanish Ministry of Science and Innovation through the 'Centro de Excelencia Severo Ochoa 2019—2023' Program (CEX2018-000806-S) and support from the Generalitat de Catalunya through the CERCA Program. No involvement of funding sources in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Author contributions: G.C.F., M.B.B., I.C., and J.G.A. performed the systematic review of research articles; G.C.F., I.S., J.G.A., and I.C. analyzed and interpreted the data; G.C.F., M.B.B., M.C., E.F., E.G.S., L.D.O., R.F., M.M.M., A.A., X.M., O.S., J.G., and J.G.A. were major contributors in writing the manuscript. All authors read and approved the final manuscript.

* Corresponding author. Barcelona Institute for Global Health, Carrer Doctor Aiguader, 88, 08003 Barcelona, Spain. Tel.: +34-93-214-7300.

E-mail address: gabriela.cardenas@isglobal.org (G. Cárdenas-Fuentes).

https://doi.org/10.1016/j.jclinepi.2023.04.011

Study Design and Setting: We conducted a systematic review in Medline (up to January 2021) of studies developing or updating a model that estimated the risk of critical COVID-19, defined as death, admission to intensive care unit, and/or use of mechanical ventilation during admission. Models were validated in two datasets with different backgrounds (HM [private Spanish hospital network], $n = 1{,}753$, and ICS [public Catalan health system], $n = 1{,}104$), by assessing discrimination (area under the curve [AUC]) and calibration (plots).

Results: We validated 18 prognostic models. Discrimination was good in nine of them (AUCs $\geq$ 80%) and higher in those predicting mortality (AUCs 65%−87%) than those predicting intensive care unit admission or a composite outcome (AUCs 53%−78%). Calibration was poor in all models providing outcome's probabilities and good in four models providing a point-based score. These four models used mortality as outcome and included age, oxygen saturation, and C-reactive protein among their predictors.

Conclusion: The validity of models predicting critical COVID-19 by using only routinely collected predictors is variable. Four models showed good discrimination and calibration when externally validated and are recommended for their use.

Keywords: COVID-19; Critical disease; Intensive care unit; Prognostic models; External validation; Epidemiology

1. Introduction

Since the beginning of the pandemic, the substantial number of critical COVID-19 cases has overwhelmed the healthcare systems around the world [1]. Despite that only about 5% of COVID-19 patients require critical care [2], the deaths from COVID-19 have already surpassed six million [3]. Therefore, a prompt and standardized identification of patients at risk of developing critical COVID-19 is crucial.

Many studies have developed models aiming to predict critical COVID-19 in patients hospitalized with the disease [4]. These models have claimed a potential to decrease the chance of fatal outcomes, assist clinicians in performing risk stratification, and optimize the use of health resources [4]. In fact, prior to the pandemic, prognostic models have been extensively used in the medical field. These rely on the use of patient-level information (e.g., demographic, clinical, and laboratory results) to estimate the probability of developing a future clinical event [5]. To apply a prognostic model in clinical settings, it is imperative to previously validate its performance in a group of patients different from the one used for model development. This process, known as external validation, helps determining whether the model properties estimated in the development dataset are real or are due to overfitting. A proper validation implies assessing the capacity of a model of, first, ordering individuals as per their risk of presenting the event (i.e., discrimination), and second, providing a risk estimate with similar magnitude to the real risk of presenting the event (i.e., calibration). Unfortunately, very few of the more than a hundred published prognostic models aiming to predict critical COVID-19 have been adequately externally validated [4], and therefore, there is no evidence on how they would perform on samples coming from different hospitals and different locations.

Hence, we aimed to (1) identify prognostic models designed to estimate the risk of critical COVID-19 in patients hospitalized because of the disease and (2) validate their performance by assessing their discrimination capacity and calibration in two external datasets.

2. Methods

2.1. Systematic identification of prognostic models

A systematic review was registered in PROSPERO repository (CRD42021235106) and performed as per a previously written protocol (Supplementary Material, Item 1) based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [6,7] and the guide to perform systematic reviews of prognostic factor studies described by Riley et al. [8].

2.1.1. Search strategy

A systematic search was performed using the Medline database (Supplementary Material, Item 2) covering the period from December 2019 (first detection of COVID-19 disease) to January 2021, including models developed prior to COVID-19 vaccination. The search was supplemented through hand search of the reference lists of relevant studies. Retrieved articles were saved in the Mendeley platform and screened for inclusion criteria using the CADIMA software [9]. Two reviewers (G.C.F. and M.B.B.) independently screened titles and abstracts of retrieved references (consistency with Kappa of 0.8) and screened the full text of those considered eligible. Where consensus was not achieved, a third reviewer (J.G.A.) was consulted.

2.1.2. Selection criteria

Eligible articles had to (1) be peer-reviewed and published in English language; (2) have a longitudinal (prospective or retrospective) design, from hospital admission to event or discharge, with any length of follow-up; (3) aim at developing or updating a multivariable prognostic model; (4) include general population (not disease subgroups) hospitalized due to COVID-19, diagnosed through real-time reverse transcription polymerase chain reaction, genetic sequencing, or radiological imaging compatible with COVID-19; (5) include only information routinely collected at hospital admission as model's predictors (i.e., demographics, clinical signs/symptoms, laboratory results, medical history, and clinical scores); (6) use critical

What is new?

Key findings

- We identified nine models that discriminate the risk of critical COVID-19 well in two external datasets.
- Four of them are additionally well calibrated, provide a point-based score, and include predictors easy to obtain.

What this adds to what was known?

- The several critical COVID-19 prognostic models that lacked external validation have been identified and validated in the context of this study.

What is the implication and what should change now?

- The use of any of these four models will assist clinicians in performing risk stratification, thus decreasing the chances of fatal outcomes and optimizing the use of health resources.

COVID-19 as the model outcome, defined by the use of invasive or noninvasive mechanical ventilation, a registered admission to the intensive care unit (ICU), in-hospital or out-hospital death, or by a combination of them; (7) provide at least one indicator of model performance; and (8) perform an internal or external validation (only in models labeled as having a high potential of over/under fitting) (Supplementary Material, Item 1). Studies including only a population subgroup (e.g., patients with cancer); excluding population subgroups (e.g., pregnant participants); not providing regression coefficients, point-based scores, or equivalent predictors' weights; or not providing predictors' units or summary statistics were discarded.

2.1.3. Data extraction

Key features of selected articles (study sample, model's predictors, outcome variables, and modeling techniques) were extracted by two independent reviewers (G.C.F. and M.B.B.). Data were organized in standardized forms designed as per the Prediction Study Risk of Bias Assessment Tool Guidelines (PROBAST) [8]. Discrepancies during the data extraction were solved by consulting a fourth reviewer (I.C.).

