ABC_2-SPH risk score for in-hospital mortality in COVID-19 patients: development, external validation and comparison with other available scores

Authors: Milena S. Marcolino PhD[1*] (Marcolino MS - milenamarc@ufmg.br, 0000-0003-4278-3771)

Magda C. Pires PhD[2] (Pires MC – magda@est.ufmg.br, 0000-0003-3312-4002)

Lucas Emanuel F. Ramos[3] (Ramos LEF - luckermos19@gmail.com, 0000-0001-7844-0581)

Rafael T. Silva[3] (Silva RT - rafaelsilva@posteo.net, 0000-0002-9270-5328)

Luana M. Oliveira MSc[4] (Oliveira LM - luanalmo19.09@gmail.com, 0000-0003-4639-4546)

Rafael L.R. Carvalho PhD[5] (Carvalho RLR - rafaelsjdr@hotmail.com, 0000-0003-3576-3748)

Rodolfo L.S. Mourato[6] (Mourato RLS - rodolfo_use@hotmail.com, 0000-0002-0251-0691)

Adrián Sánchez-Montalvá PhD[7] (Montalvá AS - adrian.sanchez.montalva@gmail.com, 0000-0002-2194-5447)

Berta Raventós MSc[8] (Raventós B - berta.raventos@vhir.org, 0000-0002-4668-2970)

Fernando Anschau PhD[9] (Anschau F - afernando@ghc.com.br, 0000-0002-2657-5406)

José Miguel Chatkin PhD[10] (Chatkin JM - jmchatkin@pucrs.br, 0000-0002-4343-025X)

Matheus C. A. Nogueira[11] (Nogueira MCA - mathnogueira42@gmail.com, 0000-0002-0241-9046)

Milton H. Guimarães Júnior MSc[12] (Guimarães Júnior MH - miltonhenriques@yahoo.com.br, 0000-0002-2127-8015)

Giovanna G. Vietta PhD[13] (Vietta GG - ggvietta@gmail.com, 0000-0002-0756-3098)

Helena Duani PhD[14] (Duani H - hduani@yahoo.com.br, 0000-0001-9345-018X)

Daniela Ponce PhD[15] (Ponce D - daniela.ponce@unesp.br, 0000-0002-6178-6938)

Patricia K. Ziegelmann PhD[16] (Ziegelmann PK - patriciakz99@gmail.com, 0000-0002-2851-2011)

Luís C. Castro PhD [17] (Castro LC - pharmlucamsc@gmail.com, 0000-0003-2379-0167)

Karen B. Ruschel PhD[18] (Ruschel KB - karenbruschel@gmail.com, 0000-0002-6362-1889)

Christiane C. R. Cimini MSc[19] (Cimini CCR - christiane.cimini@gmail.com, 0000-0002-1973-1343)

Saionara C. Francisco MSc[20] (Francisco SCF - saionaracf@gmail.com, 0000-0002-9655-6294)

Maiara A. Floriani MSc[21] (Floriani AM - maiara.floriani@hmv.org.br, 0000-0002-2981-9445)

Guilherme F. Nascimento MSc [22] (Nascimento GF - guilhermefagundesn@hotmail.com, 0000-0001-9064-7067)

Bárbara L. Farace[23] (Farace BL - barbarafarace@gmail.com, 0000-0002-6172-1093)

Luanna S. Monteiro[24] (Monteiro LS - luannasmonteiro@gmail.com, 0000-0002-6621-3338)

Maira V. R. Souza-Silva MD[25] (Souza-Silva MVR - mairavsouza@gmail.com, 0000-0003-2079-7291)

Thais L. S. Sales MSc [26] (Sales TLS - thaislorennass30@yahoo.com.br, 0000-0002-1571-3850)

Karina Paula M. P. Martins MSc[27] (Martins KPMP - kkpmprado2@gmail.com, 0000-0002-8313-7429)

Israel J. Borges do Nascimento[28] (Borges do Nascimento IJ - israeljbn@ufmg.br, 0000-0001-5240-0493)

Tatiani O. Fereguetti[29] (Fereguetti TO - tatianifereguetti@gmail.com, 0000-0001-5845-0715)

Daniel T. M. O. Ferrara[30] (Ferrara DTMO - daniel@taiar.com.br, 0000-0003-0886-9627)

Fernando A. Botoni[31] (Botoni FA - fbotoni@medicina.ufmg.br, 0000-0001-6268-8507)

Ana Paula Beck da Silva Etges[32] (Etges APBS - anabsetges@gmail.com, 0000-0002-6411-3480)

Eric Boersma PhD[33] (Boersma E - h.boersma@erasmusmc.nl, 0000-0002-2559-7128)

Carisi A. Polanczyk PhD[34] (Polanczyk CA - carisi.anne@gmail.com, 0000-0002-2447-2577)

Brazilian COVID-19 Registry Investigators**

Affiliations, positions and addresses:

[1] Associate Professor and Internal Medicine Physician. Department of Internal Medicine, Medical School; and Telehealth Center, University Hospital, Universidade Federal de Minas Gerais. Avenida Professor Alfredo Balena 190 sala 246, Belo

Horizonte, Brazil. Researcher. Institute for Health Technology Assessment (IATS/ CNPq). Rua Ramiro Barcelos, 2359. Prédio 21 | Sala 507, Porto Alegre, Brazil.

[2] Associate Professor and Statistician, Department of Statistics, Universidade Federal de Minas Gerais. Av. Presidente Antônio Carlos, 6627, ICEx, sala 4071, Belo Horizonte, Brazil

[3] Undergraduate Statistics Students, Universidade Federal de Minas Gerais. Av. Presidente Antônio Carlos, 6627, Belo Horizonte, Brazil

[4] Ph Student in Business and Administration. Center for Research and Graduate Studies in Business Administration, Universidade Federal de Minas Gerais. Av. Presidente Antônio Carlos, 6627, Belo Horizonte, Brazil. Researcher. Institute for Health Technology Assessment (IATS/ CNPq). Rua Ramiro Barcelos, 2359. Prédio 21 | Sala 507, Porto Alegre, Brazil.

[5] Nurse and Researcher. Institute for Health Technology Assessment (IATS/ CNPq). Rua Ramiro Barcelos, 2359. Prédio 21 | Sala 507, Porto Alegre, Brazil.

[6] Universidade Federal de São João del-Rei. R. Sebastião Gonçalves Coelho, 400, Divinópolis, Brazil.

[7] Physician and Researcher. Infectious Diseases Department, Vall d'Hebron University Hospita, Universitat Autònoma de Barcelona, PROSICS, Barcelona, Spain. Passeig de la Vall d'Hebron, 119-129, Edificio Mediterránea, despacho 119, 08035, Barcelona, Spain.

[8] Infectious Diseases Department, Vall d'Hebron University Hospital, Universitat Autònoma de Barcelona, Barcelona, Spain

[9] Coordinator of the Research Sector of Grupo Hospitalar Conceição. Professor of the Graduation Program on Evaluation and Production of Technologies for the Brazilian National Health System, Hospital Nossa Senhora da Conceição and Hospital Cristo Redentor. Av. Francisco Trein, 326, Porto Alegre, Brazil.

[10] Professor. School of Medicine, Pontifícia Universidade Católica do Rio Grande do Sul (RGS), Porto Alegre, Brazil; Head in Pneumology Department, Hospital São Lucas PUCRS, Porto Alegre, Brazil. Rua João Cateano, 79/503. Porto Alegre, Brazil.

[11] Physician and Researcher. Internal Medicine Department, Rede Mater Dei de Saúde. Av. do Contorno, 9000, Belo Horizonte, Brazil.

[12] Teaching and Research Coordinator, Hospital Marcio Cunha. Av. Tsunawaki Avenue, 41, Ipatinga, Brasil.

[13] Professor and epidemiologist, School of Medicine. Universidade do Sul de Santa Catarina (UNISUL). Avenida Pedra Branca, 25, Cidade Universitária Pedra Branca, Palhoça, Brazil. Epidemiologist, Dissertare Scientific Advice, Rodovia João Paulo, 1030, Lá opera, 401b, Florianópolis, Brazil. Collaborating researcher, SOS Cardio Hospital, Rodovia, SC-401, 121, Florianópolis, Brazil.

[14] Professor and Infectious Diseases Physician. Internal Medicine Departament. University Hospital, Universidade Federal de Minas Gerais. Av. Prof Alfredo Balena, 110, Belo Horizonte, Brazil.

[15] Professor and Nephrologist. Faculdade de Medicina de Botucatu - Universidade Estadual Paulista "Júlio de Mesquita Filho". Av. Prof. Mário Rubens Guimarães Montenegro, s/n - UNESP - Campus de Botucatu, Botucatu, Brazil.

[16] Professor and Statistician. Universidade Federal do Rio Grande do Sul and Institute for Health Technology Assessment (IATS/ CNPq). Hospital Tacchini. Rua Ramiro Barcelos, 2359. Prédio 21 | Sala 507, Porto Alegre, Brazil.

[17] Clinical Pharmacist - UDM Coordinator of Specialized Assistance Service. Pharmaceutical Assistance Research Center of Vale do Taquari. Av. Rio Branco, 1127, Estrela, Brazil.

[18] Researcher. Hospital Mãe de Deus, Hospital Universitário de Canoas, Universidade Federal do Rio Grande do Sul and Institute for Health Technology Assessment (IATS/ CNPq). Rua Ramiro Barcelos, 2359. Prédio 21 | Sala 507, Porto Alegre, Brazil.

[19] Adults' Intensive Care Physician. Adjunct Professor. Mucuri Medical School – FAMMUC, Universidade Federal dos Vales do Jequitinhonha e Mucuri – UFVJM. R.do Cruzeiro, 01, Teófilo Otoni, Brazil.

[20] Coordinator of the Teaching and Research Center, Hospital Metropolitano Doutor Célio de Castro. Rua Dona Luiza, 311, Belo Horizonte, Brazil.

[21] Head of Value Management Office. Moinhos Research Institute. 910 Ramiro Barcelos Street, 5 floor, Porto Alegre, Brazil.

[22] Emergency and Internal Medicine Physician, Hospital Unimed BH. Av. do Contorno, 3097, Belo Horizonte, Brazil.

[23] Internal Medicine Resident. Hospital Risoleta Tolentino Neves. Rua das Gabirobas, 01, Belo Horizonte, Brazil.

[24] Internal Medicine Physician and Master's Student. Hospital Metropolitano Odilon Behrens. Medical School and University Hospital, Universidade Federal de Minas Gerais. Avenida Professor Alfredo Balena 190 sala 246, Belo Horizonte, Brazil.

[25] Physician and Master's Student. Medical School and University Hospital, Universidade Federal de Minas Gerais. Avenida Professor Alfredo Balena 190 sala 246, Belo Horizonte, Brazil.

[26] Pharmacist and PhD student. Universidade Federal de São João del-Rey. R. Sebastião Gonçalves Coelho, 400, Divinópolis, Brazil.

[27] Internal Medicine Physican and PhD student. Medical School and University Hospital, Universidade Federal de Minas Gerais. Avenida Professor Alfredo Balena 190 sala 246, Belo Horizonte, Brazil.

[28] Undergraduate Medical Student. Medical School, Universidade Federal de Minas Gerais. Avenida Professor Alfredo Balena 190 sala 246, Belo Horizonte, Brazil.

[29] Infectious Diseases Physician. Hospital Eduardo de Menezes. R. Dr. Cristiano Rezende, 2213 - Bonsucesso, Belo Horizonte. Brazil.

[30] UX Designer. Universidade FUMEC. R. Cobre, 200, Belo Horizonte, Brazil.

[31] Professor and Physician. Medical School, Universidade Federal de Minas Gerais. Hospital Julia Kubitschek. Avenida Professor Alfredo Balena 190 sala 246, Belo Horizonte, Brazil.

[32] Researcher. Universidade Federal do Rio Grande do Sul and Institute for Health Technology Assessment (IATS/ CNPq). Rua Ramiro Barcelos, 2359. Prédio 21 | Sala 507, Porto Alegre, Brazil.

[33] Professor. Erasmus MC, University Medical Center Rotterdam, Department of Cardiology, Rotterdam, The Netherlands.

[34] Professor and Physician. Internal Medicine Department. Universidade Federal do Rio Grande do Sul. Coordinator of the Institute for Health Technology Assessment (IATS/ CNPq). Rua Ramiro Barcelos, 2359. Prédio 21 | Sala 507, Porto Alegre, Brazil.

* These authors contributed equally to the work.

** The full list of contributors (to be listed as coauthors for the MEDLINE citation) is provided at the end of the manuscript.

Running title: ABC_2-SPH risk score for mortality in COVID-19

Correspondence to: Milena Soriano Marcolino

University Hospital, Universidade Federal de Minas Gerais

Avenida Professor Alfredo Balena, 110 Room 107. Ala Sul.

Santa Efigênia – Belo Horizonte – MG. Brazil.

CEP 30130-100

Summary boxes

What is already known on this topic?

- Rapid scoring systems may be very useful for fast and effective assessment of COVID-19 patients in the emergency department.
- The majority of available scores have high risk of bias and lack benefit to clinical decision making.
- Derivation and validation studies in low- and middle-income countries, including Latin America, are scarce.

What this study adds

- ABC_2-SPH employs seven well defined variables, routinely assessed upon hospital presentation: age, number of comorbidities, blood urea nitrogen, C reactive protein, Spo2/FiO2 ratio, platelets and heart rate.
- This easy-to-use risk score identified four categories at increasing risk of death with a high level of accuracy, and displayed better discrimination ability than other existing scores.
- A free web-based calculator is available and may help healthcare practitioners to estimate the expected risk of mortality for patients at hospital presentation.

Abstract

Objective: To develop and validate a rapid scoring system at hospital admission for predicting in-hospital mortality in patients hospitalized with coronavirus disease 19 (COVID-19), and to compare this score with other existing ones.

Design: Cohort study

Setting: The Brazilian COVID-19 Registry has been conducted in 36 Brazilian hospitals in 17 cities. Logistic regression analysis was performed to develop a prediction model for in-hospital mortality, based on the 3978 patients that were admitted between March-July, 2020. The model was then validated in the 1054 patients admitted during August-September, as well as in an external cohort of 474 Spanish patients.

Participants: Consecutive symptomatic patients (≥18 years old) with laboratory confirmed COVID-19 admitted to participating hospitals. Patients who were transferred between hospitals and in whom admission data from the first hospital or the last hospital were not available were excluded, as well those who were admitted for other reasons and developed COVID-19 symptoms during their stay.

Main outcome measures: In-hospital mortality

Results: Median (25th-75th percentile) age of the model-derivation cohort was 60 (48-72) years, 53.8% were men, in-hospital mortality was 20.3%. The validation cohorts had similar age distribution and in-hospital mortality. From 20 potential predictors, seven significant variables were included in the in-hospital mortality risk score: age, blood urea nitrogen, number of comorbidities, C-reactive protein, SpO_2/FiO_2 ratio, platelet count and heart rate. The model had high discriminatory value (AUROC 0.844, 95% CI 0.829 to 0.859), which was confirmed in the Brazilian (0.859) and Spanish (0.899) validation cohorts. Our ABC_2-SPH score showed good calibration in both Brazilian cohorts, but, in the Spanish cohort, mortality was somewhat underestimated in patients with very high (>25%) risk. The ABC_2-SPH score is implemented in a freely available online risk calculator (https://abc2sph.com/).

Conclusions: We designed and validated an easy-to-use rapid scoring system based on characteristics of COVID-19 patients commonly available at hospital presentation, for early stratification for in-hospital mortality risk of patients with COVID-19.

Key Words: COVID-19; SARS-CoV-2; mortality; prognosis; risk factors; hospitalizations; score

Introduction

Coronavirus disease 2019 (COVID-19), caused by the SARS-CoV-2 virus, is still the main global health, social and economic challenge, overwhelming health care systems in many countries and heavily burdening others, with over 102 million cases and 2.2 million deaths worldwide.[1,2] While some countries have been declining in new cases, many others have been experiencing a worse surge of the disease than the first wave. Latin America is currently worst-hit region of COVID-19 cases in the world, along with Asia. [3,4] Case rates continue to rise, and some hospitals are nearly at their full capacity of intensive care unit beds. The emergence of the new variants of SARS-CoV-2 in England, South Africa and Brazil, with very high viral growth, potentially more transmissible, less detectable with the RT-PCR technique and an unknown response to the available vaccines, is currently a cause of huge concern[5-7].

Fast and efficient assessment of prognosis of the disease is needed to optimize the allocation of health care and human resources, to empower early identification and intervention of patients at higher risk of poor outcome. A proper assessment tool will guide decision making to develop an appropriate plan of care for each patient[8]. In this context, rapid scoring systems, which combine different variables to estimate the risk of a poor outcome, may be extremely helpful for quick and effective assessment of those patients in the emergency department[9].

Although different scoring systems have been proposed to assess prognosis in COVID-19 patients, the majority of them lack benefit to clinical decision making, and there is a lack of reliable prognostic prediction models[10,11]. Most scores were developed from small cohorts, at high risk for bias, with selected study samples and relatively few outcome events, without clear details of model derivation and validation, as well as unclear reporting on intended use[12-16]. These issues lead to a high risk of model overfitting, thus their predictive performance when used in clinical practice may be different than that reported.[11,12] Additionally, clinical characteristics of COVID-19 patients and disease severity vary in different studies in different countries[17], and external validation was rarely done. Derivation and validation studies in low- and middle-income countries, including Latin America, are scarce[11].

In this context, our aim was to develop and validate an easy applicable rapid scoring system that employs routinely available clinical and laboratory data at hospital presentation, to predict in-hospital mortality in patients with COVID-19, able to

discriminate high vs non-high risk patients. Additionally, we aimed to compare this score with other existing ones.

Methods

This study is part of the Brazilian COVID-19 Registry, an ongoing multicenter observational study described elsewhere[18], and a collaboration with Vall d'Hebron University Hospital, in Barcelona, Spain, for independent external validation. The Brazilian COVID-19 Registry is being conducted according to a predefined protocol, in 36 Brazilian hospitals, located in 17 cities, from four Brazilian states. With regards to the type of hospital, 25 are reference centers for COVID-19 treatment and 19 are academic hospitals. Eighteen are public hospitals; seven are private; and eleven are "mixed", hospitals that provide both public and private services. The median number of hospital beds was 316 (ranging from 60 to 936), and the median number of ICU beds for COVID patients was 22 (ranging from 0 to 105).

Model development, validation and reporting followed guidance from the Transparent Reporting of a Multivariable Prediction Model for Individual Prediction or Diagnosis (TRIPOD) checklist and Prediction model Risk Of Bias ASsessment Tool (PROBAST) (Supplementary Material)[19,20].