2.2. Validation of prognostic models

2.2.1. Study design and population

Results were reported as per the validation section of the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement [10]. Models identified in the systematic review were validated in two datasets, representing two longitudinal retrospective cohorts with different backgrounds: the private and the public health system. These datasets included anonymous information extracted from electronic health records of patients hospitalized due to COVID-19, followed from hospital admission to the development of critical COVID-19 (as defined in each validated model) or discharge. The Hospital Universitario de Madrid (HM) dataset included patients hospitalized in the dependencies of the HM network, formed by 16 private hospitals in Spain, three of which are in Catalonia (further details in Supplementary Material, Item 3). We included patients with COVID-19 admitted between February 5 and April 20, 2020, covering the first COVID-19 wave in Spain, when COVID-19 vaccination was still not available [11]. The Institut Català de la Salut (ICS) dataset included patients hospitalized in the dependencies of the Catalan Public Health System, which serves almost six million people living in Catalonia, a region in the Northeast of Spain. We included patients with COVID-19 admitted between March 1 and August 14, 2020, covering the first and second COVID-19 wave in Spain who were, therefore, also unvaccinated [11]. The study protocol was approved by the Ethics Committee of Parc de Salut Mar (2020/9206/I).

Diagnosis of COVID-19 was defined as a registration of 'confirmed COVID-19' in the HM dataset and as a positive real-time reverse transcription polymerase chain reaction test in the ICS dataset. We excluded individuals younger than 18 years ($n = 5$); those who died, were admitted to the ICU, or were discharged on the day of admission ($n = 205$); those who remained hospitalized at the time of data dumping ($n = 224$); and—because most prognostic models included signs and symptoms—those with no available information on these variables at hospital admission ($n = 21{,}951$). The final study sample included 2,857 participants (1,753 in HM dataset and 1,104 in ICS dataset) (Supplementary Material, Item 4).

2.2.2. Variables and procedures

We obtained data collected as part of routinely clinical assistance, including information on: (1) demographics, underlying comorbidities, clinical signs, symptoms, and laboratory results at hospital admission; (2) procedures performed during the hospitalization (e.g., administration of invasive mechanical ventilation); and (3) administrative registries (e.g., dates of admission and discharge of the hospital and ICU). In variables with repeated measurements during hospitalization (e.g., laboratory results, signs, or symptoms), we only included the first one.

2.2.3. Definition of critical COVID-19

Critical COVID-19 has been defined by the World Health Organization as presenting sepsis, septic shock, or other conditions that would normally require the provision of life-sustaining therapies such as mechanical ventilation or vasopressor therapy [12]. In the present validation study,

we defined critical COVID-19 as in each original study, this is: models using mortality as the outcome—regardless of being inside or outside the hospital—were validated against in-hospital mortality, respecting the original time frames (if stated); models using a registered ICU admission were validated against a registered ICU admission; and models using a composite outcome were validated against a standard composite outcome, created as having an ICU admission and/or in-hospital death.

2.3. Statistical analysis

A detailed description of the statistical analysis is provided in the Supplementary Material, Item 5. In brief, we validated the prognostic models separately in both datasets. Since missing data in each dataset accounted for $< 20\%$ of observations (Supplementary Material, Items 6-7) and were considered completely at random (e.g., a hospital did not collect a variable at all) or at random (e.g., patients admitted to the ICU had fewer missing data), we performed multiple imputations using chained equations, and performed the main analysis using the imputed datasets.

For each participant and prognostic model, we calculated (1) the predicted probability of developing critical COVID-19 (in models providing regression coefficients or similar) or (2) a final score (in models providing a point-based scoring system). Discrimination was assessed by calculating the area under the curve (AUC) in models providing outcome's probabilities based on a logistic regression or in those providing a point-based scoring system, or by calculating the Harrell's C index in models providing probabilities based on a Cox regression.

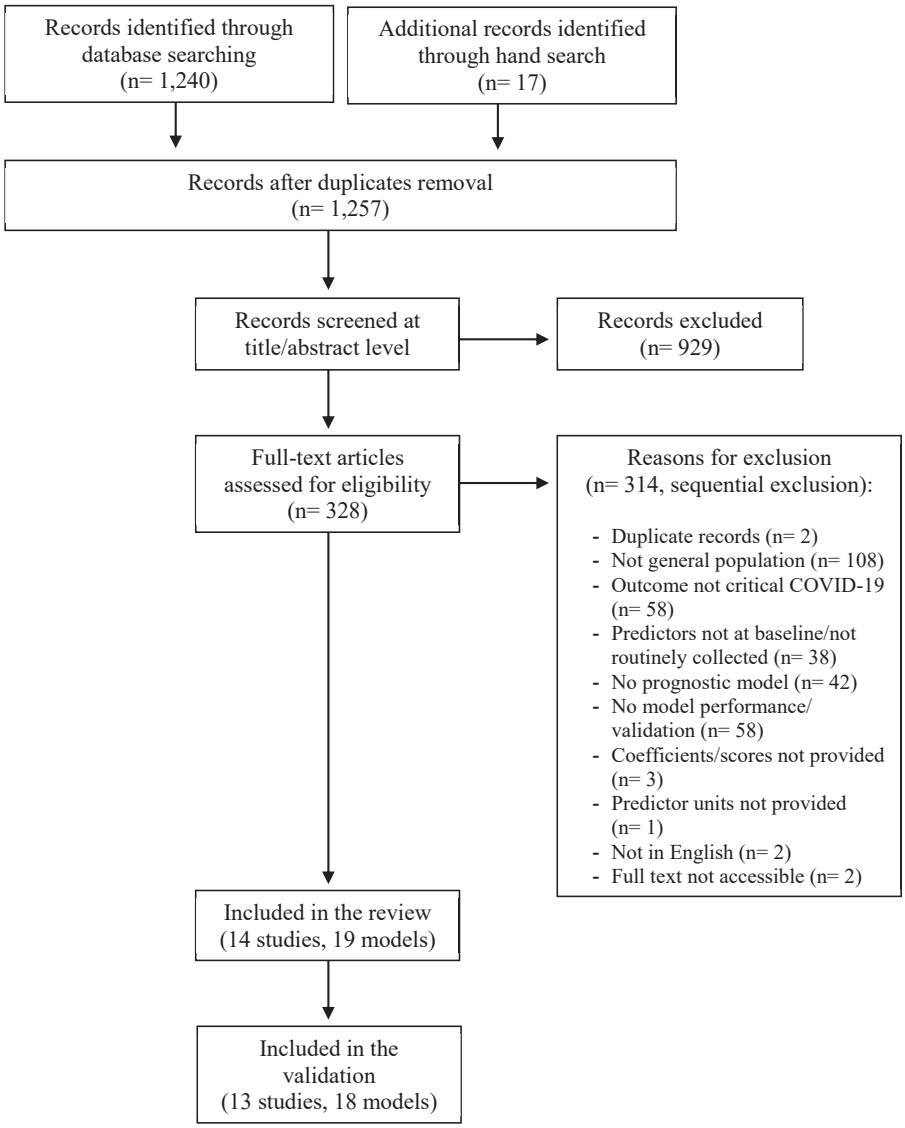

Fig. 1. PRISMA flow diagram.

Calibration was assessed by inspecting calibration plots in models providing outcome's probabilities or by plotting a final score against the observed outcome's probability in models providing a point-based scoring system.

As secondary analysis, we assessed discrimination and calibration in all models against 28-day in-hospital mortality. As sensitivity analyses, we repeated the analysis (1) in the complete-case datasets, to account for the possibility of inducing bias during imputation and (2) after excluding participants transferred to other hospitals at the moment of discharge (only in HM dataset, as this situation did not occur in the ICS dataset), to account for potential outcome misclassification. All analyses were conducted using R 4.1.2 (R Foundation for Statistical Computing, 2021).