Study subjects

Consecutive patients with laboratory-confirmed COVID-19 admitted to the participating hospitals from March 1 to September 30, 2020 were enrolled. COVID-19 diagnosis was confirmed according to the World Health Organization guidance[21]. For the purpose of the present study, eligible patients were ≥18 years-old and had completed hospitalization (i.e., discharge or death). Patients who were transferred between hospitals and admission data from the first hospital (as we aimed to develop a score to be used in the first assessment) or the last hospital was not available were excluded, as well those who were admitted for other reasons and developed COVID-19 symptoms during their stay (as their information from the first assessment would be biased, and their profile is different from the other patients) (Figure 1). Those who were admitted for other reasons were excluded. Although patients who were transferred to another hospital where we could not get the final outcome were excluded, a comparison of the

clinical characteristics with patients who were included is provided in the Supplemental Material (Table S1).

Measurement

Demographic information, clinical characteristics, laboratory and outcome data were collected from the medical records by using a prespecified case report form applying Research Electronic Data Capture (REDCap) tools[22,23] hosted at the Telehealth Center, University Hospital, *Universidade Federal de Minas Gerais*. Data were collected by trained hospital staff or interns. A detailed data management plan (DMP) was developed and provided to all participating centers. An online DMP training was mandatory before local research personnel were allowed to start collecting study data[24].

Data quality assessment

We undertook comprehensive data quality checks to ensure high quality. A code was developed in R software to identify values likely related to data entry errors for vital signs and laboratory variables, based on expert-guided rules. Data were sent to each center for checking and correction. Transfers from one participant hospital to another were merged and considered as a single visit.

Potential predictors for in-hospital mortality

All variables used to calculate the risk score were obtained at hospital admission. A set of potential predictor variables for in-hospital mortality was selected a priori, as recommended[19], taking into account the evidence in literature of association with worse prognosis in patients with COVID-19 or pneumonia, and availability of predictor measurement at the time the model would be used, i.e., hospital admission. We considered predictors that would be available in routine practice in most emergency departments worldwide. It included patient demographic characteristics, pre-existing comorbid medical conditions, home medications, clinical assessment at admission and laboratory data [12]. All laboratory tests were performed at the discretion of the treating physician. Imaging test results were not included, as X-ray and CT scan are not always performed at patient admission and their interpretation involve subjective judgement.

Candidate predictor variables which were not available for at least two thirds of patients within the derivation cohort (more than one third of missing data) were excluded.

Data analysis

Continuous variables were summarized using medians and interquartile ranges (IQR), whereas we used counts and percentages for categorical variables. We reported 95% confidence intervals, and for all two-tailed-tests performed, a p-value less than 0.05 was considered statistically significant. Statistical analysis was performed with R software (version 4.0.2) with the mgcv, finalfit, mice, glmnet, pROC, rms, rmda, and psfmi packages. Details about how missing data were handled, as well as model-building and model-validation procedures, are described below.

Missing data

Considering missing at random after analyzing missing data patterns, multiple imputation with chained equations (MICE) was used to handle missing values only on candidate variables (outcomes were not imputed). Outcome variable was considered as a predictor only in the derivation dataset. We used predictive mean matching (PMM) method for continuous predictors and polytomous regression for categorical variables (two or more unordered levels). The results of 10 imputed datasets, each with 10 iterations, were combined following Rubin's rules[25].

Development of the risk score model

Patients who were admitted before July 31 were included in the development cohort. First, we conducted predictor selection based on clinical reasoning and literature review before modeling. Second, generalized additive models (GAM) were used to examine the relationships between in-hospital mortality and continuous (through penalized thin plate splines) and categorical (as linear components) predictors. During this stage, variable selection was based on D1- (multivariate Wald test) and D2-statistic (pools test statistics from the repeated analyses).

Third, for an easier application of the risk score model at bedside, continuous variables were categorized based on widely accepted cut points, current evidence and/or categories defined in stablished rapid scoring systems from pneumonia and sepsis.

Lastly, we used least absolute shrinkage and selection operator (LASSO) logistic regression to derive the mortality score by scaling the (L1 penalized) shrunk coefficients. The penalty parameter λ in LASSO was chosen using 10-fold cross-validation methods based on mean squared error criterion.

Risk groups were proposed based on predicted probabilities: low risk (< 6.0%), intermediate risk (6.0 – 14.9%), high risk (15.0 – 49.9%), and very high risk (≥ 50.0%).

External validation

We performed an external (temporal) validation analysis using patients who were admitted from August 1 to September 30, 2020. The same investigators collected those data, and missing data were handle as described above.

Independent external validation was also performed in a cohort of patients from Vall d'Hebron University Hospital (Vall d'Hebron COVID-19 Prospective Cohort Study), a 1100-bed public tertiary care hospital with the capacity for more than 60 ICU beds, in Barcelona, Spain[26], part of the public hospital network of the Catalan Health System. Inclusion and exclusion criteria were the same as the beforementioned ones. All patients included were followed for at least 28 days.

Performance measures

We evaluated overall performance using Brier score[27]. Calibration was assessed graphically by plotting the predicted mortality probabilities against the observed mortality, testing intercept equals 0 and slope equals 1. The area under the curve for receiver operating characteristic (AUROC) described model's discrimination, i.e, its ability to predict higher risks for individuals who died than for those who were discharged. Confidence intervals (95% CI) for AUROC were obtained through 2000 bootstrap samples. We also calculated positive and negative predictive values of the derived risk groups.

Model comparisons

The developed model was compared within the validation cohort with existing rapid scores systems for in-hospital mortality in COVID-19 patients. These scores were identified through a literature search of Medline, medRxiv and BioRxiv, with no language or date restrictions, using the search terms "COVID-19," "COVID", "SARS-CoV-2," "coronavirus" combined with "score" and "mortality". The last search was

performed on November 19, 2020. Two authors independently performed article selection and data extraction. Additionally, we also included established scores for pneumonia and sepsis[28-32].

From the set of identified scores, we selected those which with predictors were available within the database and had accessible methods for calculation. Model comparisons were performed using AUROC and decision curve analysis, which describes clinical utility across a range of threshold risks, i.e, the relative value of benefits (if a true positive case is treated) and harms (if a false positive case is treated).

ABC_2-SPH risk score calculator

Risk score calculator was developed in Javascipt, using the Svelte framework while the website was developed in R language (blogdown package).

Ethics

The study protocol conforms to the ethical guidelines of the Declaration of Helsinki. It was approved by the Brazilian National Commission for Research Ethics (CAAE 30350820.5.1001.0008) Individual informed consent was waived due to the severity of the situation and the use of deidentified data, based on medical chart review only. For the independent external validation cohort, it was approved by the and Vall d'Hebron University Hospital Research Ethics Committee (PR(AG)183/2020). The institutional review board granted an informed consent waiver if patients were unable to give oral consent.

Patient and public involvement

This was an urgent public health research study in response to a Public Health Emergency of International Concern. Patients or the public were not involved in the design, conduct, interpretation or presentation of results of this research.

Results

The derivation cohort comprehended data from 3978 patients, from 267 cities of 13 states in Brazil (Figure 2). The median age was 60 [IQR, 48-72] years, 2138 (53.8%) were male, 2789 (70.1%) had at least one comorbidity and 806 (20.3%) died during hospitalization. The median follow-up time was 7 (4-14) days. Table 1 shows

demographic, clinical characteristics and laboratory findings for the derivation and validation datasets.

Development of the risk score model

Thirty-six potential predictor variables were identified (Table S2). Number of comorbidities was created as a composite of ten individual comorbidities shown to have prognostic impact in COVID-9 (hypertension, diabetes mellitus, obesity, coronary artery disease, heart failure, atrial fibrillation or flutter, cirrhosis, cancer and previous stroke)[33,34], as in other scores[35,36]. Twelve variables were excluded due to the excessive number of missing values, two for high collinearity, and one was not recorded within database. Besides that, inotrope use was combined with blood pressure. Therefore, 20 variables were tested.

Through generalized additive model (GAM), a combination of seven variables was selected as the best predictor of in hospital mortality (Table S3). For an easier application to the risk score model at bedside, continuous selected predictors were categorized for LASSO logistic regression. All categories were defined a priori, as recommended,[20] based on widely accepted cut points, current evidence and/or categories defined in stablished rapid scoring systems from pneumonia and sepsis, as follows: advanced age (60-69.9, 70-79.9 and ≥ 80 years), Spo2/Fio2 (SF) ratio (≤ 150.0, 150.1 – 235.0, 235.1 – 315.0, > 315.0), platelet count (<100x10^9/L, 100-150x10^9/L, > 150x10^9/L), C-reactive protein (≥100mg/L), blood urea nitrogen (BUN) (≥42mg/dL), heart rate (≤ 90, 91-130, ≥ 131 bpm).

All variables were statistically significant predictors for in hospital mortality (Table S4 and Figure S1). Shrunk coefficients were scaled to provide a prognostic index and we denoted it as the ABC_2-SPH risk score (Table 2). The sum of the prediction scores ranges between 0 and 20, with a high score indicating higher risk of in-hospital mortality.

Risk groups were proposed based on predicted probabilities (Table 3): low risk (0-1 score, observed in hospital mortality 2.0%), intermediate risk (2-4 score, 11.4%), high risk (5-8 score, 32.0%), and very high risk (≥ 9 score, 69.4%). Subject-specific risks can be assessed using the developed ABC_2-SPH risk score Web-based calculator (https://abc2sph.com/), freely available to the public.

As well as GAM and LASSO, ABC_2-SPH risk score showed good overall performance (Brier score: 0.114) and good discrimination (AUROC equal 0.842; 95% CI 0.840–0.843) within the derivation cohort (Table 4).

External validation – Brazilian cohort

A total of 1054 patients were included in the validation cohort. The median age was 62 (interquartile range 48-73) years, 582 (55.2%) were male and 745 (70.7%) had at least one comorbidity. The median follow-up time was 7 (4-13) days. Two hundred and eight patients (19.7%) died during hospitalization. The distribution of patients across range ABC_2-SPH Score in derivation and validation cohorts are presented in Figure 3.

We observed good discrimination (AUROC equal 0.859; 95% CI 0.833 to 0.885; Figure 4), overall performance (Brier = 0.108) and calibration (slope = 1.138, intercept = 0.114, p-value = 0.184; Figure S2a) of the ABC_2-SPH risk score under the validation cohort (Figure 4). The good performance is also demonstrated in sensitivity analyses using complete case data (Table S5).

Low, intermediate and high-risk groups showed good negative predictive values (99.7%, 88.1% and 71.0%, respectively). A positive predictive value of 73.7% was observed in patients classified as at very high mortality risk.

External validation – Spanish cohort

A second external (geographic) validation was performed within a Spanish cohort with 474 patients and 82 (17,3%) in hospital mortality. The demographic and clinical characteristics at admission are listed in Table 1. The median follow-up time was 21 (IQR, 7-40) days. Only complete cases were included.

ABC_2-SPH Score showed high discrimination (AUROC= 0.899, 95% CI 0.864 to 0.934; Figure 4), good overall performance (Brier = 0.093), but an underestimation of true mortality risk in patients with a predicted probability above 25% (intercept = 0.729, slope = 1.519, p-value = 0.001; Figure S2b).

Literature review

The literature search identified 39 scores to predict mortality in COVID-19 patients (Table 5). Most of them were still preprints (28%), in 36% the derivation cohort was from China, 21% from the United States and none from South America.

Multivariate logistic regression and LASSO regression were used in 16 and 10 studies, respectively, artificial intelligence techniques in seven studies and Cox regression analysis in 3 studies. Two scores were developed by consensus. The population of the development cohort was composed by adults-only in 51.3% of the studies, the age range was not clear in 41.4% and elderly patients in one of them. Thirteen studies developed points-based scores, three were published as nomograms and all the other ones required formulas for calculation.

From the 27 (69.2%) developed scores to predict in-hospital mortality, in three studies the full information required for proper calculation was not available, in five studies the assay used for D-dimer or troponin was not described to allow proper comparison, in two studies the variables were not clearly defined (such as "kidney failure", "elevated" CPR, and "cardiovascular and pulmonary comorbidity"), in two the variables were not applicable for other populations (such as province and coming from Wuhan), and in 12 one or more variables required were not in our study protocol.

Comparison with other scores

Based on complete case validation cohort, the ABC_2-SPH score achieved better discrimination (Table 6, Figure 5a) than other prediction scoring systems for COVID-19, pneumonia and sepsis (0.85; 95%CI: 0.82 – 0.88). Xie's and Zhang's score[8,37,38] showed good discrimination, but the number of complete cases and deaths were relatively small. Considering clinical utility (Figure 5b), ABC_2-SPH showed a better performance compared to the two most discriminating scores for in-hospital mortality that were tested in more than 700 patients (A-DROP and CURB-65[29]). COVID-AID-7 and COVID-AID-14 were not included, as they have assessed 7 and 14 day-mortality, respectively, and not in-hospital).

Discussion

Main findings

ABC_2-SPH score is simple, objective, easily available at hospital admission and easily calculated, employing seven well defined and routinely recordable variables: **a**ge, **b**lood urea nitrogen, number of **c**omorbidities, **C**-reactive protein, **S**pO2/FiO2 ratio, **p**latelet count, **h**eart rate. It has shown to be a reliable tool to estimate in-hospital

mortality in COVID-19 patients, although true mortality risk is somewhat underestimated in very high risk patients. Model performance compared surpassed other existing scores.

The pandemic of COVID-19 disease has inflicted a heavy burden on the healthcare system of numerous countries. Little is known about how long the immune system will remain protective after vaccination or recovery from infection, and scientists have been predicting that SARS-CoV-2 "is here for the long haul"[39]. Therefore, it is of utmost importance to better identify those patients with higher risk of mortality, to inform early interventions and the need of more frequent repetitive assessments, to reduce the risk of death.

Comparison with other studies

The majority of developed scores are limited by methodological bias in development cohorts.

Our prediction model was developed based on a large sample size of consecutive adult patients with confirmed COVID-19, from hospitals of different sizes, types and locations, to minimize the selection bias. Robust models require large sample sizes, which produce more reliable and accurate results[19]. It is estimated that 10-15 outcome events per predictor are required[40]. Among the models analyzed for comparison, only 30.8% used a sample with more than 1000 patients, 41.0% used a sample with less than 500 patients, and 41.0% were developed and validated in a sample with less than 100 events.

All studies have missing data[19], this reality may be due to lack of standardization of the necessary exams at hospital admission, and differences in resources available in hospitals for carrying out tests. The approach of excluding the missing data and performing the analysis with the complete cases can lead to biased results, since the complete cases may not adequately represent the entire original study sample, generating a selection bias[19]. To avoid this type of bias, in our model multiple imputation with chained equations (MICE) was used to handle missing values on potential predictors, where the foul was assumed at random. Most of the models we analyzed for comparison (69.2%) did not perform or did not describe whether imputation methods were used for the missing data, therefore, there is a high risk of bias related to the treatment of missing data.

As the accuracy of a prediction model is always high whether the model is validated on the development cohort used to derive the model only, the assessment of accuracy in those studies may be overoptimistic. It is important that studies that develop prediction models use some validation method to quantify any optimism in the predictive performance of the model developed and adjust the model for overfitting[19]. In 43.6% of the analyzed studies, external validation of the developed model was not performed. External validation is highly recommended to assess the performance of a prediction model on other participant data that was not used for the development of the model[19].

Previous studies have observed the variables included in the ABC_2-SPH score as risk factors for severe COVID-19, what shows that our results are in line with the available evidence. Age and number of comorbidities were reported as independent risk factors for developing severe COVID and mortality in several publications[10,36,41]. The strong age gradient per decade after 60 years-old is in line with other series[10,14]. One of the main causes of death in COVID-19 patients is the unregulated immune response, with an uncontrolled production and secretion of cytokines. Aging is associated with a well-known decline in the adaptive and innate immunity, which plays a major role in the increased susceptibility of infections[42]. Age-related immune imbalance is also related to an increased severity in pro-inflammatory response and increased cytokine production, what is believed to increase patient vulnerability to the unregulated inflammatory response in COVID-19[43]. Other authors hypothesize that decreased lung elasticity, increased end-expiratory and abnormal alveolar integrity related to lung senescence, which may be associated to kidney senescence play a role in the predisposition for severe COVID-19 and mortality[10].

It is important to highlight the evidence of a decreased antibody production following immunization in the elderly, as well as shortened duration of protective immunity[43]. This might be the case for COVID-19 as well. Therefore, even being a priority group for the vaccine, this age group will probably remain a major risk factor for mortality.

The number of comorbidities indicates the importance of pre-existing conditions to the severity of COVID-19. Even though comorbidities are age-dependent factors, the number of comorbidities remained as an independent risk factor in the final model.

As we aimed to use variables easily available at ED admission of any institution, we opted to evaluate the ratio of peripheral oxygen saturation over the inspired fraction of oxygen (SpO_2/FiO_2, SF ratio), instead of the ratio of arterial oxygen partial pressure over the fraction of inspired oxygen, like the COVID-AID score[44]. Arterial blood gas puncture and analysis is an invasive and complex procedure, which may be time consuming for the team. Additionally, a recent publication highlighted that despite widespread familiarity with use of PaO_2/FiO_2 ratio using blood gas analysis, clinical recognition of acute respiratory distress syndrome remains poor[45]. The authors assessed 28,758 mechanical SF ratio and PaO_2/FiO_2 ratio in mechanical ventilated patients, and observed that PaO2/FiO_2 ratios were substantially less available or even unavailable in a significant proportion of ventilated patients[45]. SF ratio was already validated as a substitute for the PaO_2/FiO_2 ratio in assessing the oxygenation criterion of patients with acute lung injury and acute respiratory distress syndrome[46].

COVID-19 associated hyperinflammation and coagulopathy are correlated to with a wide deviation in various inflammatory markers and hemostasis parameters, including C-reactive protein, thrombocytopenia, D-dimer and prothrombin time, and thus these are potential prognostic markers of increased mortality in COVID-19[47,48].

Consistent with prior studies, we also observed utility of CRP thrombocytopenia.

C-reactive protein in an acute phase reactor with established prognostic prediction role in intensive care septic and non-septic patients[49,50], and it has been included in different scores an independent predictor for mortality.[51,52]

The prognostic value of thrombocytopenia in patients with COVID-19 has shown in a recent meta-analysis[53], and it was also included in other scores[10,51]. The exact explanation is still unknown, and it is probably multifactorial, related to direct infection of bone marrow cells by the virus, resulting in abnormal hematopoiesis; platelet destruction by the immune system; endothelial damage triggering platelet activation, aggregation and microthrombi in the lung; and abnormal platelet defragmentation in the lungs[53].

A recent meta-analysis of 16 studies, which included 2783 surviving and 697 non-surviving cases, has shown significantly higher levels of D-dimer on admission in patients who died compared to the ones who were discharged[47]. This exam was included as a predictor in different scores[13,16,52,54,55]. Although D-dimer was collected in

our study, D-dimer assays varied widely among different hospitals. Ideally, the value has to be determined with the same methodology, preferably from the same manufacturer, and this information was not available in any of the studies.

A recent publication highlighted confusion and potential for misinformation in reporting D-dimer data in COVID-19[56]. The authors emphasized that the considerable variation in reporting units for D-dimer is potentially under-recognized in various studies, with at least 28 potential theoretical combinations of measuring units for D-dimer, either D-dimer units (DDU) or fibrinogen equivalent units (FEU), which are approximately 2× those of DDU. There is also possibility for misreporting of D-dimer data based on poor or incomplete reporting. The authors provided examples of serious errors in the reported values and/or units as reported in the literature related to COVID-19, even in high impact journals.

Most studies have not reported how they dealt with cases who were transferred between hospitals. Although SOFA scores tend to be low at hospital admission, Zhou et al[57] observed that age, SOFA score and D-dimer at admission were independent risk factors for mortality. However, they opted to include those patients, even patients with late stage COVID-19, using admission data from the second hospital only. It is quite likely there was a higher chance those patients were already with critical disease[57]. As the score is intended to be used at hospital admission, we opted to exclude patients who were transferred between hospitals and admission data from the first hospital was not available.

Blood urea nitrogen elevation was a strong predictor for mortality, what is in line with other scores[36,58,59]. Kidney disease has been widely described as a risk factor for in-hospital mortality. Although autopsy studies did not find conclusive evidence of SARS-CoV2 infection in the kidney, some authors hypothesize that the damage may be mediated by direct cytopathic effects of SARS-Cov2 on the kidney tissue, immune-mediated damage due to virus-induced immune complexes, as well as the effects of the inflammatory response, hypoxia and shock[60-62].

Strengths and limitations

A major strength of this rapid scoring system is its simplicity, the use of objective parameters, what helps to reduce inter-user variability, easily available at the emergency department presentation, even in under-resourced settings. A major strength of this study is that it followed strict methodological criteria, recommended by TRIPOD

checklist and PROBAST[20], and was based on robust sample of patients with laboratory confirmed SARS-CoV-2 infection, from a collaboration among researchers from 36 public, private and mixed hospitals of different sizes in four Brazilian states, to ensure diversity of the population studied and representativeness of the intended target population. The majority of published scores were developed in China or the US (56.4%) and Europe (25.6%), this is the first study in the Latin American population. Data were obtained by detailed medical chart reviews, and we were able to collect comprehensive data from a large number of patients and follow 98.5% of the patients from admission to discharge or death. Decisions about which predictors to retain in the final model did not rely on potentially biased univariable selection of predictors. They were based on clinical reasoning, previous evidence from other cohorts and systematic reviews on prognostic factors for COVID-19 patients and availability of predictor measurement at hospital admission[19].

In a huge country such as Brazil, the development of a score that truly corresponds to the reality of our population's characteristic was only possible by the collaborative work among several hospitals from all the regions of the country. The COVID-19 cause and requirement for agile answers from the scientific community motivated the fast and precise teamwork and allowed the achievement of the creation of a tool to support the daily work in the frontline to combat the pandemic. We believe that the learning regarding the development of qualified and useful research engaging several centers could allow us to generate more accurate and faster results to subside health policies in the future.

Patients who were transferred to other hospitals and thus were lost to follow-up do not characterize selection bias, as they similar characteristics of the development and validation cohorts, and a risk similar to those cohorts: of the 77 patients, 53 presented complete data and had their scores calculated - low risk 30.2%, intermediate 35.8%, high 22.6%, very high 11.3%.

With regards to study limitations, it was a retrospective analysis subject to the drawbacks of patient records review. Obesity was not directly measured by body mass index, but rather clinical defined, gathered from medical records, which may have led to underreporting. Due to the pragmatic study design, laboratory tests were performed at the discretion of the treating physician, and we did not have a full dataset on all laboratory parameters of interest available. Some laboratory parameters, which proved

to be of prognostic relevance in other studies, were not available for at least 2/3 of patients in our sample. Therefore, we cannot rule out that variables with a higher proportion of missing data would have had a significant impact on mortality prediction. Additionally, we were unable to assess the predictive ability of some scores, as some required variables were not available.

Another bias from the Spanish validation cohort is the fact that the majority of those patients came from the beginning of the pandemic, and management of the patients improved during subsequent waves. Data include 28-day mortality, which may differ from the Brazilian data, although the score was able to show very good discrimination.

Implications for clinicians and policymakers

ABC_2-SPH score may be very useful in a real-world setting, to provide healthcare practitioners the decision support that is needed to help them better identify and prioritize the care of patients who have the higher risk of death. Its development and validation followed strict methodological criteria, and the score fulfils the majority of the characteristics of an ideal score[63]. It can be used in all emergency departments, regardless of the level of resource settings. The results represent the experience of 36 hospitals in 17 cities in Brazil, and one hospital in Spain, and they are highly relevant to the current pandemic. It can be easily calculated at bedside or could be easily integrated to the electronic medical records for an automatic computation. It may help clinicians to identify high-risk patients from the triage phase, as well as to identify those most appropriate to be enrolled into therapeutic trials, may make possible to expand inclusion criteria through the early identification of patients who may benefit from therapy[64]. It might also be useful to help guiding recommendations for early palliative consultation[65,66].

Different from what has been mistakenly suggested[36,67-69], the results from this study do not suggest that patients from low-risk group may be discharged for home treatment. No score so far has specifically tested this hypothesis. A recent editorial has highlighted the importance of taking into account the “treatment paradox”: patients identified to be at the low risk group were at low risk due to the interventions received in hospital[70]. It must not be interpreted as the risk to a patient if no actions are taken.

Sperrin & McMillan counterfactual prediction modelling as a potential solution to minimize bias from treatment paradox[70]. More importantly, due to the treatment paradox, scoring systems developed and validated in in-hospital settings cannot be used in outpatient settings without further validation, as it has been mistakenly suggested[71].

Unanswered questions and future research

We believe that ABC_2-SPH score may hold potential generalizability for other countries. However, prediction models are population specific and may produce different results in different populations[72]. Considering that thresholds for admission may vary, hospitalized COVID-19 population may be different, the outcome events are different and patient management may be different, further validation (and re-calibration) in different health care settings is recommended. In particular, we learned that our model might underestimate mortality in high-risk individuals.

As we opted to develop the score focusing on information available at admission, as this would make it more useful for clinicians, other important factors during hospitalization that may impact prognosis were not included. Further analysis involving these factors are required.

ABC_2-SPH score may help clinicians to make a prompt and reasonable decision to optimize patient management and potentially reduce mortality. However, further prospective studies are needed to investigate whether the use of the score in the emergency department indeed trigger actions that result in reduced complications and hospital mortality. Additionally, due to the rapidly changing nature of the COVID-19 and the disease management, model performance should be monitored closely over time and space[70].

Future studies may also investigate risk factors for mortality among patients who develop COVID-19 symptoms during hospital admission due to other conditions.

Conclusion

In conclusion, we developed and validated the ABC_2-SPH rapid scoring system and a web-based risk calculator. This score, based on age, number of comorbidities, blood urea nitrogen, C-reactive protein, platelet count, peripheral oxygen saturation and oxygen support at admission is an inexpensive tool, showed to objectively and accurately predict in-hospital mortality in COVID-19 patients. It may be used at bedside for earlier identification of in-hospital mortality risk and, thus, inform clinical decisions

and the assignment to the appropriate level of care and treatment for COVID-19 patients.

Brazilian COVID-19 Registry investigators (in alphabetical order): Alexandre Vargas Schwarbold (0000-0002-5535-6288), Amanda de Oliveira Maurílio (0000-0002-9355-9596), Ana Luiza Bahia Alves Scotton (0000-0002-5857-2031), André Pinheiro Weber (0000-0002-1386-2553), André Soares de Moura Costa (0000-0002-9153-1186), Andressa Barreto Glaeser (0000-0002-1534-0900), Angélica Aparecida Coelho Madureira (0000-0002-8465-5674), Angelinda Rezende Bhering (0000-0002-7391-1951), Bruno Mateus de Castro (0000-0002-7013-2127), Carla Thais Cândida Alves da Silva (0000-0003-1726-4494), Carolina Marques Ramos (0000-0002-8258-0891), Caroline Danubia Gomes (0000-0002-5254-4320), Cíntia Alcantara de Carvalho (0000-0001-8240-2481), Daniel Vitório Silveira (0000-0002-7381-1651), Diego Henrique de Vasconcelos (0000-0002-8605-2121), Edilson Cezar (0000-0002-4006-4354), Elayne Crestani Pereira (0000-0002-1178-669X), Emanuele Marianne Souza Kroger (0000-0001-5108-4337), Felipe Barbosa Vallt (0000-0001-5111-1921), Fernanda Barbosa Lucas (0000-0003-3926-900X), Fernando Graça Aranha (0000-0001-9173-8892), Frederico Bartolazzi (0000-0002-9696-4685), Gabriela Petry Crestani (0000-0002-4991-4941), Gisele Alsina Nader Bastos (0000-0001-9668-1434), Glícia Cristina de Castro Madeira (0000-0002-9375-5319), Helena Carolina Noal (0000-0001-9912-2881), Heloisa Reniers Vianna (0000-0003-1144-6262), Henrique Cerqueira Guimarães (0000-0001-6511-7400), Isabela Moraes Gomes (0000-0002-4653-6447), Israel Molina Romero (0000-0001-6642-7515), Joanna d'Arc Lyra Batista (0000-0002-3703-2845), Joice Coutinho de Alvarenga (0000-0003-3536-7112), Júlia Di Sabatino Santos Guimarães (0000-0002-4738-7754), Júlia Drumond Parreiras de Morais (0000-0002-9538-6828), Juliana Machado Rugolo (0000-0003-3984-4959), Karen Cristina Jung Rech Pontes (0000-0002-8068-5331), Kauane Aline Maciel dos Santos (0000-0003-0041-7271), Leonardo Seixas de Oliveira (0000-0003-1575-6559), Lílian Santos Pinheiro (0000-0003-2957-7674), Liliane Souto Pacheco (0000-0002-0229-0692), Lucas de Deus Sousa (0000-0002-0281-0897), Luciana Siuves Ferreira Couto (0000-0002-7948-222X), Luciane Kopittke (0000-0002-6606-7756), Luis Cesar Souto de Moura (0000-0002-8138-6042), Luisa Elem Almeida Santos (0000-0003-0264-5497), Máderson Alvares de Souza Cabral (0000-0001-8556-1998), Maíra Dias Souza (0000-

0003-3546-4000), Marcela Gonçalves Trindade Tofani (0000-0003-2738-0747), Marcelo Carneiro (0000-0003-3603-1987), Marcus Vinícius de Melo Andrade (0000-0002-3716-0919), Maria Angélica Pires Ferreira (0000-0003-0961-524X), Maria Aparecida C. Bicalho PhD (0000-0001-6298-9377), Maria Clara Pontello Barbosa Lima (0000-0003-0461-2787), Mariana Frizzo de Godoy (0000-0002-6631-8826), Marilia Mastrocolla de Almeida Cardoso (0000-0002-6231-5425), Meire Pereira de Figueiredo (0000-0002-0724-6893), Natália da Cunha Severino Sampaio (0000-0002-5799-8006), Natália Lima Rangel (0000-0001-6968-2271), Natália Trifiletti Crespo (0000-0002-1009-5299), Neimy Ramos de Oliveira (0000-0001-5408-9459), Pedro Ledic Assaf (0000-0002-8419-5719), Petrônio José de Lima Martelli (0000-0001-6920-6435), Rafaela dos Santos Charão de Almeida (0000-0002-2091-5772), Raphael Castro Martins (0000-0003-2167-3645), Raquel Lutkmeier (0000-0002-9519-261X), Reginaldo Aparecido Valacio (0000-0003-2666-7092), Renan Goulart Finger (0000-0002-5569-7787), Ricardo Bertoglio Cardoso (0000-0002-4931-4081), Roberta Pozza (0000-0003-1733-6398), Roberta Xavier Campos (0000-0002-2963-8994), Rochele Mosmann Menezes (0000-0002-1548-1607), Roger Mendes de Abreu (0000-0002-2153-1340), Rufino de Freitas Silva (0000-0003-1952-9193), Silvana Mangeon Mereilles Guimarães (0000-0002-9655-6294), Silvia Ferreira Araújo (0000-0003-4782-5440), Susany Anastácia Pereira (0000-0001-5039-443X), Talita Fischer Oliveira MD (0000-0002-4614-3109), Tatiana Kurtz (0000-0003-2124-711X), Thainara Conceição de Oliveira (0000-0002-1248-2305), Thaíza Simônia Marinho Albino de Araújo (0000-0002-2082-9786), Thulio Henrique Oliveira Diniz (0000-0003-1798-4643), Veridiana Baldon dos Santos (0000-0002-9372-5785), Virginia Mara Reis Gomes (0000-0001-7592-5604), Vitor Augusto Lima do Vale (0000-0002-9910-2861), Yuri Carlotto Ramires (0000-0002-0730-8716).

Acknowlegments

We would like to thank the hospitals which are part of this collaboration, for supporting this project: Hospital Bruno Born; Hospital Cristo Redentor; Hospital das Clínicas da Faculdade de Medicina de Botucatu; Hospital das Clínicas da UFMG; Hospital das Clínicas da Universidade Federal de Pernambuco; Hospital de Clínicas de Porto Alegre; Hospital de Santo Antônio; Hospital Eduardo de Menezes; Hospital João XXIII; Hospital Julia Kubitschek; Hospital Mãe de Deus; Hospital Márcio Cunha; Hospital Mater Dei Betim-Contagem; Hospital Mater Dei Contorno; Hospital Mater Dei

Santo Agostinho; Hospital Metropolitano Dr. Célio de Castro; Hospital Metropolitano Odilon Behrens; Hospital Moinhos de Vento; Hospital Nossa Senhora da Conceição; Hospital Regional Antônio Dias; Hospital Regional de Barbacena Dr. José Américo; Hospital Regional do Oeste; Hospital Risoleta Tolentino Neves; Hospital Santa Cruz; Hospital Santa Rosália; Hospital São João de Deus; Hospital São Lucas da PUCRS; Hospital Semper; Hospital SOS Cárdio; Hospital Tacchini; Hospital Unimed-BH; Hospital Universitário Canoas; Hospital Universitário Ciências Médicas; Hospital Universitário de Santa Maria.

We also thank all the clinical staff at those hospitals, who cared for the patients, and all undergraduate students who helped with data collection.

Funding

This study was supported in part by Minas Gerais State Agency for Research and Development (*Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG*) [grant number APQ-00208-20], National Institute of Science and Technology for Health Technology Assessment (*Instituto de Avaliação de Tecnologias em Saúde – IATS*)/ National Council for Scientific and Technological Development (*Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq*) [grant number 465518/2014-1], and CAPES Foundation (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*) [grant number 88887.507149/2020-00]. AS was supported by a Postdoctoral grant “Juan Rodés” (JE18/00022) from Instituto de Salud Carlos III through the Ministry of Economy and Competitiveness, Spain.

Role of the funder/sponsor

The sponsors had no role in study design; data collection, management, analysis, and interpretation; writing the manuscript; and decision to submit it for publication. MSM and MP had full access to all the data in the study and had responsibility for the decision to submit for publication.

Conflicts of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability statement

Data are available upon reasonable request.

Transparency declaration

The lead authors (MSM and MCP) affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned (and, if relevant, registered) have been explained.

References

1. WHO. Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update. 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ (accessed 01-04 2021).
2. Emanuel EJ, Persad G, Upshur R, et al. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *New England Journal of Medicine* 2020; **382**(21): 2049-55.
3. JHU. COVID-19 Map - Johns Hopkins Coronavirus Resource Center. 2020 (accessed 12-04 2020).
4. WHO. Coronavirus disease (COVID-19). 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20201005-weekly-epi-update-8.pdf2020 (accessed 01-04 2021).
5. Conti P, Caraffa A, Gallenga C, et al. The British variant of the new coronavirus-19 (Sars-Cov-2) should not create a vaccine problem. *Journal of Biological Regulators Homeostatic Agents* 2121; **35**(1).
6. Zhang S. A Troubling New Pattern Among the Coronavirus Variants. 2021. https://www.theatlantic.com/health/archive/2021/01/coronavirus-evolving-same-mutations-around-world/617721/ (accessed 01-22 2021).
7. Faria NR, Clar IM, Candido D, et al. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. 2021. https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586 (accessed 01-22 2021).
8. Zhang S, Guo M, Duan L, et al. Development and validation of a risk factor-based system to predict short-term survival in adult hospitalized patients with COVID-19: a multicenter, retrospective, cohort study. *Critical Care* 2020; **24**(1): 1-13.
9. Leeuwenberg AM, Schuit E. Prediction models for COVID-19 clinical decision making. *The Lancet Digital Health* 2020; **2**(10): e496-e7.
10. Fumagalli C, Rozzini R, Vannini M, et al. Clinical risk score to predict in-hospital mortality in COVID-19 patients: a retrospective cohort study. *BMJ Open* 2020; **10**(9): e040729.
11. Gupta RK, Marks M, Samuels THA, et al. Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: An observational cohort study. *European Respitatory Journal* 2020; **56**(6).
12. Wynants L, Calster BV, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *British Medical Journal Publishing Group* 2020; **369**: m1328.
13. Wang J, Zhang H, Qiao R, et al. Thrombo-inflammatory features predicting mortality in patients with COVID-19: The FAD-85 score. *Journal of International Medical Research* 2020; **48**(9): 1-14.
14. Allenbach Y, Saadoun D, Maalouf G, et al. Development of a multivariate prediction model of intensive care unit transfer or death: A French prospective cohort study of hospitalized COVID-19 patients. *PloS One* 2020; **15**(10): e0240711.
15. Kim I-C, Song JE, Lee HJ, et al. The Implication of Cardiac Injury Score on In-hospital Mortality of Coronavirus Disease 2019. *Journal of Korean Medical Science* 2020; **35**(39): e349.
16. Zhou J, Huang L, Chen J, et al. Clinical features predicting mortality risk in older patients with COVID-19. *Current Medical Research and Opinion* 2020; **36**(11): 1753-9.