3. Results

3.1. Identified prognostic models of critical COVID-19

The search retrieved 1,257 articles (fig 1) (1,240 from Medline and 17 from the hand search). Among them, 328 were eligible for full-text screening: 327 due to fulfilling the title and abstract screening criteria and one—despite including participants with confirmed and nonconfirmed COVID-19 diagnosis—because of its large sample size, international coverage, and statistical rigor [13]. During full-text screening, and after removing duplicate records ($n = 2$), we excluded those not meeting participant's, outcome's, or predictor's selection criteria ($n = 204$); those not deriving a prognostic model ($n = 42$); those not assessing model's performance or validity ($n = 58$, 67% of the articles deriving a model); and those not providing model coefficients or predictor units ($n = 3$, 3% of the articles deriving a model). Fourteen studies were selected after full-text screening, from which we extracted 19 prognostic models [13−26]. Models differed in their study populations, selection of predictors and outcomes, and modeling techniques (Table 1 and Supplementary Material, Item 8). They were developed in seven countries (China, England, Italy, Scotland, Spain, the United States, and Wales), had samples sizes ranging between 176 and 35,463 individuals (median 893 individuals), and were mostly developed in retrospective cohorts. In the identified models, the most prevalent critical COVID-19 definition was death (74%), followed by ICU admission (21%) and composite outcome ($n = 5\%$). Forty four different predictors were included in the models, the most frequent being age (89%), oxygen saturation (63%), C-reactive protein (37%), lactate dehydrogenize (32%), and lymphocyte count (32%). Of the 19 models, 17 provided the components of the regression formula allowing the estimation of the outcome probability, from which eight also provided a scoring system. The remaining two studies provided only a scoring system, which also permitted validation. Of the 19 models, nine had been only internally validated, whereas 10 had been externally (and, in some cases, also internally) validated.

3.2. Characteristics of the external validation samples

We included a total of 2,857 COVID-19 patients from the HM and ICS datasets (Table 2). On average, patients were aged more than 60 years (mean [standard deviation] 66.9 [15.8] years in HM; 63.6 [16.2] years in ICS), there were slightly more men (HM: 61%; ICS: 58%), and they were mostly nonsmokers (HM: 97%; ICS: 86%). A considerable proportion of patients had a recorded diagnosis of high blood pressure (HM: 35%; ICS: 49%), diabetes (HM: 16%; ICS: 21%), and/or obesity (HM 5%; ICS: 42%). At hospital admission, more than half of the participants reported dyspnea as the main (in HM) or as one of the symptoms (in ICS); and 7% and 70% reported cough as the main (in HM) or as one of the symptoms (in ICS), respectively. By the end of follow-up, in the HM and ICS datasets, respectively, roughly 7% and 28% of the patients received some form of mechanical ventilation; 4% and 18% were admitted to the ICU; and a 16% and 13% died. In the crude analysis (Supplementary Material, Item 9), an increased risk of 28-day mortality was observed among participants who were older, male, had an underlying comorbidity, high respiratory rate, low oxygen saturation, and/or altered laboratory results at hospital admission.

3.3. External validation of the prognostic models

Due to the unavailability of some predictors in the validation datasets (Supplementary Material, Item 10), we validated 18 (of 19) models: four using both datasets, 13 using one of the datasets (seven in HM and six in ICS), and one using only the complete case ICS dataset.

Nine of 17 models showed a good discrimination capacity (AUC $\geq$ 80%) (Table 3, Supplementary Material, Item 11). All these models included age among their predictors, six included oxygen saturation, and seven included at least one laboratory variable obtained at hospital admission. In both datasets, discrimination was higher in models predicting mortality (AUCs ranging from 65% to 87%) than in those predicting ICU admission or a composite outcome (AUCs ranging from 53% to 78%).

All models estimating outcome's probabilities exhibited an overall poor calibration (Supplementary Material, Item 12), with some models showing a general underestimation of the risk of critical COVID-19 and others showing an overestimation at small probabilities and underestimation at large probabilities. Of the eight models using a point-based scoring system, four showed a good calibration (Figs. 2 and 3). These models (Torres Macho [20], Li [26], 4C Mortality [13], and Altschul [18]) showed good discrimination, used mortality as the outcome variable,

Table 1. Characteristics of the 19 identified prognostic models

Model	Country	Events (sample size)[a]	Age, mean, or median (years)[a]	Sex, % of male[a]
Zhao—ICU	USA	195 (454)	60	60%
Zhao—mortality		82 (454)		
Weng	China	21 (176)	47	42%
Torres-Macho	Spain	325 (1,968)	67	56%
Mei—nonlaboratory	China	103 (1,088)	58	50%
Mei—laboratory				
Wang—nonlaboratory	China	19 (296)	47	47%
Wang—laboratory				
Berzuini—nonlaboratory	England	110 (392)	71	65%
Berzuini—laboratory				
Gude	Spain	51 (229)	68	61%
Zhou	China	68 (763)	51	49%
Li—ICU	USA	271 (1,108)	ICU: 59, non-ICU admitted: 62	57%
Li—mortality		142 (1,022)	Death: 76, alive: 59	57%
Altschul	USA	621 (2,355)	65	47%
Chen	China	50 (1,590)	Death: 69, alive: 48	Death: 60%, alive: 57%
Baronio	Italy	157 (191)	65	78%
4C Mortality	England, Scotland, Wales	11,426 (35,463)	73	58%
Magro	Italy	495 (1,810)	67	71%

AST, aspartate aminotransferase; BUN, blood urea nitrogen; CKD, chronic kidney disease; CRP, C-reactive protein; COPD, chronic obstructive pulmonary disease; COVID-19, coronavirus 2019; GFR, glomerular filtration rate; HR, hazard ratio; ICU, intensive care unit; IMV, invasive mechanical ventilation; LASSO, least absolute shrinkage and selection operator; LDH, lactate dehydrogenize; MAP, mean arterial pressure; NLR, neutrophil/lymphocyte ratio; OR, odds ratio; PCT, procalcitonin; SpO_2, oxygen saturation; USA, United States of America; WBC, white blood cell count.

[a] Described for the total sample. In articles performing an external validation, these parameters are described for the derivation cohort.
[b] Day of diagnosis considered day 1.
[c] Severe COVID-19 defined as ICU admission, IMV, and/or all-cause mortality.
[d] Calculated as (prothrombin time/normal range of prothrombin time) international sensitivity index.
[e] We validated the model using 28-day survival.

and included age, oxygen saturation, and C-reactive protein among their predictors (Table 4).

3.4. Secondary and sensitivity analysis

The validation of all models against 28-day mortality as a unique outcome showed a 1% to 2% increase in the AUCs of most models predicting mortality and a more than 10% increase in the AUCs of those predicting ICU admission (Supplementary Material, Item 13). The validation of the models in the complete case datasets (including one additional model not validated in the multiple imputed datasets [21]), resulted in slightly higher AUCs in the HM dataset (0% to 10% more) and in slightly lower or higher AUCs in the ICS dataset (3% to 4%) (Supplementary Material, Item 14). After excluding participants who were transferred to another hospital, there were no substantial changes in the discrimination ability of the models (Supplementary Material, Item 15). The four models highlighted above—Torres Macho [20], Li [26], 4C Mortality [13], and Altschul [18]—exhibited very similar validation properties in the sensitivity analyses.

4. Discussion

This systematic review identified 19 models aimed at predicting critical COVID-19 in hospitalized individuals with the disease and validated 18 of them separately in datasets obtained from public and private hospitals. Main results showed that (1) half of the validated models presented good discrimination, all using mortality as the outcome; (2) of those, only four models had a good calibration, and these were all based on a scoring system (not providing outcome's probabilities); and (3) the best performing models included parameters easy to obtain in clinical settings.

4.1. Performance of prognostic models during the external validation

In the present study, half of the models estimating COVID-19 mortality showed good discrimination. These

Predictors	Outcome	Statistical method/validated parameters
Smoking status, SpO_2, LDH, lymphocyte count, PCT	ICU admission	Logistic/risk score points
Age, COPD, heart failure, heart rate, SpO_2, LDH, PCT	In-hospital mortality	
Age, CRP, D-dimer, NLR	All-cause mortality	Logistic/intercept, beta coefficients
Age, smoking status, SpO_2, creatinine, CRP, hemoglobin, lymphocyte count, platelet count, sodium	In-hospital mortality	Logistic/risk score points
Age, respiratory failure, renal failure, coronary heart disease, heart failure, interaction age-renal failure	60-day all-cause mortality[b]	Logistic/intercept, beta coefficients
Age, respiratory failure, D-dimer, LDH, lymphocyte count, platelet count, WBC, interaction WBC platelets, interaction D-dimer-LDH		
Age, coronary heart disease, hypertension	In-hospital mortality	Logistic/OR
Age, SpO_2, AST, CRP, D-dimer, GFR, lymphocyte count, neutrophil count		
Age, smoking status, respiratory rate, SpO_2	21-day mortality[b]	Logistic/intercept, beta coefficients
Age, smoking status, respiratory rate, SpO_2, NLR		
Age, diabetes, SpO_2, lymphocyte count, pH	Severe COVID-19[c]	Logistic/intercept, beta coefficients
Age, smoking status, CKD, respiratory rate, systolic blood pressure, fever	ICU admission	Logistic/risk score points
SpO_2, CRP, ferritin, LDH, PCT	ICU admission	Deep learning neural network/risk score points
Age, SpO_2, CRP, LDH, PCT, troponin	In-hospital mortality	
Age, MAP, SpO_2, BUN, CRP, international normalized ratio[d]	In-hospital mortality	Logistic/risk score points
Age, cerebrovascular disease, coronary heart disease, AST, PCT, dyspnoea	14-day, 21-day, and 28-day survival[e]	Cox/HR
Age, BMI, cardiovascular disease, SpO_2, lymphocyte count	ICU admission	Logistic/OR
Age, sex, number of comorbidities, respiratory rate, SpO_2, Glasgow Coma Scale, BUN, CRP	In-hospital mortality	LASSO/risk score points
Age, sex, chronic liver disease, coronary heart disease, diabetes, LDH, duration of symptoms before hospital admission	40-day in-hospital mortality	Competing risks/none

results are in contrast with most findings in the field of predictive modeling, where the discrimination of a model substantially decreases when being validated in an external sample [27]. Indeed, our results confirm the importance of conducting a systematic review and of including strict selection criteria to identify the most adequate models, prior to their validation. Previous studies have validated several critical COVID-19 prognostic models [28–30], but compared to our study, they have found worse discrimination results. This could be explained by the fact that they have selected the models using more permissive criteria and have used smaller sample sizes for the validation (from 169 to 1,612). The lower discrimination capacity of models using ICU admission or composite outcomes to define critical COVID-19, compared to those using mortality, is in line with other reports [29,31] and might be explained by the context of the pandemic itself, where the criteria to be admitted to an ICU may have been modified due to an increased demand or to the unavailability of hospital resources [32]. Finally, our finding of poor calibration in all models providing an outcome's probability (not going through a calibration in the large) agrees with the idea that prognostic models should recalibrate the intercept based on the frequency of the outcome in the setting where they are to be applied [33]. Overall, our study went a step further by providing a list of models that are applicable to the clinical settings and by identifying four models exhibiting the best discrimination and calibration properties when validated externally in two diverse and large datasets.

4.2. *Components of the well-performing prognostic models*

We identified age, oxygen saturation, and laboratory indicators (C-reactive protein, lactate dehydrogenize, and lymphocyte count) as variables consistently included in the models with best prediction properties for critical COVID-19, in agreement with previous literature identifying them as individual predictors of critical COVID-19 [32,34–37]. It is important to highlight that these indicators are easy and rapid to obtain, something that enhances the clinical applicability of the models. Some of the studies

Table 2. Characteristics of the study population according to study dataset

	HM dataset (N = 1,753)		ICS dataset (N = 1,104)	
Characteristics of the Study Population	***N***		***N***	
Age, years — mean (SD)	1,753	66.9 (15.8)	1,104	63.6 (16.2)
Male gender, *n* (%)	1,753	1,064 (60.7%)	1,104	639 (57.9%)
Current smoker, *n* (%)	1,502	47 (3.13%)	902	124 (13.7%)
BMI, kg/m^2—mean (SD)	n.a.	n.a.	373	29.5 (5.52)
Presence of comorbidity, *n* (%)				
Diabetes Mellitus	1,502	239 (15.9%)	1,104	228 (20.7%)
COPD	1,502	72 (4.79%)	1,104	89 (8.06%)
Asthma	1,502	69 (4.59%)	1,104	80 (7.25%)
Obesity	1,502	77 (5.13%)	1,104	459 (41.6%)
Hypertension	1,502	525 (35.0%)	1,104	535 (48.5%)
Cardiovascular disease (any)	n.a.	n.a.	1,104	117 (10.6%)
Cerebrovascular disease	1,502	33 (2.20%)	1,104	45 (4.08%)
Heart failure	1,502	24 (1.60%)	1,104	55 (4.98%)
Coronary heart disease	1,502	82 (5.46%)	1,104	87 (7.88%)
Vital signs at admission — mean (SD) or median [IQR]				
Systolic blood pressure, mmHg	1,462	125 (19.4)	1,066	130 (20.2)
Diastolic blood pressure, mmHg	1,462	71.5 (12.4)	1,062	74.5 (13.9)
Temperature, °C	1,746	36.7 (0.86)	918	37.2 (1.00)
Heart rate, bpm	1,721	81.3 (14.4)	1,066	91.8 (17.8)
Respiratory rate, breath/min	n.a.	n.a.	981	20.0 [18.0–25.0]
SpO_2, %	1,489	94.0 [91.0-96.0]	1,030	96.0 [94.0–98.0]
Clinical scores at admission — mean (SD) or median [IQR]				
Glasgow coma scale, units	n.a.	n.a.	899	15.0 [15.0–15.0]
Laboratory parameters at admission — mean (SD) or median [IQR]				
D-dimer, mg/L	1,325	2.17 (7.10)	491	1.69 (16.9)
Platelet count, 10^9/L	1,605	220 (94.6)	642	207 (85.6)
International normalized ratio[a]	1,305	1.20 [1.12–1.31]	597	1.11 [1.04–1.21]
Arterial O_2, mmHg	599	59.0 [51.0–67.0]	175	75.0 [60.0–100]
Arterial CO_2, mmHg	599	34.3 (6.57)	175	66.7 (45.2)
WBC, 10^9/L	1,605	6.47 [4.93–8.59]	575	6.72 [5.12–8.89]
Lymphocyte count, 10^9/L	1,605	1.02 [0.73–1.41]	554	1.00 [0.70–1.40]
Neutrophil count, 10^9/L	1,605	4.68 [3.26–6.73]	548	4.90 [3.48–7.00]
NLR	1,605	6.85 (11.1)	547	7.01 (7.75)
Creatinine, mg/L	1,570	0.90 [0.72–1.10]	643	0.85 [0.68–1.07]
BUN, mmol/L	1,538	5.61 [4.20–7.91]	509	6.30 [4.33–9.00]
GFR, mL/min	5	25.1 [19.7–27.9]	575	86.0 [61.0–90.0]
AST, U/L	1,498	34.0 [24.0-51.5]	621	41.0 [30.0–60.0]
CRP, mg/L	1,573	73.2 [31.1–138]	445	12.4 [5.86–24.5]
PCT, ng/mL	142	0.14 [0.09–0.27]	42	0.21 [0.14–0.32]
Troponin, ng/L	272	11.2 [7.08–29.5]	180	10.0 [8.43–20.0]
LDH, U/L	1,527	541 [427–698]	363	352 [288–453]
Sodium, mmol/L	1,552	137 (4.55)	644	136 (3.56)
Ferritin, ng/mL— median [IQR]	373	936 [426–1,553]	322	582 [315–1,089]
pH, units— mean (SD)	788	7.45 (0.06)	284	7.42 (0.08)
Symptoms at admission, *n* (%)				
Main symptom				