17. Goel S, Jain T, Hooda A, et al. Clinical Characteristics and In-Hospital Mortality for COVID-19 Across The Globe. *Cardiology and Therapy* 2020; **9**(2): 553-9.
18. Marcolino MS, Ziegelmann PK, Souza-Silva MV, et al. Clinical characteristics and outcomes of patients hospitalized with COVID-19 in Brazil: results from the Brazilian COVID-19 Registry. *International Journal of Infectious Diseases* 2021.
19. 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**(1): W1-73.
20. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine* 2019; **170**(1): 51-8.
21. Organization WH. Diagnostic testing for SARS-CoV-2: interim guidance, 11 September 2020: World Health Organization, 2020.
22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics* 2009; **42**(2): 377-81.
23. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics* 2019; **95**: 103208.
24. Gregory KE, Radovinsky L. Research strategies that result in optimal data collection from the patient medical record. *Appl Nurs Res* 2012; **25**(2): 108-16.
25. Rubin DB. Multiple imputation for nonresponse in surveys: John Wiley & Sons; 2004.
26. Montalva AS, Nadal JS, Pereiro JE, et al. Early outcomes of tocilizumab in adults hospitalized with severe COVID19. An initial report from the Vall dHebron COVID19 prospective cohort study. *MedRxiv* 2020.
27. Rufibach K. Use of Brier score to assess binary predictions. *Journal of Clinical Epidemiology* 2010; **63**(8): 938-9.
28. Lim WS, van der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax* 2003; **58**(5): 377-82.
29. Liu JL, Xu F, Zhou H, et al. Expanded CURB-65: a new score system predicts severity of community-acquired pneumonia with superior efficiency. *Scientific Reports* 2016; **6**: 22911.
30. Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med* 1997; **336**(4): 243-50.
31. Cag Y, Karabay O, Sipahi OR, et al. Development and validation of a modified quick SOFA scale for risk assessment in sepsis syndrome. *PloS One* 2018; **13**(9): e0204608.
32. Olsson T, Terént A, Lind L. Rapid Emergency Medicine Score: a new prognostic tool for in□hospital mortality in nonsurgical emergency department patients. *Journal of Internal Medicine* 2004; **255**(5): 579-87.
33. Harrison SL, Fazio-Eynullayeva E, Lane DA, Underhill P, Lip GYH. Comorbidities associated with mortality in 31,461 adults with COVID-19 in the United States: A federated electronic medical record analysis. *PLoS Med* 2020; **17**(9): e1003321.
34. Harrison SL, Fazio□Eynullayeva E, Lane DA, Underhill P, Lip GY. Atrial fibrillation and the risk of 30□day incident thromboembolic events, and mortality in adults≥ 50 years with COVID□19. *Journal of Arrhythmia* 2020; (00): 1–7.

35. Tuty Kuswardhani RA, Henrina J, Pranata R, Anthonius Lim M, Lawrensia S, Suastika K. Charlson comorbidity index and a composite of poor outcomes in COVID-19 patients: A systematic review and meta-analysis. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 2020; **14**(6): 2103-9.
36. Knight SR, Ho A, Pius R, 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.
37. Xie J, Hungerford D, Chen H, et al. Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19. *MedRxiv* 2020.
38. Zhang H, Shi T, Wu X, et al. Risk prediction for poor outcome and death in hospital in-patients with COVID-19: derivation in Wuhan, China and external validation in London, UK. *Preprints with The Lancet* 2020.
39. Scudellari M. How the pandemic might play out in 2021 and beyond. *Nature* 2020; **584**(6): 22-5.
40. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020; **368**: m441.
41. Liang W, Liang H, Ou L, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Internal Medicine* 2020; **180**(8): 1081-9.
42. Fuentes E, Fuentes M, Alarcon M, Palomo I. Immune system dysfunction in the elderly. *Anais da Academia Brasileira de Ciências* 2017; **89**(1): 285-99.
43. Sherwani S, Khan MWA. Cytokine Response in SARS-CoV-2 Infection in the Elderly. *Journal of Inflammation Research* 2020; **13**: 737-47.
44. Hajifathalian K, Sharaiha RZ, Kumar S, et al. Development and external validation of a prediction risk model for short-term mortality among hospitalized U.S. COVID-19 patients: A proposal for the COVID-AID risk tool. *PloS One* 2020; **15**(9): e0239536.
45. Adams JY, Rogers AJ, Schuler A, et al. Association between peripheral blood oxygen saturation (SpO2)/fraction of inspired oxygen (FiO2) ratio time at risk and hospital mortality in mechanically ventilated patients. *The Permanente Journal* 2020; **24**(19): 113.
46. Rice TW, Wheeler AP, Bernard GR, et al. Comparison of the SpO2/FIO2 ratio and the PaO2/FIO2 ratio in patients with acute lung injury or ARDS. *Chest* 2007; **132**(2): 410-7.
47. Gungor B, Atici A, Baycan OF, et al. Elevated D-dimer levels on admission are associated with severity and increased risk of mortality in COVID-19: A systematic review and meta-analysis. *The American Journal of Emergency Medicine* 2021; **39**: 173-9.
48. Nascimento IJBd, Groote TCv, Mathúna DPO, et al. Clinical, laboratory and radiological characteristics and outcomes of novel coronavirus (SARS-CoV-2) infection in humans: A systematic review and series of meta-analyses. *PloS One* 2020; **15**(9): e0239235.
49. Qu R, Hu L, Ling Y, et al. C-reactive protein concentration as a risk predictor of mortality in intensive care unit: a multicenter, prospective, observational study. *BMC Anesthesiology* 2020; **20**(292): 1-9.
50. Koozi H, Lengquist M, Frigyesi A. C-reactive protein as a prognostic factor in intensive care admissions for sepsis: A Swedish multicenter study. *Journal of Critical Care* 2020; **56**: 73-9.

51. Nicholson CJ, Wooster L, Sigurslid HH, et al. Estimating Risk of Mechanical Ventilation and Mortality Among Adult COVID-19 patients Admitted to Mass General Brigham: The VICE and DICE Scores. *Medrxiv* 2020; **17**: 1-33.
52. Weng Z, Chen Q, Li S, et al. ANDC: an early warning score to predict mortality risk for patients with Coronavirus Disease 2019. *Journal of Translational Medicine* 2020; **18**(328): 1-10.
53. Bashash D, Hosseini-Baharanchi FS, Rezaie-Tavirani M, et al. The Prognostic Value of Thrombocytopenia in COVID-19 Patients; a Systematic Review and Meta-Analysis. *Archives of Academic Emergency Medicine* 2020; **8**(1): e75.
54. Shang Y, Liu T, Wei Y, et al. Scoring systems for predicting mortality for severe patients with COVID-19. *EClinicalMedicine* 2020; **24**: 100426.
55. Hu C, Liu Z, Jiang Y, et al. Early prediction of mortality risk among severe COVID-19 patients using machine learning. *MedRxiv* 2020.
56. Favaloro EJ, Thachil J. Reporting of D-dimer data in COVID-19: some confusion and potential for misinformation. *Clinical Chemistry and Laboratory Medicine* 2020; **58**(8): 1191–9.
57. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* 2020; **395**(10229): 1054-62.
58. Ko H, Chung H, Kang WS, et al. An Artificial Intelligence Model to Predict the Mortality of COVID-19 Patients at Hospital Admission Time Using Routine Blood Samples: Development and Validation of an Ensemble Model. *Journal of Medical Internet Research* 2020; **22**(12): e25442.
59. Levy TJ, Richardson S, Coppa K, et al. Development and validation of a survival calculator for hospitalized patients with COVID-19. *MedRxiv* 2020.
60. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 2020; **395**(10223): 497-506.
61. Lemos JA, McGuire DK, Drazner MH. B-type natriuretic peptide in cardiovascular disease. *The Lancet* 2003; **362**(9380): 316-22.
62. Yao XH, Li TY, He ZC, et al. [A pathological report of three COVID-19 cases by minimal invasive autopsies]. *Zhonghua Bing Li Xue Za Zhi* 2020; **49**(5): 411-7.
63. Oprita B, Aignatoaie B, Gabor-Postole D. Scores and scales used in emergency medicine. Practicability in toxicology. *Journal of Medicine and Life* 2014; **7**(Spec Iss 3): 4-7.
64. Haimovich AD, Ravindra NG, Stoytchev S, et al. Development and validation of the quick COVID-19 severity index: a prognostic tool for early clinical decompensation. *Annals of Emergency Medicine* 2020; **76**(4): 442-53.
65. Altschul DJ, Unda SR, Benton J, et al. A novel severity score to predict inpatient mortality in COVID-19 patients. *Scientific Reports* 2020; **10**(1): 16726, 1-8.
66. Sheehan J, Ho KS, Poon J, Sarosky K, Fung JY. Palliative care in critically ill COVID-19 patients: the early New York City experience. *BMJ Supportive & Palliative Care* 2020; **0**: 1-5.
67. Gavelli F, Castello LM, Bellan M, et al. Clinical stability and in-hospital mortality prediction in COVID-19 patients presenting to the Emergency Department. *Minerva Medica* 2020: 1-18.
68. Halalau A, Imam Z, Karabon P, et al. External validation of a clinical risk score to predict hospital admission and in-hospital mortality in COVID-19 patients. *Annals of Medicine* 2021; **53**(1): 78-86.

69. Galloway JB, Norton S, Barker RD, et al. A clinical risk score to identify patients with COVID-19 at high risk of critical care admission or death: an observational cohort study. *Journal of Infection* 2020; **81**(2): 282-8.
70. Sperrin M, McMillan B. Prediction models for covid-19 outcomes. BMJ; 2020. p. m3777.
71. Zhang C, Qin L, Li K, et al. A Novel Scoring System for Prediction of Disease Severity in COVID-19. *Front Cell Infect Microbiol* 2020; **10**: 318.
72. KGM M, DG A, JB R, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine* 2015; **162**(1): W1-W73.
73. Garibaldi BT, Fiksel J, Muschelli J, et al. Patient Trajectories Among Persons Hospitalized for COVID-19: A Cohort Study. *Annals of Internal Medicine* 2021; **174**(1): 33-41.
74. Sourij H, Aziz F, Bräuer A, et al. COVID□19 fatality prediction in people with diabetes and prediabetes using a simple score upon hospital admission. *Diabetes, Obesity and Metabolism* 2020; **23**: 589–98.
75. Kazemi MA, Ghanaati H, Moradi B, et al. Prognostic factors of chest CT findings for ICU admission and mortality in patients with COVID-19 pneumonia. *MedRxiv* 2020.
76. Nunez-Gil IJ, Fernandez-Perez C, Estrada V, et al. Mortality risk assessment in Spain and Italy, insights of the HOPE COVID-19 registry. *Intern Emerg Med* 2020.
77. Gomez NFP, Lobo IM, Cremades IG, et al. [Potential biomarkers predictors of mortality in COVID-19 patients in the Emergency Department]. *Rev Esp Quimioter* 2020; **33**(4): 267-73.
78. Bello-Chavolla OY, Bahena-López JP, Antonio-Villa NE, et al. Predicting mortality due to SARS-CoV-2: A mechanistic score relating obesity and diabetes to COVID-19 outcomes in Mexico. *The Journal of Clinical Endocrinology & Metabolism* 2020; **105**(8): 2752-61.
79. Yoo E, Percha B, Tomlinson M, et al. Development and calibration of a simple mortality risk score for hospitalized COVID-19 adults. *MedRxiv* 2020.
80. Yadaw AS, Li YC, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *The Lancet Digit Health* 2020; **2**(10): e516-e25.
81. Faisal M, Mohammed MA, Richardson D, Fiori M, Beatson K. Development and validation of automated computer aided-risk score for predicting the risk of in-hospital mortality using first electronically recorded blood test results and vital signs for COVID-19 hospital admissions: a retrospective development and validation study. *MedRxiv* 2020.
82. Mei Q, Wang AY, Bryant A, et al. Development and validation of prognostic model for predicting mortality of COVID-19 patients in Wuhan, China. *Sci Rep* 2020; **10**(1): 22451.
83. Lu J, Hu S, Fan R, et al. ACP risk grade: a simple mortality index for patients with confirmed or suspected severe acute respiratory syndrome coronavirus 2 disease (COVID-19) during the early stage of outbreak in Wuhan, China. *Preprints with The Lancet* 2020.
84. Soto-Mota A, Marfil-Garza BA, Rodriguez EM, et al. The low-harm score for predicting mortality in patients diagnosed with COVID-19: A multicentric validation study. *Am Coll Emerg Physicians Open* 2020; **1**(6): 1436-43.
85. Yan L, Zhang H-T, Goncalves J, et al. An interpretable mortality prediction model for COVID-19 patients. *Nat Mach Intell* 2020; **2**: 283-8.

86. Williams RD, Markus AF, Yang C, et al. Seek COVER: Development and validation of a personalized risk calculator for COVID-19 outcomes in an international network. *MedRxiv* 2020.
87. Gue YX, Tennyson M, Gao J, Ren S, Kanji R, Gorog DA. Development of a novel risk score to predict mortality in patients admitted to hospital with COVID-19. *Sci Rep* 2020; **10**(1): 21379.
88. Das AK, Mishra S, Gopalan SS. Predicting CoVID-19 community mortality risk using machine learning and development of an online prognostic tool. *PeerJ* 2020; **8**: e10083.
89. Chen R, Liang W, Jiang M, et al. Risk Factors of Fatal Outcome in Hospitalized Subjects With Coronavirus Disease 2019 From a Nationwide Analysis in China. *Chest* 2020; **158**(1): 97-105.
90. Sarkar J, Chakrabarti P. A machine learning model reveals older age and delayed hospitalization as predictors of mortality in patients with COVID-19. *MedRxiv* 2020.
91. Miyashita N, Matsushima T, Oka M, Japanese Respiratory S. The JRS guidelines for the management of community-acquired pneumonia in adults: an update and new recommendations. *Intern Med* 2006; **45**(7): 419-28.
92. Salah HM, Sharma T, Mehta J. Smoking Doubles the Mortality Risk in COVID-19: A Meta-Analysis of Recent Reports and Potential Mechanisms. *Cureus* 2020; **12**(10): e10837.
93. Physicians RCo. National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. London: RCP, 2017.
94. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996; **22**(7): 707-10.
95. Choi KJ, Hong HL, Kim EJ. Association between oxygen saturation/fraction of inhaled oxygen and mortality in patients with COVID-19 associated pneumonia requiring oxygen therapy. *Tuberc Respir Dis (Seoul)* 2020.
96. Lim ZJ, Subramaniam A, Ponnapa Reddy M, et al. Case Fatality Rates for Patients with COVID-19 Requiring Invasive Mechanical Ventilation. A Meta-analysis. *Am J Respir Crit Care Med* 2021; **203**(1): 54-66.
97. Lippi G, Mattiuzzi C. Hemoglobin value may be decreased in patients with severe coronavirus disease 2019. *Hematol Transfus Cell Ther* 2020; **42**(2): 116-7.
98. Aujesky D, Fine MJ. The pneumonia severity index: a decade after the initial derivation and validation. *Clin Infect Dis* 2008; **47 Suppl 3**: S133-9.
99. Li P, Chen L, Liu Z, et al. Clinical features and short-term outcomes of elderly patients with COVID-19. *International Journal of Infectious Diseases* 2020; **97**: 245-50.

Table 1. Demographic and clinical characteristics for derivation and validation cohorts of patients admitted to hospital with COVID-19

Characteristics	Derivation cohort (n=3,978)		Brazilian validation cohort (n=1,054)		Spanish validation cohort (n=474)	
	Frequency (%) or median (IQR)	Non missing cases (%)	Frequency (%) or median (IQR)	Non missing cases (%)	Frequency (%) or median (IQR)	Non missing cases (%)
In hospital mortality	806 (20.3%)	3,978 (100%)	208 (19.7%)	1,054 (100%)	82 (17.3%)	474 (100%)
Age (years)	60.0 (48.0, 72.0)	3,978 (100%)	62.0 (48.2, 73.0)	1,054 (100%)	59.5 (49.0, 71.0)	474 (100%)
Sex at birth		3,976 (99.9%)		1,054 (100%)		474 (100%)
Male	2,138 (53.8%)		582 (55.2%)		276 (58.2%)	
Comorbities						
Hypertension	2,147 (54.0%)	3,978 (100%)	563 (53.4%)	1,054 (100%)	193 (40.7%)	474 (100%)
Coronary artery disease	215 (5.4%)	3,978 (100%)	60 (5.7%)	1,054 (100%)	32 (6.8%)	474 (100%)
Heart failure	269 (6.8%)	3,978 (100%)	58 (5.5%)	1,054 (100%)	23 (4.9%)	474 (100%)
Atrial fibrillation or flutter	139 (3.5%)	3,978 (100%)	27 (2.6%)	1,054 (100%)	44 (9.3%)	474 (100%)
Stroke	146 (3.7%)	3,978 (100%)	43 (4.1%)	1,054 (100%)	18 (3.8%)	474 (100%)
COPD	253 (6.4%)	3,978 (100%)	60 (5.7%)	1,054 (100%)	24 (5.1%)	474 (100%)
Diabetes mellitus	1,151 (28.9%)	3,978 (100%)	297 (28.2%)	1,054 (100%)	83 (17.5%)	474 (100%)
Obesity (BMI>30kg/m^2)	696 (17.5%)	3,978 (100%)	181 (17.2%)	1,054 (100%)	112 (23.6%)	474 (100%)
Cirrhosis	25 (0.6%)	3,978 (100%)	9 (0.9%)	1,054 (100%)	3 (0.6%)	474 (100%)
Cancer	194 (4.9%)	3,978 (100%)	65 (6.2%)	1,054 (100%)	19 (4.0%)	474 (100%)
Number of comorbidities		3,978 (100%)		1,054 (100%)		474 (100%)
0	1,189 (29.9%)		309 (29.3%)		195 (41.1%)	
1	1,173 (29.5%)		328 (31.1%)		111 (23.4%)	
2	1,013 (25.5%)		269 (25.5%)		95 (20.0%)	
3	429 (10.8%)		106 (10.1%)		53 (11.2%)	

4	131 (3.3%)		33 (3.1%)		14 (3.0%)	
≥ 5	43 (1.1%)		9 (0.9%)		6 (1.2%)	
Clinical assessment at admission						
SF ratio	428.6 (332.1, 452.4)	3,845 (96.7%)	433.3 (339.3, 452.4)	1,034 (98.1%)	459.5 (428.6, 471.4)	474 (100%)
Respiratory rate (irpm)	20.0 (18.0, 24.0)	3,236 (81.3%)	20.0 (18.0, 24.0)	870 (82.5%)	20.0 (18.0, 28.0)	452 (95.3%)
Heart rate (bpm)	88.0 (78.0, 100.0)	3,787 (95.2%)	88.0 (77.0, 100.0)	1,020 (96.8%)	95.0 (82.0, 108.0)	474 (100%)
Glasgow coma score	15.0 (15.0, 15.0)	3,695 (92.9%)	15.0 (15.0, 15.0)	982 (93.2%)	15.0 (15.0, 15.0)	466 (98.3%)
Systolic blood pressure		3,762 (94.6%)		1,014 (96.2%)		471 (99.4%)
≥ 90 (mm Hg)	3,076 (81.8%)		825 (81.4%)		466 (98.9%)	
< 90 (mm Hg)	510 (13.6%)		146 (14.4%)		5 (1.1%)	
Inotrope requirement	176 (4.7%)		43 (4.2%)		0	
Diastolic blood pressure		3,776 (94.9%)		1,022 (97.0%)		471 (99.4%)
> 60 (mm Hg)	3,541 (93.8%)		962 (94.1%)		405 (86.0%)	
≤ 60 (mm Hg)	59 (1.6%)		17 (1.7%)		66 (14.0%)	
Inotrope requirement	176 (4.7%)		43 (4.2%)		0	
Laboratory parameters						
Hemoglobin (g/L)	13.3 (12.1, 14.4)	3,871 (97.3%)	13.3 (11.9, 14.5)	1,021 (96.9%)	13.4 (12.2, 14.7)	474 (100%)
Platelet count (10^9/L)	196.0 (154.0, 257.0)	3,824 (96.1%)	203.0 (154.0, 260.2)	1,016 (96.4%)	197.5 (155.3, 257.0)	474 (100%)
NLR	4.7 (2.8, 7.8)	3,759 (94.5%)	4.9 (3.0, 8.4)	989 (93.8%)		

Lactate (mmol/L)	1.4 (1.1, 1.9)	2,742 (68.9%)	1.5 (1.2, 2.1)	720 (68.3%)		
C-reactive protein (mg/L)	77.0 (38.0, 143.0)	3,487 (87.7%)	74.1 (33.8, 143.0)	881 (83.6%)	102.4 (43.9, 189.3)	474 (100%)
BUN (mg/dL)	16.3 (11.5, 24.3)	3,636 (91.4%)	17.3 (12.9, 25.2)	942 (89.4%)	15.9 (11.7, 24.3)	474 (100%)
Creatinine (mg/dL)	0.9 (0.8, 1.2)	3,765 (94.6%)	1.0 (0.8, 1.3)	967 (91.7%)	0.8 (0.7, 1.1)	474 (100%)
Sodium (mmol/L)	137.0 (135.0, 140.0)	3,550 (89.2%)	137.0 (134.3, 140.0)	930 (88.2%)	136.0 (134.2, 138.0)	474 (100%)
Bicarbonate (mEq/L)	23.0 (21.0, 25.0)	3,222 (81.0%)	23.0 (20.6, 25.0)	807 (76.6%)	NA	NA
pH	7.4 (7.4, 7.5)	3,232 (81.2%)	7.4 (7.4, 7.5)	808 (76.7%)	NA	NA
pO2 (mmHg)	75.0 (63.0, 96.0)	3,183 (80.0%)	73.4 (63.0, 94.6)	800 (75.9%)	NA	NA
pCO2 (mmHg)	35.0 (31.3, 39.0)	3,194 (80.3%)	34.0 (30.0, 38.0)	801 (76.0%)	NA	NA

BMI: body mass index; BUN: blood urea nitrogen; COPD: chronic obstructive pulmonary disease; NA: not available; NLR: neutrophils-to-lymphocytes ratio; SF ratio: SpO_2/FiO_2 ratio

Table 2. ABC_2-SPH Score for in-hospital mortality in patients with COVID-19

	Variable	ABC_2-SPH score
A	**A**ge (years)	
	< 60	0
	60 - 69	1
	70 - 79	3
	≥ 80	5
B	**B**lood urea nitrogen (mg/dL)*	
	< 42	0
	≥ 42	3
C_2	**C**omorbidities	
	0 – 1	0
	≥ 2	1
	C reactive protein (mg/L)	
	< 100	0
	≥ 100	1
S	SF ratio (%)	
	> 315.0	0
	235.1 – 315.0	1
	150.1 – 235.0	3
	≤ 150.0	6
P	**P**latelet count ($x10^9$/L)	
	> 150	0
	100 -150	1
	< 100	2
H	Heart rate (bpm)	
	≤ 90	0
	91 – 130	1
	≥ 131	2

* When converted to urea, the cut-off is 90 mg/dL.

Table 3. Predicted mortality and mortality rates for ABC_2-SPH Score risk groups

Risk Group	Predicted mortality	Derivation cohort		Validation cohort	
		No of patients	No of deaths (%)	No of patients	No of deaths (%)
Low (0-1)	< 6%	1133	23 (2.0%)	290	1 (0.3%)
Intermediate (2-4)	6 - 14.9%	1470	168 (11.4%)	394	47 (11.9%)
High (5-8)	15 - 49.9%	907	290 (32.0%)	252	73 (29.0%)
Very high (≥9)	≥ 50%	468	325 (69.4%)	118	87 (73.7%)
Overall	-	3978	806 (20.3%)	1054	208 (19.7%)

Table 4. Discrimination and model overall performance in derivation and validation cohorts

Model	Derivation cohort		Brazilian validation Cohort	
	AUROC (95%CI)	Brier Score	AUROC (95%CI)	Brier Score
GAM	0.884 (0.879; 0.888)	0.101	0.871 (0.862; 0.879)	0.102
LASSO	0.844 (0.842; 0.846)	0.115	0.859 (0.855; 0.862)	0.110
ABC_2-SPH	0.842 (0.840; 0.843)	0.114	0.857 (0.854; 0.860)	0.108

GAM: generalized additive models; LASSO: least absolute shrinkage and selection operator logistic regression

Table 5. Main characteristics of the studies

Study	Study design	Patient time span	Country of derivation	Country of validation	Sample size (n)	Develoment sample (n) (for mortality)	Validation sample (n) (for mortality)	Develoment population	Validation population
Halalau[68]	Retrospective cohort	March 1, 2020 to April 1, 2020	United States of America	United States of America	2025	Not clear	1290	Not clear	Confirmed SARS-CoV-2 patients who required hospital admission at 8 hospitals in Beamount, excluding patients who remained hospitalized beyond May 12, 2020
Fumagalli[10]	Retrospective cohort	February 22, 2020 to April 10, 2020	Italy	Italy	516	516	NA	Consecutive adult patients with COVID-19 from 2 Italian tertiary hospitals	
Knight[36]	Prospective cohort	May 21, 2020 to June, 29 2020	England, Scotland, and Wales	England, Scotland, and Wales	57824	35463	22361	Consecutive adult patients with COVID-19 from 260 hospitals, admitted up to May 20, 2020	The same as the development population, admitted after May 20, 2020
Liang[41]	Retrospective cohort	November 21, 2019 to January 31, 2020	China	China	2300	1590	710	Patients with COVID-19 from 575 hospitals in 31 provincial administrative regions	Data from hospitals not included in the development cohort
Nicholson[51]	Retrospective cohort	First patient to May 19, 2020	United States of America	United States of America	1042	578	464	Consecutive adult patients with laboratory-confirmed COVID-19 patients from Mass General Brigham hospitals	
Garibaldi[73]	Retrospective cohort	March 4, 2020 to April 24, 2020, with follow-up through June 27, 2020	United States of America	United States of America	832	832	NA	Consecutive confirmed COVID-19 patients from 5 hospitals (John Hopkins Medicine)	

Sourij[74]	Prospective and retrospective cohort	April 15, 2020 to June 30, 2020	Austria	NA	238	238	NA	Adult patients with confirmed COVID-19 and diabetes or pre-diabetes	NA
Gavelli[67]	Retrospective single-center cohort	March 16, 2020 to April 22, 2020	Italy	Italy	480	Apparently, it was developed by expert consensus	480	NA	Adult patients with confirmed COVID-19 patients admitted to one university hospital
Kazemi[75]	Retrospective cohort	February 25, 2020 to April 25, 2020	Iran	NA	91	91	NA	Adult patients with confirmed COVID-19 who had undergone CT scan <8 days from the beginning of symptoms, excluding the ones with RT-PCR more than 7 days from CT. CT score developed not based on the data. Authors tested CT score and clinical variables in a model	NA
Núñez-Gil[76]	Retrospective cohort	February 8, 2020 to April 1, 2020	Spain and Italy	NA	908	908	NA	Patients with confirmed COVID-19 from centers in Italy (n=88) and Spain (n=820)	
Allenbach[14]	Prospective single-center cohort	March 16, 2020 to April 4, 2020	France	France	152	152	131	Adult patients with confirmed COVID-19 from one tertiary care university hospital	Not described
Kim[15]	Retrospective single-center cohort	February 19, 2020 to March 15, 2020	Korea	NA	38	38	NA	Adult patients with confirmed COVID-19 admitted to a tertiary university hospital	NA
Altschul[65]	Retrospective single-center cohort	March 1, 2020 to April 16, 2020	United States of America	United States of America	4711	2355	2356	Patients with confirmed COVID-19 from an academic hospital	The same as the development population (spitted 50/50%, apparently by admission date)

Hajifathalian[44]	Retrospective cohort	March 4, 2020 to April 9, 2020	United States of America	United States of America	929	664	265	Adult patients with confirmed COVID-19 patients presenting to emergency department of 2 hospitals in Manhattan (did not exclude patients who were discharged within 24 hours)	Adult patients with confirmed COVID-19 patients presenting to emergency department of 9 hospitals in Massachusetts (did not exclude patients who were discharged within 24 hours)
Wang[13]	Retrospective single-center cohort	January 28, 2020 to March 4, 2020	China	China	243	199	44	Adult patients with confirmed COVID-19 from one university hospital	The same as the development population (the criteria used to divide patients in training and testing sets was not clear)
Zhou[16]	Retrospective single-center cohort	January 12, 2020 to February 26, 2020	China	NA	118	118	NA	Elderly patients (>60 years) with "clinically diagnosed" COVID-19 (RT-PCR or chest CT) from one university hospital	NA
Goméz[77]	Retrospective single-center cohort	February 24, 2020 to March 16, 2020	Spain	NA	163	163	NA	Adult patients with suspected COVID-19 admitted to one university hospital	NA
Galloway[69]	Retrospective cohort	March 24, 2020 to April 17, 2020	England	NA	1157	1157	NA	Patients with confirmed COVID-19 from 2 academic hospitals	NA
Bello-Chavolla[78]	Registry data from na open source database from the Mexican Ministry of Health	First patient up to May 18, 2020	Mexico	Mexico	51633	41307	10326	Patients with confirmed COVID-19 from the open source Mexican Ministry of Health database (inpatients and outpatients)	The same as the development population (split by random sampling stratified by mortality status)
Weng[52]	Retrospective cohort	January 1, 2020 to February 15, 2020	China	China	301	176	125	Adult patients with laboratory-confirmed COVID-19 from 2 hospitals	The same as the development population (the criteria used to divide

									patients in training and testing sets was not clear)
Ko[52]	Retrospective cohort	Development cohort: January 10, 2020 to February 24, 2020; Validation cohort: February to July 2020	China	China	467	361	106	Patients with COVID-19 (not clear if laboratory-confirmed) from one hospital, excluding 14 patients without a blood test within 1 day after the hospital admission	Patients with COVID-19 (not clear if (laboratory-confirmed) from 3 hospitals
Xie[37]	Retrospective cohort	January and February 2020	China	China	444	299	145	Patients with confirmed COVID-19 from one hospital in Wuhan who had been discharged or died	Patients with confirmed COVID-19 from another hospital in Wuhan, excluding 6 patients who died quickly
Yoo[79]	Retrospective cohort	March 1, 2020 to April 28, 2020	United States of America	United States of America	4.840	1.613	1.614	Adult patients with confirmed COVID-19 from 5 hospitals, up to 99 years-old. The sample was randomly split in 3 datasets, the second one was used for development	The same as the development population: andomly split in 3 datasets, the third one was used for validation
Zhang[38]	Retrospective cohort	Not reported	China	United Kingdom	1001	775	226	Adult patients with confirmed COVID-19 from one hospital	Adult patients with confirmed COVID-19 from another hospital
Yadaw[80]	Retrospective and prospective cohort	March 9, 2020 toApril 7, 2020	United States of America	United States of America	5051	3841	961	Inpatients and outpatients (including those attended by telehealth) with confirmed COVID-19 from the Mont Sinai Health System (8 hospitals and over 400 ambulatory practices) until April 6, 2020	The same as the development population (randomly split 80/20%) and patients admitted to Mont Sinai Hospitals who were included in the database (with the outcome) on April 7, 2020
Shang[54]	Retrospective Cohort	January 1, 2020 to March 27,2020	China	China	452	113	339	Consecutive patients with confirmed COVID-19 from 2 hospitals in Wuhan, who had severe or critical illness	The same definition as the development population, but from a third hospital in Wuhan

Faisal[81]	Registry data	March 11, 2020 to June 13, 2020	United Kingdom	United Kingdom	6444	3924	2520	Consecutive adult non-elective or emergency medical admissions (COVID-19 and non-COVID-19 patients) from one hospital, who were discharged over a course of three months and had electronic NEWS2 recorded	Consecutive adult non-elective or emergency medical admissions (COVID-19 and non-COVID-19 patients) from another hospital, who were discharged over a course of three months and had electronic NEWS2 recorded
Mei[82]	Retrospective cohort	January 21, 2020 to February 27, 2020	China	China	492	237	Validation 1 = 120 and validation 2 = 135	Adult patients with confirmed COVID-19, diagnosed with pneumonia by CT scan, from one hospital in Wuhan. Patients who died within the first 24 hours, with not clinical outcome available or who refused to participate were excluded	The same as the development population, from other 3 hospitals
Zhang[8]	Retrospective cohort	January 12, 2020 to February 9, 2020	China	China	828	516	312	Adult patients with confirmed COVID-19 from one hospital	Adult patients with confirmed COVID-19 from the same hospital in a different time span (February 8-9, 2020) and from another hospital
Lu[83]	Retrospective single-center cohort	January 21, 2020 to February 5, 2020	China	NA	577	577	NA	Patients with confirmed or suspected COVID-19 from one hospital	NA
Soto-Mota[84]	Retrospective Cohort	April 30, 2020 to May 20, 2020	Mexico	NA	400	Score developed by consensus	400	NA	Consecutive patients with confirmed COVID-19 from 12 hospitals, with complete clinical information and outcome

Yan[85]	Retrospective cohort	Development cohort: January 10, 2020 to February 18, 2020; Validation cohort: February 19-24, 2020	China	China	485	375	110	Adult patients with COVID-19 (not clear if patients had laboratory-confirmed disease), from one hospital, excluding patients with >20% missing values and breast-feeding women	The same as the development population, admitted after February 18, 2020
Williams [86]	Retrospective cohort	Development cohort: any time prior to 2020; validation cohort: January 1st 2020 to April 20, 2020	United States of America, South Korea, Spain, Australia, Japan, Netherlands	South Korea, Spain, United States of America	2,126,784	2,082,277	44.507	Healthcare database of 6 countries, in which adult patients with GP, EP or OP visit with influenza or flu-like symptoms, at least 365 days of prior observation, and no symptoms in the preceding 60 days	Adult patients with confirmed with COVID-19, presenting at an initial healthcare provider interaction in a GP, ER or OP visit, and who had no diagnosis of influenzae or pneumonia and no flu-like symptoms in the preceding 60 days
Gue[87]	Retrospective single-center cohort	March 10, 2020 to May 30, 2020	United Kingdom	NA	316	316	NA	Consecutive patients with confirmed COVID-19 from a general hospital, who had clinical symptoms at admission	NA
Das[88]	Retrospective cohort	January 20, 2020 to May 30, 2020.	South Korea	South Korea	3,524	3,524	NA	Data shared by Korea Centers for Disease Control and Prevention, from 17 provinces. Patients with confirmed COVID-19, with availability of demographic, exposure and diagnosis confirmation features along with the outcome	NA

Levy[59]	Retrospective and prospective cohort	March 1, 2020 to May 12, 2020	United States of America	United States of America	8391	6162	2229	Adult patients with confirmed COVID-19 from 11 acute care hospitals in New York, from March 1, 2020 to April 23, 7 2020. Patients were excluded if they were still in the hospital at the study end point with a length of stay less than 7 days; if they were transferred to a hospital outside of the health system and their outcomes were unknown; or if they expired but were not marked as discharged in the EH	The same as the development cohort from another hospital in New York from March 1, 2020 to May, 7 2020, and all 12 hospitals from April 24, 2020 to May 6, 2020.
Chen[89]	Retrospective cohort	The first patient to January 31, 2020	China	China	1590	1590	NA	Patients with confirmed COVID-19 from 575 hospitals throughout China, excluding cases with incomplete medical records (20.8%)	NA
Sarkar[90]	Registry data	13th January, 2020 to 28th February, 2020	22 countries in Asia, Australia, Europe and North America	NA	115	115	NA	Open source databased of COVID-19 patients (inclusion criteria is not clear)	NA
Hu[55]	Retrospective cohort	28 January 2020 and 11 March 2020	China	China	247	183	64	Patients with severe confirmed COVID-19 infection admitted to one hospital in Wuhan. patients who had >10% missing values, stayed in the hospital <7□days, were afflicted by a severe disease before admission (e.g. cancer, aplastic anaemia or uraemia), were unconscious at admission or were directly admitted to the intensive care unit (ICU) were excluded	The same as the development population, admitted at another hospital

Table 5. Continued

Study	Model outcome	Outcome time	Original modelling approach	Imputation	Use of AI techniques	Was a score produced?	Number of variables were tested in the development cohort	Univariate analysis	How many patients died in the development dataset?
Halalau[68]	Hospital admission and in-hospital mortality	In-hospital	Multivariate logistic regression	No	No	Yes	Not clear	No	Not clear
Fumagalli[10]	Mortality	In-hospital	Cox regression analysis	No	No	Yes	20	Yes	120
Knight[36]	Mortality	In-hospital	LASSO logistic regression	Yes. Multiple imputation with chained equations	Yes. ML	Yes (4C mortality score)	21	No	11426
Liang[41]	Composite of ICU admission, need of invasive mechanical ventilation or death	In-hospital	LASSO logistic regression	Yes (if <20%). Predictive mean matching to impute numeric features, logistic regression to impute binary variables, and Bayesian polytomous regression to impute factor features	No	Yes (COVID-GRAM)	72	No	51 (3.2%)
Nicholson[51]	Need of mechanical ventilation and in-hospital mortality	In-hospital	Multivariate logistic regression	No	No	Yes: one to predict ventilation need (VICE score) and another one for death (DICE score)	49	Yes	Not reported
Garibaldi[73]	In-hospital mortality and a composite of disease severity (WHO scale) or	In-hospital	Cox regression analysis	Yes. Imputed missing values by chained equations (MICE) with predictive mean matching	Yes. NLP was used to identify presenting symptoms	Yes: COVID-19 Inpatient Risk Calculator (CIRC)	24	No	131

	in-hospital mortality								
Sourij[74]	Mortality	In-hospital	Multivariate logistic regression	No	No	Yes	Not clear	Yes	58
Gavelli[67]	In-hospital mortality and in-hospital clinical stability	In-hospital	Multivariable logistic regression and Cox Regression Hazard models	No	No	Yes (NOVARA score)	NA	No	NA (consensus)
Kazemi[75]	Mortality	In-hospital	Multivariate logistic regression	No	No	Yes (authors created a CT score not based on the data)	Not available	No	11
Núñez-Gil[76]	Mortality	In-hospital	Multivariate logistic regression	No	No	Yes	Not clear	Yes	311
Allenbach[14]	Composite of ICU admission or death	14 days	Multivariate logistic regression	No	No	Yes	42	Yes	32
Kim[15]	Mortality	In-hospital	Consensus	No	No	Yes	3	No	7
Altschul[65]	Mortality	In-hospital	Multivariate logistic regression	No	No	Yes	Not clear	Yes	621
Hajifathalian[44]	Mortality	7 days and 14 days	Multivariable logistic regression	Yes. Imputation by chained equations	No	Yes (COVID-AID)	38	Yes	93