(*Continued*)

Table 2. Continued

Characteristics of the Study Population	HM dataset (N = 1,753)		ICS dataset (N = 1,104)	
	N		*N*	
Dyspnoea	1,714	940 (54.8%)	n.a.	n.a.
Fever	1,714	231 (13.5%)	n.a.	n.a.
Cough	1,714	123 (7.18%)	n.a.	*n*.a.
Presence of symptoms				
Dyspnoea	n.a.	n.a.	876	440 (50.2%)
Fever	n.a.	n.a.	1,023	794 (77.6%)
Cough	n.a.	n.a.	876	642 (73.3%)
Complications, *n* (%)				
NIMV	1,502	66 (4.39%)	802	103 (12.8%)
IMV	1,502	33 (2.20%)	802	121 (15.1%)
ICU admission	1,753	77 (4.39%)	1,104	205 (18.6%)
In-hospital mortality	1,753	278 (15.9%)	1,104	145 (13.1%)
ICU admission and/or death	1,753	319 (18.2%)	1,104	325 (29.4%)
Time from hospital admission to — mean (SD)				
Discharge/death, days	1,753	8.39 (5.62)	1,104	10.7 (12.9)
ICU admission, days	77	5.60 (5.99)	203	4.57 (6.78)

AST, aspartate aminotransferase; BUN, blood urea nitrogen; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; COVID-19, coronavirus disease 2019; CRP, C-reactive protein; GFR, glomerular filtration rate; ICU, intensive care unit; IMV, invasive mechanical ventilation; LDH, lactate dehydrogenase; NIVM, noninvasive mechanical ventilation; NLR, neutrophil/lymphocyte ratio; SpO_2, oxygen saturation; WBC, white blood cell count.

[a] Calculated as (prothrombin time/normal range of prothrombin time) international sensitivity index.

Table 3. Discriminatory ability of 16 critical COVID-19 prognostic models in two external validation samples

Critical COVID-19 Prognostic Models	HM dataset	ICS dataset
	AUC (95% CI)	AUC (95% CI)
Outcome: In-hospital mortality		
Zhao — mortality	0.78 (0.75 to 0.81)	-
Weng	0.84 (0.81 to 0.86)	0.78 (0.74 to 0.82)
Torres-Macho	0.87 (0.85 to 0.89)	0.82 (0.79 to 0.86)
Mei — laboratory	0.80 (0.77 to 0.84)	-
Wang — nonlaboratory	0.83 (0.80 to 0.85)	0.81 (0.78 to 0.84)
Wang — laboratory	-	0.65 (0.59 to 0.70)
Berzuini — nonlaboratory	-	0.85 (0.82 to 0.88)
Berzuini — laboratory	-	0.86 (0.83 to 0.89)
Li — mortality	0.81 (0.77 to 0.85)	-
Altschul	0.83 (0.81 to 0.86)	0.82 (0.78 to 0.87)
Chen	0.73 (−1.48 to 2.93)[a]	-
4C Mortality	-	0.84 (0.81 to 0.87)
Outcome: ICU admission		
Zhao — ICU	0.64 (0.58 to 0.71)	-
Zhou	-	0.53 (0.49 to 0.59)
Li — ICU	0.59 (0.52 to 0.66)	-
Baronio	-	0.62 (0.55 to 0.69)
Outcome: composite		
Gude	0.78 (0.75 to 0.81)	-

ICU, intensive care unit; AUC, area under the curve.

[a] Harrell's C index.

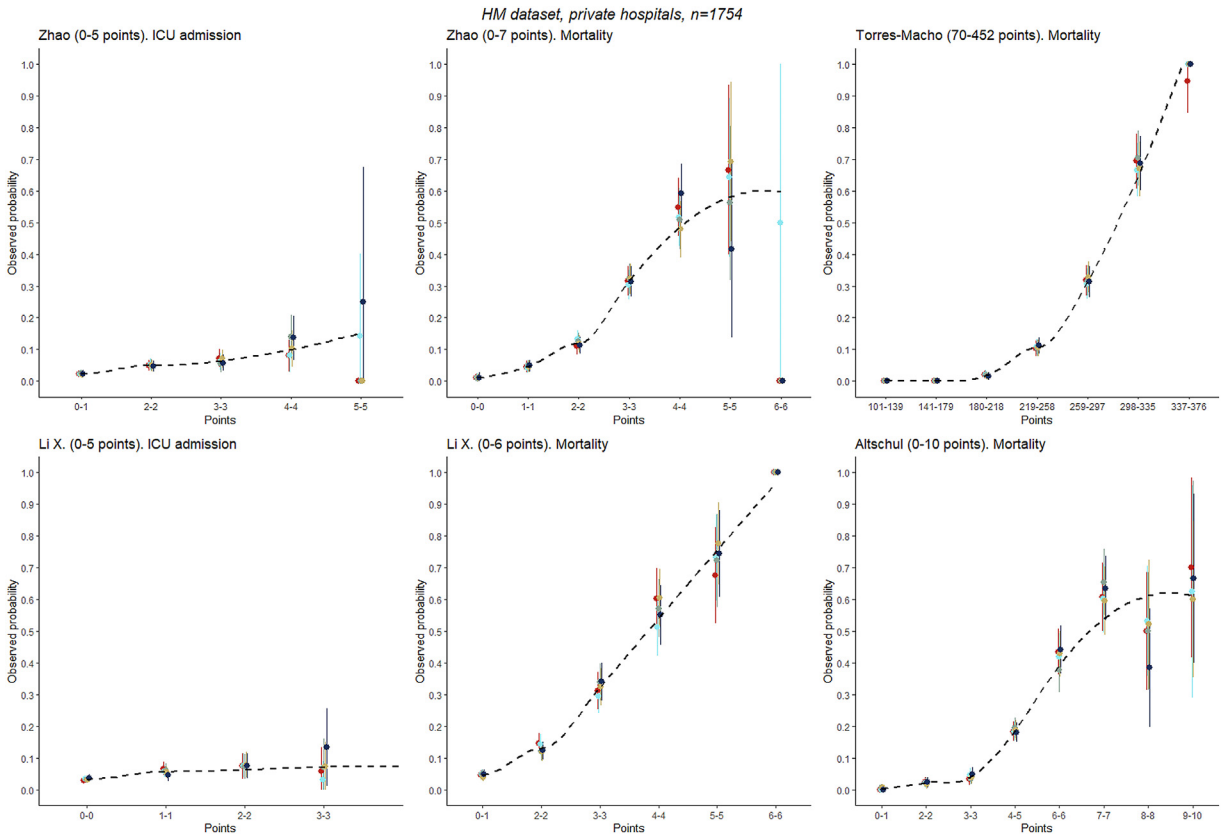

Fig. 2. Calibration of six critical COVID-19 prognostic models providing a point-based scoring system in the HM dataset. The dashed black line indicates the LOESS fit. The five imputed datasets are differentiated in colors.

developing prognostic models emphasized that laboratory results and vital signs (i.e., dynamic indicators of acute disease) exhibited better predictive abilities than chronic comorbidities or demographic data (i.e., fixed individual characteristics) [35,38]. This is of great importance at the moment of hospital admission, where diagnosis of comorbidities may not be available or accurate in electronic health records, and where interviewing compromised incoming patients may not be possible. In this context, two of our four best performing models, based only on age and parameters obtained from physical examination and laboratory analysis, might be particularly useful [18,26].

4.3. Clinical applicability of the identified models

Based on our results, half of the tested models are suitable for use in the general hospitalized population and for a prompt application at hospital admission, a key moment in decision-making. The selection of one of the four models with best prediction properties should be based mainly on the target population and on the availability of predictors (Table 4). In brief, the Torres-Macho score [20] is suitable for a broad population (as we found it valid in both private and public hospitals), is especially useful when a comprehensive—but still routinely collected—set of predictors including age, smoking status, and results from the physical examination and laboratory analysis are available (see full list of predictors in Table 4), and can be computed online (www.pandemyc-score.com). The Li [26] and the Altschul [18] scores include the smallest set of predictors (age, one to two clinical signs, and three to four laboratory indicators), making them especially useful at emergency settings; however, they do not provide an online application and, in the case of Li score, it was validated (and proved suitable) only in our private hospitals dataset. Finally, the application of the 4C Mortality score [13] needs a compressive set of predictors and is available as an online calculator (www.isaric4c.net/risk); however, in this study, its suitability was tested (and proved) only in our public hospitals dataset.

It could be argued whether the mentioned models, developed and validated in prevaccination conditions, can be applicable to current and future patients. Research has consistently reported that the severity of COVID-19 (and therefore, the proportion of cases with critical disease) was higher in the first than in subsequent waves, in alpha and beta than in following variants, and in unvaccinated than vaccinated patients [39–43]. However, the predictors of critical COVID-19 have been reported to be the same regardless of wave, variant, and vaccination status. Indeed, a recent study validating one prevaccination model found

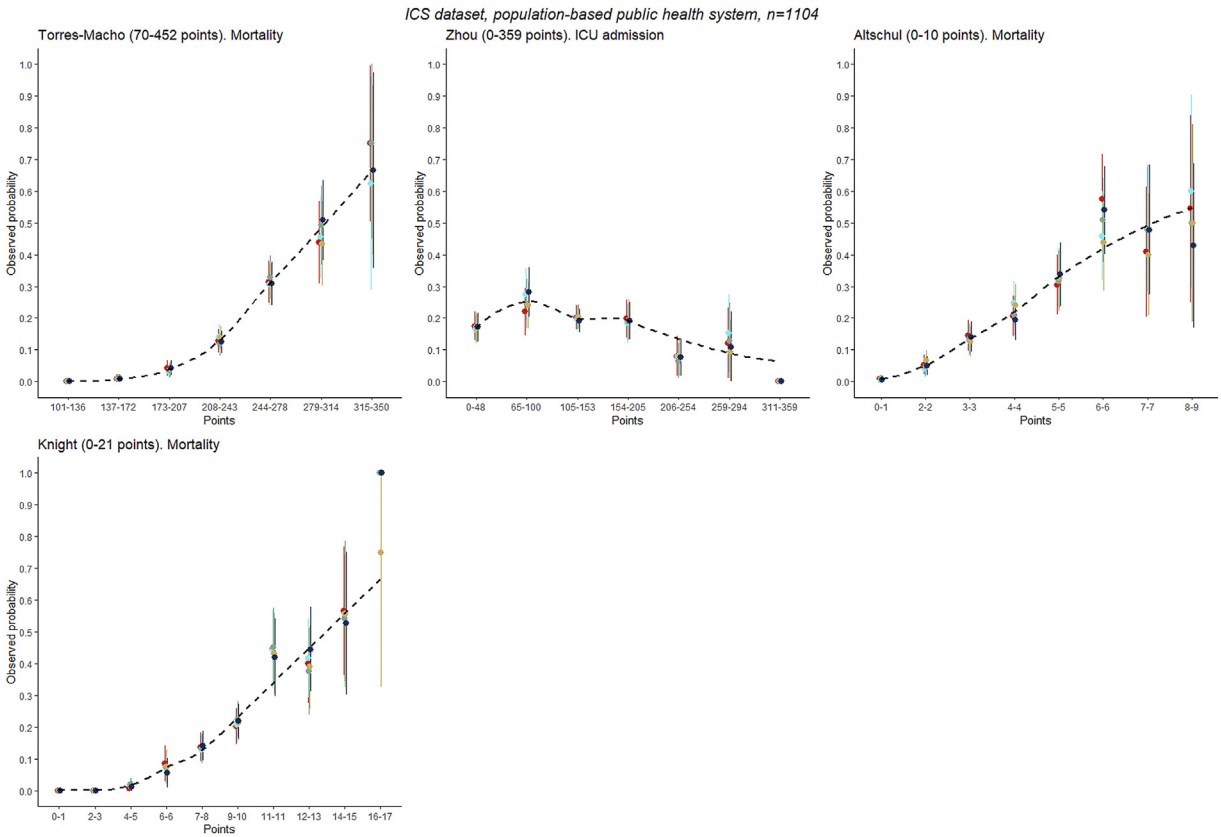

Fig. 3. Calibration of four critical COVID-19 prognostic models providing a point-based scoring system in the ICS dataset. The dashed black line indicates the LOESS fit. The five imputed datasets are differentiated in colors.

good performance in a sample of patients infected with delta and omicron variants [44], and a recently developed model including vaccination as a predictor did not include any interaction term between vaccine and remaining predictors [45]. All in all, data suggest that models developed prior to vaccination maintain their discrimination properties (i.e., the ability to order individuals as per their risk of critical COVID-19 when used for risk stratification) and would only need to be recalibrated when used to estimate the risk of critical COVID-19, as anyways recommended when using prognostic models as mentioned previously [33].