Wang[13]	Mortality	28 days	Multivariable logistic regression	No	No	Yes (FAD-85)	41	No	24
Zhou[16]	Mortality	In-hospital	Multivariable logistic regression	No	No	Yes (NLAUD)	37	No	51
Goméz[77]	Mortality	30 days	Multivariable logistic regression	No	No	Yes (COVEB)	20	No	33
Galloway[69]	Composite of transfer to ICU or death	In-hospital	LASSO logistic regression	No	No	Yes	19	No	244
Bello-Chavolla[78]	Mortality	30 days	Cox proportional risk regression analysis	No	No	Yes	12	No	4276
Weng[52]	Mortality	In-hospital	LASSO logistic regression	Yes, for variables with <10% missing values (>10% were excluded from model development). RF.	No	Yes (ANDC)	24	No	21
Ko[52]	Mortality	In-hospital	Machine learning techniques	Yes, imputed with mean values for development and trainig datasets	Yes, DLN and RF model	Yes (EDRnet)	73	Yes	212 (58.7%)
Xie[37]	Mortality	In-hospital	Multivariate logistic regression	No	No	Yes	28	No	155
Yoo[79]	Mortality	In-hospital	Gray`s K-sample tests, DeLong's test	No	No	Yes	48	Yes	Not reported

Zhang[38]	Death and poor outcome (developing ARDS, receiving intubation ou ECMO treatment, ICU admission or death)	In-hospital	LASSO logistic regression	No	No	Yes (DCS, DCSL, DL)	19	No	33 (4.3%)
Yadaw[80]	Mortality	In-hospital	Artificial inteligence techniques	Yes, using means	Yes. Recursive feature elimination method for feature selection, and logistic regression, SVM, RF model, and XGBoost algorithms for prediction	Yes (17F and 3F models)	17	No	313 (8.15%)
Shang[54]	Mortality	In-hospital	LASSO logistic regression	Yes, multiple imputation methods for variables with <10% missing values	No	Yes (CSS score)	52	No	49
Faisal[81]	Mortality	In-hospital	Multivariable logistic regression	No	No	Yes (CARMc19_N and CARMc19_NB)	Not clear	No	323
Mei[82]	Mortality	In-hospital	LASSO logistic regression	No	No	Yes	43	No	105

Zhang[8]	Mortality	14 days and 28 days	Cox regression analyses	Yes. Multiple imputations (method not reported)	No	Yes	30	No	96
Lu[83]	Mortality	12 days	Cox regression analysis	No	No	Yes	Not clear	Yes	39
Soto-Mota[84]	Mortality	In-hospital	Consensus	No	No	Yes (LOW-HARM)	NA	No	200 (50%)
Yan[85]	Mortality	In-hospital	Machine learning techniques	No	Yes, XGBoost machine learning algorithm	Yes	75	No	174
Williams [86]	Hospitalization with pneumonia, hospitalization with pneumonia requiring intensive services or death and death in the 30 days after index date	In-hospital and 30 days after index rate	LASSO logistic regression	No	Yes, ML (train-test-split)	Yes, 3 scores (COVER-F for death)	31,917	No	11407
Gue[87]	Mortality	30 days	Multivariable logistic regression	No	No	Yes (COVID-19 Mortality Socre)	15	No	145
Das[88]	Mortality	In-hospital	Logistic regression and machine learning techniques	No	Yes. SVM, K nearest neighbor, RFM and gradient boosting	Yes (CoCoMoRP)	4	No	74
Levy[59]	Mortality	7 days	LASSO logistic regression	Yes, imputation of means. Variables with >50% missing values were excluded.	No	Yes (NOCOS Calculator)	42	No	Not clear

Chen[89]	Mortality	14, 21 and 28 days	Multivariate Cox regression analysis	No	No	Yes (nomogram)	37	No	50
Sarkar[90]	Mortality	In-hospital	Machine learning techniques	No	Yes, RF classification algorithm	Yes	6	No	37
Hu[55]	Mortality	In-hospital	LASSO logistic regression	Yes, using bagging tree. Variables with >30% missing values were excluded	Yes. Logistic regression, PLS regression, EN model, random forest and bagged flexible discriminant analysis (FDA).	Yes	51	No	68

Table 5. Continued

Study	Variables included in the final model (for mortality)	External validation	How are predictors combined?	AUC in derivation cohort	AUC in validation cohort	Limitations
Halalau[68]	Age, male sex, congestive heart failure, end-stage renal disease, chronic pulmonary disease, DM, hypertension, obesity, nursing home residence, immunocompromised status, congenital heart disease, coronary artery disease, end-stage liver disease and pregnancy	Yes	Points-based score	Not available	0.75 (0.71 – 0.78)	Selection bias: Excluded patients who were hospitalized beyond May 12, 2020. Data on how the score was developed not reported. Absence of an initial validation cohort. Uniform scoring weights of different risk factors. Complete case analysis.
Fumagalli[10]	Age, number of comorbidities (CV disease, hypertension, DM, depression, dementia and cancer), respiratory rate, PaO2/FiO2, serum creatinine and platelet count obtained on admission	No	Points-based score	0.90 (0.87 - 0.93)	NA	Modest sample size. No external validation. Variables were selected by univariate analysis. Complete case analysis.
Knight[36]	Age, sex, number of comorbidities (chronic cardiac disease, chronic respiratory disease excluding asthma, chronic renal disease defined as estimated glomerular filtration rate ≤30, mild to severe liver disease, dementia, chronic neurological conditions, connective tissue disease, DM, HIV or AIDS, and malignancy), respiratory rate, SpO2, level of consciousness, urea and CPR obtained on admission	Yes	Points-based score	0.786 (0.781 - 0.790)	0.767 (0.760 - 0.773)	Several potentially relevant comorbidities, such as hypertension, previous myocardial infarction, and stroke, were not included in data collection. The authors considered that inclusion of these comorbidities might have impacted upon or improved the performance and generalizability of the 4C Mortality Score. Secondly, a proportion of recruited patients (3.3%) had incomplete episodes, so there is a possibility of selection bias, if patients with incomplete episodes, such as those with prolonged hospital admission, had a differential mortality risk to those with completed episodes.

Liang[41]	Chest radiographic abnormality, age, hemoptysis, dyspnea, unconsciousness, number of comorbidities (COPD, hypertension, DM, coronary heart disease, chronic kidney disease, cancer, cerebrovascular disease, hepatitis B, immunodeficiency), cancer history, neutrophil-to-lymphocyte ratio, lactate dehydrogenase and direct bilirubin obtained on admission	Yes	Logistic Regression	0.88 (0.85 - 0.91)	0.88 (0.84 - 0.93)	Modest sample size for score development and a relatively small sample for validation. The data for score development and validation are entirely from China, which could potentially limit the generalizability of the risk score in other areas of the world. Mortality was quite low (3.2%). Apparently, patients with cancer should gain points for both cancer history and number of comorbidities, not clear.
Nicholson[51]	Age, sex, diabetes mellitus, chronic statin use, albumin, C-reactive protein, neutrophil-lymphocyte ratio, mean corpuscular volume, platelet count, and procalcitonin obtained on admission	Yes	Logistic Regression	0.87 (0.83 – 0.91)	0.80 (0.75 – 0.85)	Modest sample sizes in both our derivation and validation cohorts. The number of events on the derivation and validation cohort separately was not informed (211 in total). Variables were selected by univariate analysis. Complete case analysis.
Garibaldi[73]	Age, nursing home residence, sex, BMI, Charlson Comorbidity Index, SaO2/FiO2 ratio obtained on admission	No	Cox regression analysis	Not available	Not available	Modest sample size. No external validation. Too many variables tested in the model for the number of events (24/131). To try to overcome that, authors tested variables "in blocks"
Sourij[74]	Age, arterial occlusive disease, CRP, estimated GFR and aspartate AST levels obtained on admission	No	Nomogram	0.889 (0.837 - 0.941)	NA	Small sample size and number of events. Number of variables tested not clear. Complete case analysis, and predictors with >20% missing values were excluded. No external validation
Gavelli[67]	Presence of comorbidity (any disease on active therapy), SpO2 and respiratory rate after a trial of 15 minutes with oxygen at a FiO2 0.5	No	Points-based score	NA	Not reported	Score developed by consensus. Modest sample size. Number of events is not clear. Single-center study. No external validation. AUC and accuracy not presented.

Kazemi[75]	Age, sex, comorbidity (cardiovascular and pulmonary), diffused distribution of CT abnormality, total CT-score and dyspnea at admission	No	Logistic Regression	0.73 (95% CI not reported)	NA	Small sample size and number of events. Too many variables tested for the low number of events. Comorbidities were not well defined, percentage of involvement included in CT score is subjective and peripheral involvement is not well defined. Complete case analysis. High risk of selection bias: All 3 hospitals were referral centers for COVID-19 patients, so it is possible that the overall CT- score of the patients in this study would not be representative of the general population
Núñez-Gil[76]	Age, hypertension, obesity, renal insufficiency, any immunosuppressive condition, SpO2, CRP obtained on admission	No	Points-based score	0.88 (0.85 – 0.91)	NA	No external validation. Variables were selected by univariate analysis. Complete case analysis. Variables included in the model not clearly defined. Authors reported that some incident events in the participating centers may not have been diagnosed and/or not been reported. The data analysis and modeling focused on only two countries (Italy and Spain) of the four initially considered, since as previously mentioned heterogeneity among countries with regard to clinical features and death-risk assessment could limit the representative nature of the sampling.
Allenbach[14]	Age, WHO clinical scale, CRP and lymphocytes count obtained on admission	No	Points-based score (but AUC presented based on the logistic regression model)	0.786 for the composite outcome and 0.803 for death (after correction for over-optimism; IC95% not reported)	0.787 for the composite outcome and 0.827 for death (after correction for over-optimism; IC95% not reported)	Small sample size of both development and validation samples. Too many predictors tested for small number of events. Complete case analysis. External validation sample not described. The external sample consisted of patients from a regional non-university hospital, which could explain the differences on catchment area and patient recruitment. In the acute context of the first SARS-CoV-2 epidemic wave in France, we relied on a sample prospectively defined by consecutive eligible patients in the study center.

Kim[15]	Myocardial damage marker (creatine kinase-MB [CK-MB] or troponin-I > the 99th percentile upper reference limit) + Heart failure marker (NT-proBNP ≥ 125 pg/mL) + Electrical abnormality marker (first detected or newly developed supraventricular tachycardia, ventricular tachycardia, ventricular fibrillation, atrial fibrillation, bundle branch block, ST-segment elevation/depression, T-wave flattening/inversion, and QT interval prolongation on ECG)	No	Points-based score	Not reported	NA	Score developed by consensus. Small sample size and small number of events. Accuracy not assessed. The protocol for the evaluation of cardiac injury was not controlled. The attending physician decided each category of the test according to the patient's condition at the time of the management. When the test was not performed, it is assumed as a negative result because the physician considered it as an unnecessary test or the result might be negative.
Altschul[65]	Age, sex, SpO2, MAP, INR, creatinine, BUN, interleukin-6 (IL-6), CRP and procalcitonin obtained on admission	Yes	Points-based score	0.824 (0.814 to 0.851)	0.798 (0.789 to 0.818)	Complete case analyses, variables selected by univariate analyses
Hajifathalian[44]	Age, mean arterial pressure, serum creatinine and severity of hypoxia at hospital presentation.	Yes	Multivariate logistic regression	7 days: 0.877 (95%CI 0.831–0.923); 14 days: 0.847 (95%CI 0.806–0.888)	7 day (0.851 [0.781 to 0.921]); 14 day (0.825 [0.764 to 0.887])	Modest sample size for development and validation, less than 100 events both in the development and validation cohorts, short follow-up time
Wang[13]	Age, ferritin and D-dimer obtained on admission	Yes	Logistic regression and nomogram	0.871 (based on its optimal cut-off value = 85)	Not available (link for supplemental material does not work)	Single-center study, with small sample for development and validation, less than 100 events both in the development and validation cohorts. Complete-case analysis. D-dimer assay not described. AUC for external validation not available to the readers.
Zhou[16]	Lactate dehydrogenase, albumin, BUN, NLR and D-dimer obtained on admission	No	Nomogram	0.955 (95% CI not provided)	NA	Single-center study, with small sample size, including cases not confirmed by RT-PCR, and less than 100 events. Complete-case analysis and tests too many variables for the number of events. D-dimer assay not

						described.
Goméz[77]	Age, creatin, glucose and white blood cells obtained on admission	No	Not clear	0.874 (0.816-0.933)	NA	Single-center study, with small sample size, including cases not confirmed by RT-PCR, and less than 100 events. Complete-case analysis and tests too many variables for the number of events.
Galloway[69]	Age, sex, ethnicity, DM, hypertension, chronic lung disease, SpO2, radiographic severity score, neutrophil count, respiratory rate, CRP, albumin, creatinine obtained on admission	No	Points-based score	0.697 (0.652,0.741)	NA	Modest sample size. No external validation. Complete case analysis. AUC < 0.70
Bello-Chavolla[78]	Age, diabetes, obesity, CKD, COPD, hypertension, immunosuppression and COVID-19 pneumonia	Yes	Points-based score	0.823 (95% CI not reported)	0.830 (95% CI not reported)	The use of data collected from a sentinel surveillance system model, what raises concern about data quality. The same score for inpatient and outpatients and sensitivity analysis was not performed to assess accuracy for patients who were hospitalized. Apparently, complete case analysis.
Weng[52]	Age, neutrophil-to-lymphocyte ratio, D-dimer and C-reactive protein obtained on admission	Yes	Nomogram and logistic regression	0.921 (0.835-0.968)	0.975 (0.947-1.0)	Small sample size for development and validation, with <100 events in both cohorts. Variables with >10% missing values were excluded. D-dimer assay was not reported.
Ko[52]	Lymphocytes, neutrophils, albumin, LDH, neutrophil count (?), CRP, prothrombin activity, calcium, urea, estimated GFR, monocytes, globulin, eosinophils, glucose, RDW, bicarbonate, RDW standard deviation, platelet count, mean platelet volume, platelet large-cell ratio, prothrombin time, total protein, platelet distribution width, aspartate aminotransferase,	Yes	AI model	Not reported	Not reported	Small sample size for development and validation, too many variables tested for the limited number of events, high mortality rate, with possibility of selection bias. Not clear if included laboratory-confirmed COVID-19 patients only. The number of predictors make it difficult to be applicable at bedside.

	thrombocytocrit, eosinophil count, alkaline phosphatase, INR					
Xie[37]	Age, lymphocyte count, lactate dehydrogenase and SpO2 obtained on admission	Yes	Logistic regression and nomogram	0.880 (95% CI not reported)	0.980 (0.958-1.00)	High risk of selection bias: the cohort was conducted early in the pandemic, there was a high mortality rate (51.8% in development cohort and 47.6% in the validation cohort), and it may not accurately represent patients with mild or asymptomatic COVID-19 (as they were not being tested). Small sample size for development and validation, less than 100 events both. Complete case analysis.
Yoo[79]	Glasgow coma scale, oxygen support level, BUN, age, lymphocyte percentage, troponin	Yes	Points-based score	Not reported, as AUC was used to define the variables for the score.	At admission 0.81; maximum through admission 0.91; mean through admission0.92	The authors reported that documentation of all kinds was inconsistent during the first wave of covid-19 and the environments at different hospitals varied substantially. While it is unlikely that a laboratory result or medication administration was missed, inconsistencies in flowsheet documentation during this period could mean that the timings of different modes of oxygen administration were not always accurately capture. The statistical test used to produce the score is not adequate according to the TRIPOD and may lead to overoptimism.
Zhang[38]	DCS (demographic, comorbidities and symptoms): age, sex, chronic lung disease, DM, hypertension, immunosuppression, cancer, CKD, heart disease, cough, dyspnea, diarrhea; DCSL (demographic, comorbidities, symptoms and laboratory tests): age, sex, chronic lung disease, DM, cancer, cough, dyspnea, CRP, creatinine, platelets, neutrophils and lymphocytes counts; DL (demographic and laboratory	Yes	Logistic regression	DCS: 0.79; DCS: 0.89; DL: 0.91 (95% CI not reported)	DL: 0.74 (95% CI not reported)	Authors reported that clinical datasets were collected when healthcare services were under severe strain. Data extraction sought to ensure consistency and accuracy, but there is missing data in both datasets and the analysis was complete case based. Sample sizes for development and validation were small, with <100 events. Clinical assessments at admission such as SpO2 were not available in either dataset. The external validation dataset has very different case-mix, and only had follow-up to a fixed date (6-39 days). Although the Wuhan cohort includes many people with less severe disease, in the validation

	tests): age, sex, CRP, creatinine, platelets, neutrophils and lymphocytes counts (around admission)					cohort most admitted patients are likely to have severe disease. Although the authors reported all variables were included in the model, for most of the included ones the 95% CI of the OR included 1.0
Yadaw[80]	17F: age, sex, ethnicity, encounter type, temperature, diastolic blood pressure, oxygen saturation at presentation, minimum oxygen saturation, smoking, asthma, COPD, obesity, DM, HIV, cancer; 3F: age, minimum oxygen saturation, and type of patient encounter, obtained the day of admission	Yes	Artificial intelligence (XGBoost)	0.91 (95% CI not provided)	0.91 (95% CI not provided)	As it includes inpatients and outpatients, important laboratory parameters were not tested. The authors reported that the clinical features available were limited to those routinely collected during hospital encounters, and they pointed out that development of even better prediction models should be possible using a richer set of features.
Shang[54]	Age, coronary heart disease, % of lymphocytes, procalcitonin, D-dimer	Yes	Points-based score	0.919 (95% CI 0.870-0.970)	0.938 (95% CI 0.902-0.973)	Small sample size in development (113 participants) and validation cohorts, with <100 events in the development one. Too many variables tested for the number of events.
Faisal[81]	CARMc19_N: 10 [age, sex, COVID-19 (yes/no), NEWS2 score and subcomponents] and CARMc19_NB: 18. All variables from CARMc19_N + 7 blood test results + AKI score	Yes	Points-based score	CARMc19_NB = 0.87 (95% CI 0.85-0.89) vs CARMc19_N 0.86 (95% CI 0.84-0.87)	CARMc19_NB = 0.88 vs CARMc19_N = 0.86	Not exclusively for COVID-19 patients. COVID-19 was identified by ICD-10 code which depends on clinical judgment. Risk of selection bias, as only patients with NEWS2 recorded were included. Complete case analysis.
Mei[82]	Age, NLR, admission body temperature, AST, total protein	Yes	Points-based score	0.912 (95% CI 0.878-0.947)	VC1 = 0.928 (95% CI 0.884-0.971) and VC2 = 0.883 (0.815-0.952)	Risk of selection bias due to inclusion/exclusion criteria, included only patients from Wuhan. Small sample size for development and validation. Complete case analysis.

Zhang[8]	Age, LDH, NLR and direct bilirubin obtained on admission	Yes	Nomogram	0.886 (95% CI 0.873–0.899)	0.879 (95% CI, 0.856–0.900) and 0.839 (95% CI [0.798–0.880) for each one of the hospitals	Small sample size for evelopment and validation, <100 events for both cohorst. The amount of missing data differed between the survivor and non-survivor groups. The study included a high population of patients who were severely ill, the authors pointed out there may be a selection bias when identifying the risk factors of mortality
Lu[83]	Age, CPR	No	Cox regression analysis, decision tree	Not reported	NA	Included both patients with confirmed and not confirmed disease, small sample size with <100 events, number of potential predictors tested was not clear. No external validation.
Soto-Mota[84]	Age, hypertension, white blood cell count, lymphocyte count, myocardial necrosis marker, creatinine, SpO2 (not clear in which moment)	No	Logistic regression	NA	Provided by different cut-offs, ranging from 0.61 to 0.90 (95% ranges from 0.59 to 0.93), with best AUC for 25 points (0.90 [95% CI 0.87-0.93])	Score developed by consensus. Not clear the moment it is meant to be used. Risk of selection bias, high mortality in the cohort (50%)
Yan[85]	LDH, lymphocytes and CRP obtained at hospital admission	Yes	Multi-tree XGBoost model	0.978 (IC 95% not provided)	0.951 (CI 95% not provided)	Single-center study, with small sample for development and validation, less than 100 events in the validation cohort. Apparently, complete-case analysis.

Williams[86]	Age, sex, history of cancer, COPD, diabetes, heart disease, hypertension, hyperlipidemia and kidney disease.	Yes	Points-based score	0.896 (95% CI 0.72 - 0.90)	CUIMC database 0.820 (95% CI 0.796-0.840); HIRA database 0.898 (95% CI 0.857-0.940); SIDIAP 0.895 (95% CI 0.881-0.910); VA 0.717 (0.642-0.791)	The auhtors reported they were unable to develop a model on COVID-19 patient data due the scarcity of databases that contains this information in sufficient numbers. Based on seconday data, with possibility of misclassifications of predictors (diseases is incorrectly recorded in a patient's history, incorrect recording of influenza or COVID-19, and authors were unable to include some suspected diseases predictors such as BMI/obesity in the analysis due the inconsistency with which these mesures are collected and reporte across the databases included in the study. Patients may day after 30 days, and this will be recorded as a non-event. Apparently, complete case analysis.
Gue[87]	Age, sex, hypertension, coronary artery disease, heart failure, atrial fibrilation, oral anticoagulants, modified sepsis-induced coagulopathy (mSIC) score (INR, platelet count, qSOFA score)	No	Points-based score	0.793 (95% CI 0.745–0.841)	NA	Small sample size from a single center, no external validation. Complete case analysis. Authors pointed out that patients at the highest risk may be deemed too sick for maximal intervention and may be denied ICU treatment; predictors and their assigned weights in the final model.
Das[88]	Age, sex, province (in South Korea) and exposure (nursing home, hospital, religious gathering, call center, community center, shelter and apartment, gym facility, overseas inflow, contact with patients and others)	No	Logistic regression (SMOTE)	0.830 (95% CI not reported)	NA	Risk of selecion bias (only patients wit complete data were included), unavailability of crucial clinical information on symptoms, risk factors and clinical parameters. Less than 100 events. No external validation
Levy[59]	Age, length of stay, SpO2, neutrophil, RDW, sodium urea (on admission and every 2 days)	Yes	Logistic regression	0.86 (95% CI not reported)	0.82 (95% CI not reported)	Data were imputed for variables with up to 50% missing values. Follow up was too short (7 days), what causes a high risk of bias, as a significant proportion of patients may die after 7 days. Authors did not show how to calculate the score.

Chen[89]	Age, coronary heart disease, cerebrovascular disease, dyspnea, procalcitonin, aspartate aminotransferase, total bilirubin upon admission	No	Nomogram	0.91 (95% CI, 0.85-0.97)	NA	High risk of selectionbias (20.8% patients with incomplete data were excluded), modest sample size, with <100 events. No external validation. Complete case analysis. Authors did not show how to calculate the score.
Sarkar[90]	Age, sex, from Wuhan, visit to Wuhan, days from symptom onset to hospitalization	No	RF classificatio n algorithm	0.97 (95% CI not reported)	NA	Small sample size, with <100 events. High risk of selection bias: from 1085 patients, 652 (60.1%) were excluded due to missing values, and the model was developed using one 115 patients(10.6%). Data quality is questionable, as the study is based in open source database.
Hu[55]	Age, CRP, D-dimer, lymphocyte count at admission	Yes	Points-based score	0.895 (95% CI not reported)	0.881 (95% CI not reported)	Small sample size of both development and validation samples, with <100 events. Too many predictors tested for a small number of events. The authors did not exclude patients transferred from other hospitals (so the assessment was not the first hospital assessment in all patients). Single center study, patients from both derivation and validation sets were from Tongji Hospital, which is one of the hospitals with a high level of medical care in China (the authors reported that some critically ill patients who recovered there might die in other hospitals with suboptimal or typical levels of medical care).

AUC: area under the curve; BMI: body mass index; CI: confidence interval; CPOD: chronic obstructive pulmonary disease; CPR: C-reactive protein; CT: computed tomography; DLN: deep learning networks; DM: diabetes mellitus; GFR: glomerular filtration rate; ICU: intensive care unit; LASSO: least absolute shrinkage and selection operator logistic regression; NA: not applicable; RDW: red blood cell distribution width; PLS: partial least squares RF: Random Forest; SF ratio: SpO_2/FiO_2 ratio; SVM: support-vector machine; XGBoost: eXtreme Gradient Boosting; WHO: World Health Organization.

Table 6. Discrimination of risk scores within validation cohort (complete case)

Score	Number of patients	Number of deaths (%)	AUROC (95%CI)
A-DROP	704	148 (21%)	0.780 (0.740-0.820)
ABC_2SPH	779	148 (19%)	0.853 (0.822-0.885)
AID-14	929	187 (20.1%)	0.752 (0.714-0.790)
AID-7	929	187 (20.1%)	0.751 (0.713-0.789)
CURB65	770	165 (21.4%)	0.748 (0.709-0.786)
E-CURB65	146	33 (22.6%)	0.768 (0.682-0.853)
NEWS-FAST	578	112 (19.4%)	0.739 (0.692-0.786)
NEWS2	425	90 (21.2%)	0.746 (0.687-0.804)
NOVARA	865	176 (20.3%)	0.656 (0.613-0.699)
qSOFA	850	172 (20.2%)	0.653 (0.609-0.697)
REMS	780	145 (18.6%)	0.753 (0.712-0.793)
SOFA	288	59 (20.5%)	0.778 (0.712-0.843)
Xie	475	93 (19.6%)	0.816 (0.768-0.863)
Yan	431	81 (18.8%)	0.650 (0.603-0.697)
Zhang	279	67 (24%)	0.810 (0.751-0.869)

Table S1. Demographic and clinical characteristics for patients admitted to hospital with COVID-19 and were transferred to other hospitals (n=77)

Characteristic	Frequency (%) or median (IQR)	Non missing cases (%)
Age (years)	55.0 (51.0, 70.0)	77 (100%)
Sex at birth		77 (100%)
Male	48 (62.3%)	
Comorbities		
Hypertension	41 (53.2%)	77 (100%)
Coronary artery disease	4 (5.2%)	77 (100%)
Heart failure	5 (6.5%)	77 (100%)
Atrial fibrillation or flutter	2 (2.6%)	77 (100%)
Stroke	3 (3.9%)	77 (100%)
COPD	4 (5.2%)	77 (100%)
Diabetes mellitus	22 (28.6%)	77 (100%)
Obesity (BMI>30kg/m^2)	8 (10.4%)	77 (100%)
Cirrhosis	2 (2.6%)	77 (100%)
Cancer	5 (6.5%)	77 (100%)
Number of comorbidities		77 (100%)
0	23 (29.9%)	
1	24 (31.2%)	
2	20 (26.0%)	
3	8 (10.4%)	
4	2 (2.6%)	
Clinical assessment at admission		
SF ratio	433.3 (350.0, 447.6)	75 (97.4%)
Respiratory rate (irpm)	22.0 (18.0, 24.0)	61 (79.2%)
Heart rate (bpm)	89.0 (78.2, 99.8)	70 (90.9%)
Glasgow coma score	15.0 (15.0, 15.0)	75 (97.4%)
Systolic blood pressure (mmHg)		70 (90.9%)
< 90	2 (2.9%)	
≥ 90	68 (97.1%)	

Diastolic blood pressure		70 (90.9%)
≤ 60	12 (17.1%)	
> 60	58 (82.9%)	
Inotrope need at admission	0 (0%)	
Laboratory		
Hemoglobin (g/L)	13.6 (12.2, 14.9)	71 (92.2%)
Platelet count (10^9/L)	196.0 (144.0, 250.0)	71 (92.2%)
Neutrophils-to-lymphocytes ratio	5.7 (4.0, 8.4)	62 (80.6%)
Lactate (mmol/L)	1.3 (1.1, 1.9)	45 (58.4%)
C-reactive protein (mg/L)	87.5 (61.2, 134.5)	62 (80.6%)
BUN (mg/dL)	41.0 (19.1, 28.5)	69 (89.6%)
Creatinine (mg/dL)	1.1 (0.8, 1.4)	73 (94.8%)
Sodium (mmol/L)	138.0 (135.0, 141.0)	65 (84.4%)
Bicarbonate (mEq/L)	21.9 (20.0, 23.2)	59 (76.6%)
pH	7.4 (7.4, 7.5)	60 (77.9%)
pO2 (mmHg)	78.0 (62.1, 99.7)	59 (76.6%)
pCO2 (mmHg)	32.0 (27.9, 35.5)	59 (76.6%)

BMI: body mass index; BUN: blood urea nitrogen; COPD: chronic obstructive pulmonar disease; SF ratio: SpO_2/FiO_2 ratio.

Table S2. Evaluating potential predictors for the model development

Variables	Scientific evidence	Model development (derivation cohort)
Demographics characteristics		
Sex at birth	Halalau *et. al*[68]; 4C Mortality Score[36]; VICE and DICE[51]; COVID-19 Inpatient Risk Calculator (CIRC)[73]; Kazemi *et.al*[75]; Altschul *et. al*[65]; Galloway *et. al*[69]; DCS, DCSL and DL[38]; 17F[80]; CARMc19_N and CARMc19_NB[81]; COVER-F for death[86]; COVID-19 Mortality Socre[87]; CoCoMoRP[88]; Sarkar and Chakrabarti[90].	Included as candidate predictor
Age (years)	A-DROP[91]; Halalau *et. al*[68]; COVID-19MRS[10]; 4C Mortality Score[36]; COVID-GRAM[41]; VICE and DICE[51]; COVID-19 Inpatient Risk Calculator (CIRC)[73]; Sourij *et. al*[74]; Kazemi *et.al*[75]; Núñez-Gil *et. al*[76]; Allenbach *et. al*[14]; Altschul *et. al*[65]; COVID-AID[44]; FAD-85[13]; COVEB[77]; Galloway *et. al*[69]; Bello-Chavolla *et. al*[78]; ANDC[52]; Xie *et.al*[37]; Yoo *et. al*[79]; DCS, DCSL and DL[38]; 17F and 3F models[80]; CSS score[54]; CARMc19_N and CARMc19_NB[81]; Mei *et. al*[82]; Zhang *et. al*[8]; ACP risk grade[83]; LOW-HARM[84]; COVER-F for death[86]; COVID-19 Mortality Socre[87]; CoCoMoRP[88]; NOCOS Calculator[59]; Chen *et. al*[89]; Sarkar and Chakrabarti[90]; Hu *et. al*[55].	Included as candidate predictor
Ethnicity	17F[80]; Galloway *et.* al[69].	Not recorded within database
Hypertension	Halalau *et. al*[68]; COVID-19MRS[10]; Núñez-Gil *et. al*[76]; Galloway *et. al*[69]; Bello-Chavolla *et. al*[78]; DCS[38]; LOW-HARM[84]; COVER-F for death[86]; COVID-19 Mortality Socre[87].	Combined with other comorbities
Coronary artery disease	Halalau *et. al*[68]; COVID-GRAM[41]; CSS score[54]; COVID-19 Mortality Socre[87]; Chen *et. al*[89].	Combined with other comorbities
Heart failure	Halalau *et. al*[68]; Kim *et. al*[15]; COVID-19	Combined with other comorbities

	Mortality Socre[87].	
Atrial fibrillation or flutter	Kim *et. al*[15]; COVID-19 Mortality Socre[87].	Combined with other comorbities
Stroke	Charlson Comorbidity Index[35]; COVID-GRAM[41].	Combined with other comorbities
COPD	COVID-GRAM[41]; Bello-Chavolla *et. al*[78]; 17F[80]; COVER-F for death[86].	Combined with other comorbities
Diabetes mellitus	VICE and DICE[51].	Combined with other comorbities
Obesity (BMI>30kg/m^2)	Halalau *et. al*[68]; 17F[80]; Núñez-Gil *et. al*[76]; Bello-Chavolla *et. al*[78].	Combined with other comorbities
Cirrhosis	Charlson Comorbidity Index[35], 4C Mortality Score[36].	Combined with other comorbities
Cancer	COVID-19MRS[10]; COVID-GRAM[41]; DCS and DCSL[38]; 17F[80]; COVER-F for death[86].	Combined with other comorbities
Smoking	Salah, Sharma and Mehta[92].	High collinearity with COPD, not included
Number of comorbidities	COVID-19MRS[10]; 4C Mortality Score[36]; COVID-GRAM[41].	Included as candidate predictor
Clinical characteristics		
Respiratory rate (irpm)	COVID-19MRS[10]; 4C Mortality Score[36]; Gavelli *et. al*[67]; Galloway *et.* al[69].	Included as candidate predictor
Heart rate (bpm)	NEWS2[93].	Included as candidate predictor
Systolic blood pressure (mm Hg)	CURB65[29].	Combined with inotrope requirement and included as candidate predictor
Diastolic blood pressure (mm Hg)	17F[80]; CURB65[29].	High collinearity with systolic blood pressure, not included
Inotrope use	SOFA[94].	Combined with systolic and diastolic blood pressure
Glasgow coma score	Yoo *et. al*[79].	Included as candidate predictor
Temperature (ºC)	17F[80]; Mei *et. al*[82].	Too many missing values, not included
SF ratio	Choi, Hong and Kim[95]; Choi *et. al*[95].	Included as candidate predictor predictor
Laboratory	Lim *et. al*[96].	
Mechanical ventilation		Included as candidate predictor
C reactive protein (mg/L)	VICE and DICE[51]; ANDC[52].	Included as candidate predictor
Hemoglobin (g/L)	Lippi and Mattiuzzi[97].	Included as candidate predictor

Neutrophils-to-lymphocytes ratio	COVID-GRAM[41]; ANDC[52]; VICE and DICE[51].	Included as candidate predictor
Platelet count (10^9/L)	SOFA[94]; VICE and DICE[51]; EDRnet[58]; COVID-19 Mortality Socre[87].	Included as candidate predictor
Creatinine (mg/dL)	COVID-19MRS[10]; COVID-AID[44]; Altschul *et. al*[65]; Galloway *et.* al[69]; DCSL and DL[38]; LOW-HARM[84]; SOFA[94].	Included as candidate predictor
Urea (mg/dL)	4C Mortality Score[36]; EDRnet[58]; NOCOS Calculator[59], CURB65[29].	Included as candidate predictor
Lactate (mmol/L)	COVID-GRAM[41]; NLAUD[16]; Xie *et.al*[37].	Included as candidate predictor
Sodium (mmol/L)	PSI[98].	Included as candidate predictor
Bicarbonate (mEq/L)	EDRnet[58].	Included as candidate predictor
pH	Li *et. al*[99].	Included as candidate predictor
pO2 (mmHg)	SOFA[94].	Included as candidate predictor
pCO2 (mmHg)	Li *et. al*[99].	Included as candidate predictor
Ferritin (mcg/L)	FAD-85[13].	Too many missing values, not included
NT-proBNP (pg/mL)	Kim *et. al*[15].	Too many missing values, not included
Creatine kinase (U/L)	Kim *et. al*[15].	Too many missing values, not included
Troponin (ng/mL)	Yoo *et. al*[79].	Too many missing values, not included
Bilirubin (mg/dL)	SOFA[94]; COVID-GRAM[41]; Zhang *et. al*[8]; Chen *et. al*[89].	Too many missing values, not included
Partial thromboplastin time (times the control value in seconds)	Zhou *et. al*[57].	Too many missing values, not included
Lactate dehydrogenase (U/L)	COVID-GRAM[41]; Xie *et.al*[37].	Too many missing values, not included
International normalized ratio	Zhou *et. al*[57].	Too many missing values, not included
Alanine aminotransferase (U/L)	EDRnet[58]; Chen *et. al*[89]; Sourij *et. al*[74]; Mei *et. al*[82].	Too many missing values, not included
Aspartate aminotransferase (U/L)		Too many missing values, not included
D-dimer	FAD-85[13]; NLAUD[16]; ANDC[52]; CSS score[54]; Hu *et. al*[55].	Different assays may compromise assessment, not incluided

Table S3. Variable selection based on Generalized Additive Model

Variable	Deviance explained (%)	R-sq.(adj)	UBRE	D1-statistics (p-value)	D2-statistics (p-value)
All variables included	0.354	0.361	-0.324		
Sex at birth	0.354	0.361	-0.325	0.773	0.785
Age (years)	0.314	0.320	-0.284	0.000**	0.000**
Number of comorbities	0.353	0.361	-0.323	0.011**	0.011**
Respiratory rate (irpm)	0.351	0.358	-0.321	0.246	0.131
Heart rate (bpm)	0.350	0.357	-0.320	0.047**	0.122
Systolic blood pressure (mm Hg)	0.353	0.361	-0.324	0.217	0.244
Glasgow coma score	0.353	0.360	-0.324	0.995	1.000
SF ratio	0.333	0.339	-0.303	0.000**	0.000**
C-reactive protein (mg/L)	0.347	0.355	-0.318	0.006**	0.019**
Hemoglobin (g/L)	0.348	0.358	-0.321	0.069	0.087
NL ratio	0.351	0.359	-0.323	0.966	0.840
Platelet count (10^9/L)	0.335	0.344	-0.308	0.000**	0.000**
Creatinine (mg/dL)	0.354	0.361	-0.325	1.000	1.000
BUN (mg/dL)	0.347	0.355	-0.320	0.000**	0.001**
Lactate (mmol/L)	0.348	0.356	-0.320	0.144	0.459
Sodium (mmol/L)	0.352	0.359	-0.324	0.689	0.957
Bicarbonate (mEq/L)	0.353	0.360	-0.325	0.999	1.000
pH	0.352	0.360	-0.323	0.805	0.925
pO2 (mmHg)	0.349	0.358	-0.321	0.554	0.678
pCO2 (mmHg)	0.353	0.361	-0.324	0.996	1.000

BUN: blood urea nitrogen; UBRE: Unbiased risk estimator; D1: multivariate Wald test; D2: pools test statistics from the repeated analyses; NL: neutrophils-to-lymphocytes count ratio; SF: SpO_2/FiO_2 ratio

** Variable included in final model (p-value < 0.05)

Table S4. L1 penalized shrunk coefficients and scaled coefficients from LASSO logistic regression

Variable	Coefficients	Scaled coefficients (× 3)
Age (years)		
< 60	-	0
60 - 69	0.413	1
70 - 79	0.935	3
≥ 80	1.666	5
Number of comorbidities		
≤ 1	-	0
> 1	0.353	1
SF ratio		
> 315.0	-	0
235.1 – 315.0	0.431	1
150.1 – 235.0	1.001	3
≤ 150.0	1.880	6
C reactive protein (mg/L)		
< 100	-	0
≥ 100	0.476	1
Blood urea nitrogen (mg/dL)		
< 42	-	0
≥ 42	0.905	3
Platelet count (10^9/L)		
> 150	-	0
100 -150	0.288	1
< 100	0.667	2
Heart rate (bpm)		
≤ 90	-	0
91 – 130	0.185	1
≥ 131	0.503	2
Intercept	-2.965	-9

LASSO: least absolute shrinkage and selection operator logistic regression, SF ratio: SpO_2/FiO_2 ratio

Table S5. Sensitivity analysis - Discrimination and model overall performance within complete cases

Model	Derivation Cohort		Brazilian Validation Cohort	
	AUROC (95%CI)	Brier Score	AUROC (95%CI)	Brier Score
GAM	0.871 (0.866; 0.875)	0.108	0.880 (0.878; 0.887)	0.094
LASSO	0.824 (0.792; 0.856)	0.115	0.858 (0.793; 0.922)	0.092
ABC_2-SPH	0.841 (0.824; 0.858)	0.114	0.852 (0.820; 0.884)	0.107

GAM: generalized additive models; LASSO: least absolute shrinkage and selection operator logistic regression

Table S6. TRIPOD checklist for transparent reporting on a multivariable prognostic model.