4.4. Applicability of study results to prognostic model research

In our systematic review, more than two-thirds of the studies that had developed a prognostic model did not report their results in adequate detail to allow individual prediction nor subsequent validation despite this is strongly recommended in Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis guidelines [10]. Incomplete reporting has been considered a form of research waste and, in the case of prognostic model research, precludes any further validation and use of research results. Our findings sustain the idea that, in the COVID-19 field, further efforts should be directed toward validating, recalibrating, and making available existing models rather than toward developing new ones.

4.5. Strengths and limitations

A strength of this study is the systematic identification of prognostic models and the focus on those applicable to the general population (not to specific subgroups) and on those using only routinely collected data. For the external validation, we used two different datasets (one from a public health system and one from a private hospitals network, thus representing different patients' background) with relatively high sample sizes, which broadens the applicability of our results. Additionally, the HM dataset was generated during the first COVID-19 epidemiological period (i.e., first wave), which was dominated by the original SARS-CoV-2 strain [46], whereas the ICS dataset included data not only from the first wave but also from the second, which in Spain was dominated by the B.1.177. variant [47]. This adds confidence regardless of the SARS-CoV-2 variant in circulation, in the predictive ability of models that performed well in the two validation datasets, such as the Torres-Macho score.

The main limitation of this study is that some predictors (e.g., signs/symptoms at hospital admission) were missing, which prevented us from validating some of the identified

Table 4. Components of the four prognostic models showing the best discrimination and calibration properties: Torres-Macho, Li, 4C Mortality, and Altschul mortality scores

Torres-Macho [20] (range 70–452)		**Li [26] (range 0–6)**	
Categories	**Pts**	**Categories**	**Pts**
Age, years		Age, years	
<46	−46	≤71	0
≥46 and ≤60	4	>71	1
>60 and ≤78	29	SpO_2, %	
>78	50	≥88	0
Missing	28	<88	1
Smoking status		CRP, mg/L	
No	25	≤17	0
Yes	43	>17	1
Missing	22	PCT, ng/mL	
SpO_2, %		≤1.1	0
<88	59	>1.1	1
≥88 and ≤92	36	Troponin, ng/mL	
>92 and ≤96	21	≤0.03	0
>96	−2	>0.03	1
Missing	30	LDH, U/L	
Platelet count, 10^9/L		≤487	0
<124	48	>487	1
≥124	27	BUN, mmol/L	
Missing	18	≤19.6	0
Hemoglobin, g/dL		>19.6 and ≤39.2	1
<11.2	51	>39.2	3
≥11.2 and ≤12.2	36		
>12.2	25		
Missing	20		
Lymphocyte count, cells/μL			
<500	55		
≥500 and ≤799	37		
>799 and ≤1,000	28		
>1,000	22		
Missing	19		
Creatinine, mg/dL			
<0.77	14		
≥0.77 and ≤1.08	21		
>1.08 and ≤1.71	37		
>1.71	48		
Missing	29		
CRP, mg/L			
<9.7	8		
≥9.7 and ≤65.9	20		
>65.9 and ≤188.5	35		
>188.5	45		
Missing	18		
Sodium, mmol/L			
<144	28		

(*Continued*)

Table 4. Continued

Torres-Macho [20] (range 70–452)		Li [26] (range 0–6)	
Categories	**Pts**	**Categories**	**Pts**
≥144	53		
Missing	17		
4C Mortality [13] (range 0–21)		Altschul [18] (range 0–10)	
Age, years		Age, years	
<50	0	<60	0
≥50 and <60	2	≥60 and <70	1
≥60 and <70	4	≥70 and <80	2
≥70 and <80	6	≥80	3
≥80	7	>80	0
Sex		Mean arterial pressure, mmHg	
Female	0	>70 and ≤80	1
Male	1	>60 and ≤70	2
Number of comorbidities		≤60	3
0	0	SpO_2, %	
1	1	≥94	0
≥2	2	<94	1
Respiratory rate, breath/min		International normalized ratio[a]	
<20	0	≤1.2	0
≥20 and <30	1	>1.2	1
≥30	2	BUN, mg/dl	
SpO_2, %		≤30	0
≥92	0	>30	1
<92	2	CRP, mg/L	
Glasgow Coma Scale		≤10	0
15	0	>10	1
<15	2		
BUN, mmol/L			
≤19.6	0		
>19.6 and ≤39.2	1		
>39.2	3		
CRP, mg/L			
<50	0		
≥50 and <100	1		
≥100	2		

BUN, blood urea nitrogen; CRP, C-reactive protein; LDH, lactate dehydrogenase; SpO_2, oxygen saturation; Pts, points.
[a] Calculated as (prothrombin time/normal range of prothrombin time) international sensitivity index.

models in both datasets. However, it is likely that these predictors are not available in most clinical settings, something that hinders the clinical applicability of models including them and support the idea that future studies should focus only on models including easily available predictors. Another limitation was the retrospective nature of our validation cohorts, which increases the chances of bias (e.g., due to a non-standardized data collection and data entry procedure).

4.6. Conclusion

The validity of models predicting critical COVID-19 in the general population by using only routinely collected predictors is variable, and only half of them showed good discrimination when externally validated. Four models (Torres Macho [20], Li [26], 4C Mortality [13], and Altschul [18]) exhibiting good discrimination and calibration included routinely collected patient's parameters and therefore are proposed for their use in clinical settings.

Acknowledgments

This study was carried out using anonymized data provided by the Catalan Agency for Quality and Health

Assessment, within the framework of the PADRIS Program.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2023.04.011.

References

[1] Islam N, Shkolnikov VM, Acosta RJ, Klimkin I, Kawachi I, Irizarry RA, et al. Excess deaths associated with covid-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries. BMJ 2021;373:n1137.

[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 72314 cases from the Chinese center for disease control and prevention. JAMA 2020;323:1239−42.

[3] European centre for disease prevention and control. COVID-19 situation update worldwide, as of week 14, updated 14 April 2022. 2020. Available at: https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases.

[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: systematic review and critical appraisal. BMJ 2020;369:m1328.

[5] Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019;170:W1−33.

[6] Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: elaboration and explanation. BMJ 2015;349:1−25.

[7] Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015 statement. Syst Rev 2015;4:1.

[8] Riley RD, Moons KGM, Snell KIE, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. BMJ 2019;364:k4597.

[9] Kohl C, McIntosh EJ, Unger S, Haddaway NR, Kecke S, Schiemann J, et al. Correction to: online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. Environ Evid 2018;7(8). https://doi.org/10.1186/s13750-018-0124-4.

[10] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:1−9.

[11] Centro Nacional de Epidemiología, Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación. Informe no112. Situación de COVID-19 en España a 5 de enero de 2022. Equipo COVID-19. 2022. Available at: https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Documents/INFORMES/Informes.

[12] World Health Organization (WHO). Living guidance for clinical management of COVID-19. World Heal Organ; 2021.

[13] 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.

[14] Zhao Z, Chen A, Hou W, Graham JM, Li H, Richman PS, et al. Prediction model and risk scores of ICU admission and mortality in COVID-19. PLoS One 2020;15:e0236618.

[15] Zheng Y, Zhang Y, Chi H, Chen S, Peng M, Luo L, et al. The hemocyte counts as a potential biomarker for predicting disease progression in COVID-19: a retrospective study. Clin Chem Lab Med 2020;58:1106−15.

[16] Baronio M, Freni-Sterrantino A, Pinelli M, Natalini G, Tonini G, Marri M, et al. Italian SARS-CoV-2 patients in intensive care: towards an identikit for subjects at risk? Eur Rev Med Pharmacol Sci 2020;24:9698−704.

[17] Magro B, Zuccaro V, Novelli L, Zileri L, Celsa C, Raimondi F, et al. Predicting in-hospital mortality from coronavirus disease 2019: a simple validated app for clinical use. PLoS One 2021;16:e0245281.

[18] Altschul DJ, Unda SR, Benton J, de la Garza Ramos R, Cezayirli P, Mehler M, et al. A novel severity score to predict inpatient mortality in COVID-19 patients. Sci Rep 2020;10:16726.

[19] Weng Z, Chen Q, Li S, Li H, Zhang Q, Lu S, et al. ANDC: an early warning score to predict mortality risk for patients with coronavirus disease 2019. J Transl Med 2020;18:328.

[20] Torres-Macho J, Ryan P, Valencia J, Pérez-Butragueño M, Jiménez E, Fontán-Vela M, et al. The PANDEMYC score. An easily applicable and interpretable model for predicting mortality associated with COVID-19. J Clin Med 2020;9:3066.

[21] Mei J, Hu W, Chen Q, Li C, Chen Z, Fan Y, et al. Development and external validation of a COVID-19 mortality risk prediction algorithm: a multicentre retrospective cohort study. BMJ Open 2020;10:e044028.

[22] Wang K, Zuo P, Liu Y, Zhang M, Zhao X, Xie S, et al. Clinical and laboratory predictors of in-hospital mortality in patients with COVID-19: a cohort study in Wuhan, China. Clin Infect Dis 2020;71:2079−88.

[23] Berzuini C, Hannan C, King A, Vail A, O'Leary C, Brough D, et al. Value of dynamic clinical and biomarker data for mortality risk prediction in COVID-19: a multicentre retrospective cohort study. BMJ Open 2020;10:e041983.

[24] Gude F, Riveiro V, Rodríguez-Núñez N, Ricoy J, Lado-Baleato Ó, Lourido T, et al. Development and validation of a clinical score to estimate progression to severe or critical state in COVID-19 pneumonia hospitalized patients. Sci Rep 2020;10:19794.

[25] Zhou Y, He Y, Yang H, Yu H, Wang T, Chen Z, et al. Exploiting an early warning Nomogram for predicting the risk of ICU admission in patients with COVID-19: a multi-center study in China. Scand J Trauma Resusc Emerg Med 2020;28:106.

[26] Li X, Ge P, Zhu J, Li H, Graham J, Singer A, et al. Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables. PeerJ 2020;8:e10337.

[27] Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol 2015;68:25−34.

[28] 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;56:2003498.

[29] Wickstrøm KE, Vitelli V, Carr E, Holten AR, Bendayan R, Reiner AH, 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.

[30] Yamada G, Hayakawa K, Asai Y, Matsunaga N. External validation and update of prediction models for unfavorable outcomes in hospitalized patients with COVID-19 in Japan. J Infect Chemother 2021;27:1043−50.

[31] Xu Y, Becker N, Ferres J, Lee A. Machine learning-based derivation and external validation of a tool to predict death and development of organ failure in hospitalized patients with COVID-19. Sci Rep 2022;12:16913.

[32] Kim HJ, Hwang H, Hong H, Yim JJ, Lee J. A systematic review and meta-analysis of regional risk factors for critical outcomes of

COVID-19 during early phase of the pandemic. Sci Rep 2021;11:9784.

[33] Janssen KJM, Moons KGM, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol 2008;61:76−86.

[34] dong GY, Ding M, Dong X, jin ZJ, Kursat Azkur A, Azkur D, et al. Risk factors for severe and critically ill COVID-19 patients: a review. Allergy 2021;76:428−55.

[35] Incerti D, Rizzo S, Li X, Lindsay L, Yau V, Keebler D, et al. Prognostic model to identify and quantify risk factors for mortality among hospitalised patients with COVID-19 in the USA. BMJ Open 2021;11:1−11.

[36] Meintrup D, Borgmann S, Seidl K, Stecher M, Jakob CEM, Pilgram L, et al. Specific risk factors for fatal outcome in critically ill covid-19 patients: results from a European multicenter study. J Clin Med 2021;10:1−14.

[37] Shcherbak SG, Anisenkova AY, Mosenko SV, Glotov OS, Chernov AN, Apalko SV, et al. Basic predictive risk factors for cytokine storms in COVID-19 patients. Front Immunol 2021;12:1−8.

[38] Adderley NJ, Taverner T, Price MJ, Sainsbury C, Greenwood D, Chandan JS, et al. Development and external validation of prognostic models for COVID-19 to support risk stratification in secondary care. BMJ Open 2022;12:e049506.

[39] Buttenschøn HN, Lynggaard V, Sandbøl SG, Glassou ENHA. Comparison of the clinical presentation across two waves of COVID-19: a retrospective cohort study. BMC Infect Dis 2022;22:493.

[40] Robinson ML, Morris CP, Betz JF, Zhang Y, Wang N, Thiemann DR, et al. Impact of severe acute respiratory syndrome coronavirus 2 variants on inpatient clinical outcome. Clin Infect Dis 2023;76(9):1539−49.

[41] Bieńkowski C, Kowalska JD, Paciorek M, Wasilewski P, Uliczny P, Garbacz-Łagożna E, et al. The clinical course and outcomes of patients hospitalized due to COVID-19 during three pandemic waves in Poland: a single center observational study. J Clin Med 2022;11:1−13.

[42] Butt AA. Severity of illness in persons infected with the SARS-CoV-2 delta variant vs beta variant in Qatar. JAMA Intern Med 2022;182:1−9.

[43] Vito A De, Colpani A, Saderi L, Puci M, Zauli B, Fiore V, et al. Impact of early SARS-CoV-2 antiviral therapy on disease progression. Viruses 2023;15:71.

[44] Wei S, Ong X, Tham SM, Koh LP, Dugan C, Khoo BY, et al. External validation of the PRIORITY model in predicting COVID-19 critical illness in vaccinated and unvaccinated patients. Clin Microbiol Infect 2022;28:884.e1−3.

[45] Zhang Y, Han J, Sun F, Guo Y, Guo Y, Zhu H, et al. A practical scoring model to predict the occurrence of critical illness in hospitalized patients with SARS-CoV-2 omicron infection. Front Microbiol 2022;13:1−10.

[46] López M, Chiner-Oms Á, García de Viedma D, Ruiz-Rodriguez P, Bracho M, Cancino-Muñoz I, et al. The first wave of the COVID-19 epidemic in Spain was associated with early introductions and fast spread of a dominating genetic variant. Nat Genet 2021;53:1405−14.

[47] González-Candelas F, Shaw MA, Phan T, Kulkarni-Kale U, Paraskevis D, Luciani F, et al. One year into the pandemic: short-term evolution of SARS-CoV-2 and emergence of new lineages. Infect Genet Evol 2021;92:104869.