Section/topic	Item	Checklist item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	8
Introduction			
Background and objective	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models	10
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both	10-11
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable	11
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up	11
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres	11
	5b	Describe eligibility criteria for participants	11, 12
	5c	Give details of treatments received, if relevant	NA
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed	12
	6b	Report any actions to blind assessment of the outcome to be predicted	NA
Predictors	7a	Clearly define all predictors used in developing the multivariable prediction model, including how and when they were measured	12
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors	NA
Sample size	8	Explain how the study size was arrived at	NA

Section/topic	Item	Checklist item	Page
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method	13
Statistical analysis methods	10a	Describe how predictors were handled in the analyses	13, 14
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation	13
	10c	For validation, describe how the predictions were calculated	14
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models	14
	10e	Describe any model updating (e.g., recalibration) arising from the validation, if done	NA
Risk groups	11	Provide details on how risk groups were created, if done	14
Development vs. validation	12	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors	14
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful	15, Figure 1
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome	15, Table 1
	13c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome)	Table 1
Model development	14a	Specify the number of participants and outcome events in each analysis	Table 1
	14b	If done, report the unadjusted association between each candidate predictor and outcome	NA
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point)	Table S4
	15b	Explain how to use the prediction model	Page 16, Table 2
Model performance	16	Report performance measures (with CIs) for the prediction model	Table 4, Table S5
Model updating	17	If done, report the results from any model updating (i.e., model specification, model performance)	NA
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data)	22, 23

Interpretation	19a	For validation, discuss the results with reference to performance in the development data, and any other validation data	18-25
	19b	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence	18, 19
Implications	20	Discuss the potential clinical use of the model and implications for future research	23, 24
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets	16, 28
Funding	22	Give the source of funding and the role of the funders for the present study	27, 28

Table S7. Risk of bias assessment using PROBAST checklist

Domain and Item	Checklist item	Development	Brazilian validation
Participants			
1.1	Were appropriate data sources used, e.g., cohort, RCT, or nested case–control study data?	Yes (a cohort design has been used)	Yes (a cohort design has been used)
1.2	Were all inclusions and exclusions of participants appropriate?	Yes (participants correspond to unselected participants of interest)	Yes (participants correspond to unselected participants of interest)
Risk of bias introduced by participants or data sources: low risk of bias.			
Predictors			
2.1	Were predictors defined and assessed in a similar way for all participants?	Yes (definitions of predictors and their assessment were similar for all participants)	Yes (definitions of predictors and their assessment were similar for all participants)
2.2	Were predictor assessments made without knowledge of outcome data?	Yes (outcome information was stated as not used during predictor assessment)	Yes (outcome information was stated as not used during predictor assessment)
2.3	Are all predictors available at the time the model is intended to be used?	Yes (all included predictors were available at the time the model was intended to be used for prediction)	Yes (all included predictors were available at the time the model was intended to be used for prediction)
Risk of bias introduced by predictors or their assessment: low risk of bias.			
Outcome			
3.1	Was the outcome determined appropriately?	Yes (objective outcome was used: mortality)	Yes (objective outcome was used: mortality)
3.2	Was a prespecified or standard outcome definition used?	Yes (objective outcome was used: mortality)	Yes (objective outcome was used: mortality)
3.3	Were predictors excluded from the outcome definition?	Yes (none of the predictors are included in the outcome definition)	Yes (none of the predictors are included in the outcome definition)
3.4	Was the outcome defined and determined in a similar way for all participants?	Yes (outcomes were defined and determined in a similar way for all participants)	Yes (outcomes were defined and determined in a similar way for all participants)
3.5	Was the outcome determined without knowledge of predictor information?	Yes (predictor information was not known when determining the	Yes (predictor information was not known when determining the

		outcome status)	outcome status)
3.6	Was the time interval between predictor assessment and outcome determination appropriate?	Yes (time interval between predictor assessment and outcome determination was appropriate)	Yes (time interval between predictor assessment and outcome determination was appropriate)
Risk of bias introduced by predictors or their assessment: low risk of bias.			
Analysis			
4.1	Were there a reasonable number of participants with the outcome?	Yes (high number of events per variable).	Yes (number of participants with the outcome is ≥100)
4.2	Were continuous and categorical predictors handled appropriately?	Yes (continuous predictors are examined for nonlinearity using thin-plate splines and then categorical predictor groups were defined using widely accepted cut points, current evidence and/or categories defined in stablished rapid scoring systems).	Yes (predictors were used as in the development model).
4.3	Were all enrolled participants included in the analysis?	Yes (all participants enrolled in the study were included in the data analysis).	Yes (all participants enrolled in the study are included in the data analysis).
4.4	Were participants with missing data handled appropriately?	Yes (missing values were handled using multiple imputation methods)	Yes (missing values are handled using multiple imputation methods)
4.5	Was selection of predictors based on univariable analysis avoided?	Yes (the predictors were not selected on the basis of univariable analysis prior to multivariable modeling)	NA
4.6	Were complexities in the data (e.g., censoring, competing risks, sampling of control participants) accounted for appropriately?	Yes (a full cohort approach was used - median follow-up time was 7 days)	Yes (a full cohort approach was used - median follow-up time was 7 days)
4.7	Were relevant model performance measures evaluated appropriately?	Yes (both calibration and discrimination were evaluated appropriately)	Yes (both calibration and discrimination were evaluated appropriately)
4.8	Were model overfitting and optimism in model performance accounted for?	Yes (10-fold cross-validation have been used).	NA
4.9	Do predictors and their assigned weights in	Yes (the predictors and regression	NA

	the final model correspond to the results from the reported multivariable analysis?	coefficients in the final model correspond to reported results from multivariable analysis)
Risk of bias introduced by the analysis: low risk of bias.		

Table S8. Justifications for inclusion or exclusion

Study	Included?
Halalau[68]	No. Ccongenital heart disease not available.
Fumagalli[10]	No. Depression and dementia were not categorical variables in the present study.
Knight[36]	No. Dementia was collected as a free-text field, and could not be cathegorized up to the data this study was submited.
Liang[41]	No. Composite outcome.
Nicholson[51]	No. Mean corpuscular volume not available.
Garibaldi[73]	No. Nursing home resident and BMI not available.
Sourij[74]	No. Arterial onclusive disease not available.
Gavelli[67]	No. SpO2 and respiratory rate after 15-minute trial with oygen not available.
Kazemi[75]	No. Comorbidities were not well defined, percentage of involvement included in CT score is subjective and peripheral involvement is not well defined.
Núñez-Gil[76]	No. Variables not cleary defined (renal failure and elevated C-reactive protein).
Allenbach[14]	No. Composite outcome.
Kim[15]	No. CK-MB not available.
Altschul[65]	No. IL-6 not available, intercept not provided for calculation.
Hajifathalian[44]	Yes
Wang J[13]	No. D-dimer assay not described by the authors.
Zhou[16]	No. D-dimer assay not described by the authors.
Goméz[77]	No. The authors did not provide all ifnormation necessary to calculate the score.
Galloway[69]	No. Ethnicity not available.
Bello-Chavolla[78]	No. As the score was developed considering outpatients and inpatients, the comparison would not be appropriate.
Weng[52]	No. D-dimer assay not described by the authors.
Ko[52]	No. Not all predictors are availabe, such as RDW.
Xie[37]	Yes
Yoo[79]	No. Troponin assay not described by the authors.
Zhang[38]	No. Very limited study, most included variables had OR with 95% CI including 1.0.
Yadaw[80]	No. Ethnicity not available.
Shang[54]	No. D-dimer assay not described by the authors.

Faisal[81]	No. Information to allow calculation was not provided.
Mei[82]	No. Total protein not available.
Zhang[8]	Yes
Lu[83]	No. Score development included patients with confirmed and suspected COVD-19, a comparison would not be appropriate.
Soto-Mota[84]	No. Not clear the moment the score is meant to be used.
Yan[85]	Yes
Williams[86]	No. Hyperlipidemia not available as a categorical variable.
Gue[87]	Yes
Das[88]	No. Variables such as province not applicable for other populations.
Levy[59]	No. Authors did not show how to calculate the score.
Chen[89]	No. Authors did not show how to calculate the score.
Sarkar[90]	No. Some variables applicable only to the Chinese population, in the beggiing og the pandemic.
Hu[55]	No. D-dimer assay not described by the authors.

Figure legends

Figure 1. Flowchart of COVID-19 patients included in the study

Figure 2. City of residence of patients within (a) development and (b) validation cohorts

Figure 3. ABC_2-SPH Score in derivation and validation cohorts

Figure 4. Discrimination of ABC_2-SPH Score in external validation cohorts

Figure 5. ROC curves (a) and decision curve for best performing scores

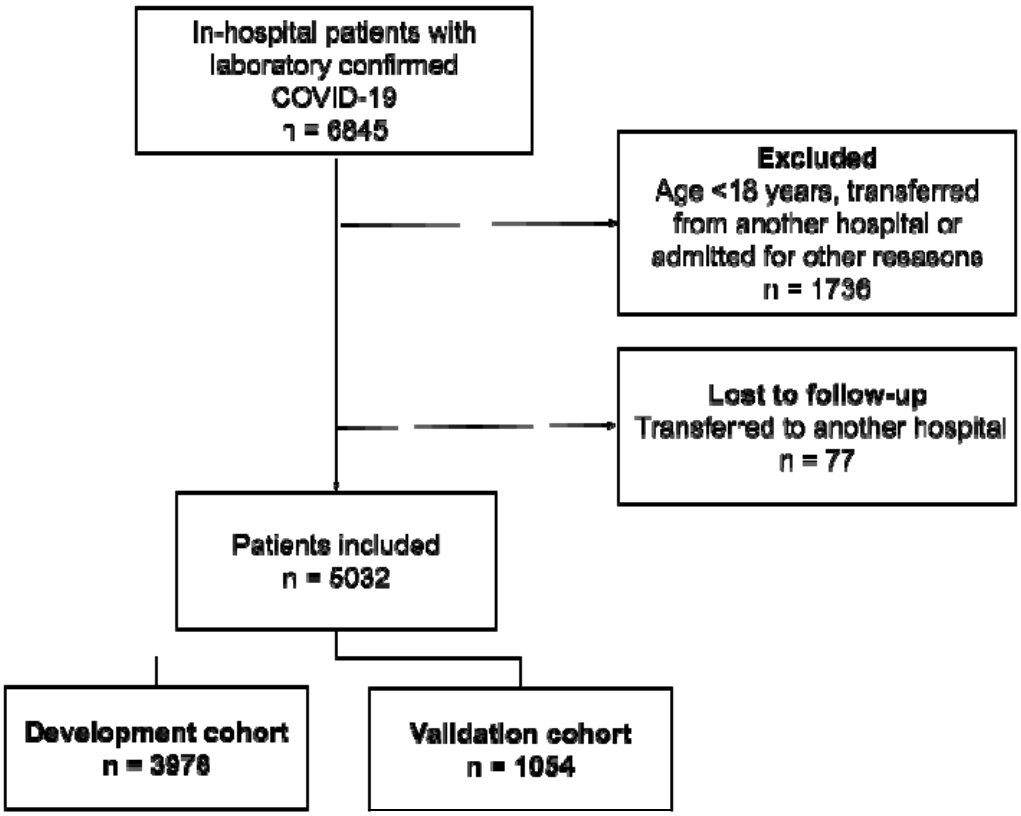

Figure 1. Flowchart of COVID-19 patients included in the study

(a) (b)

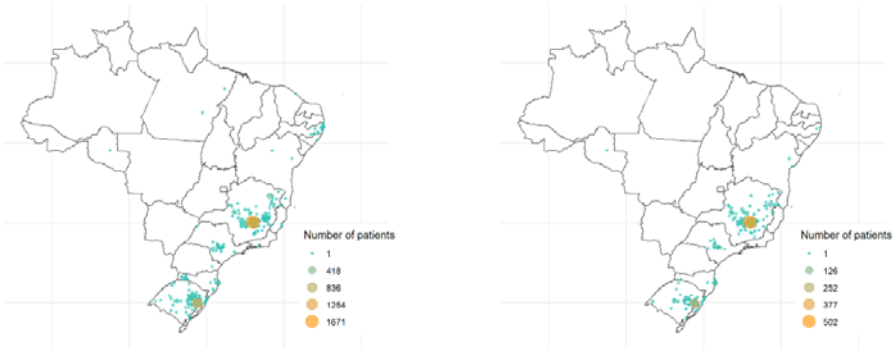

Figure 2. City of residence of patients within (a) development and (b) validation cohorts

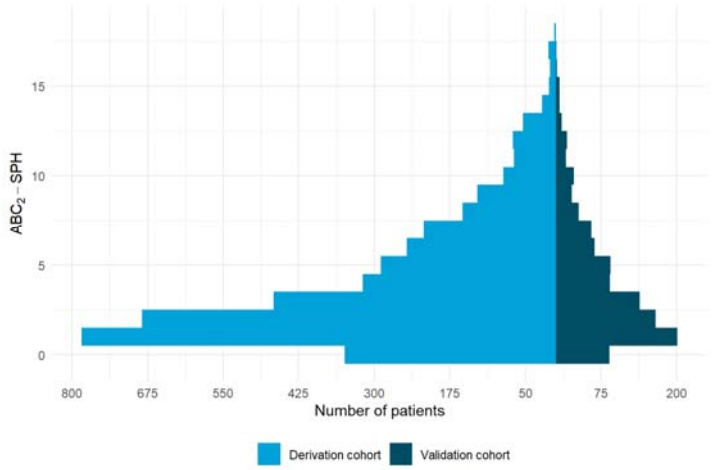

Figure 3. ABC_2-SPH Score in derivation and validation cohorts

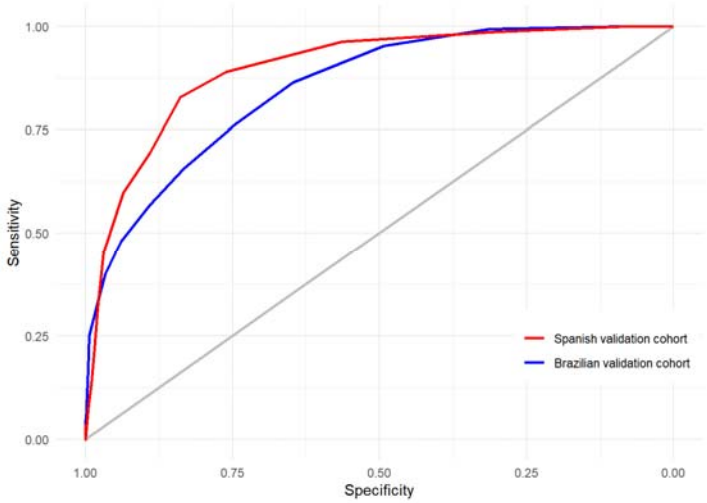

Figure 4. Discrimination of ABC_2-SPH Score in external validation cohorts

(a) (b)

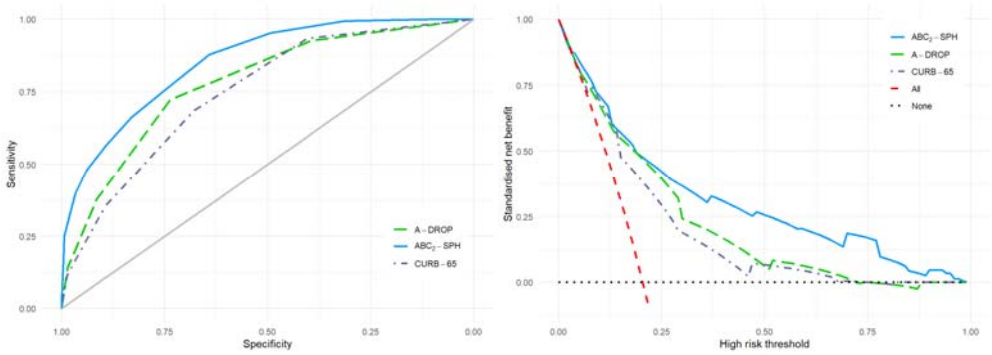

Figure 5. ROC curves (a) and decision curve for best performing scores

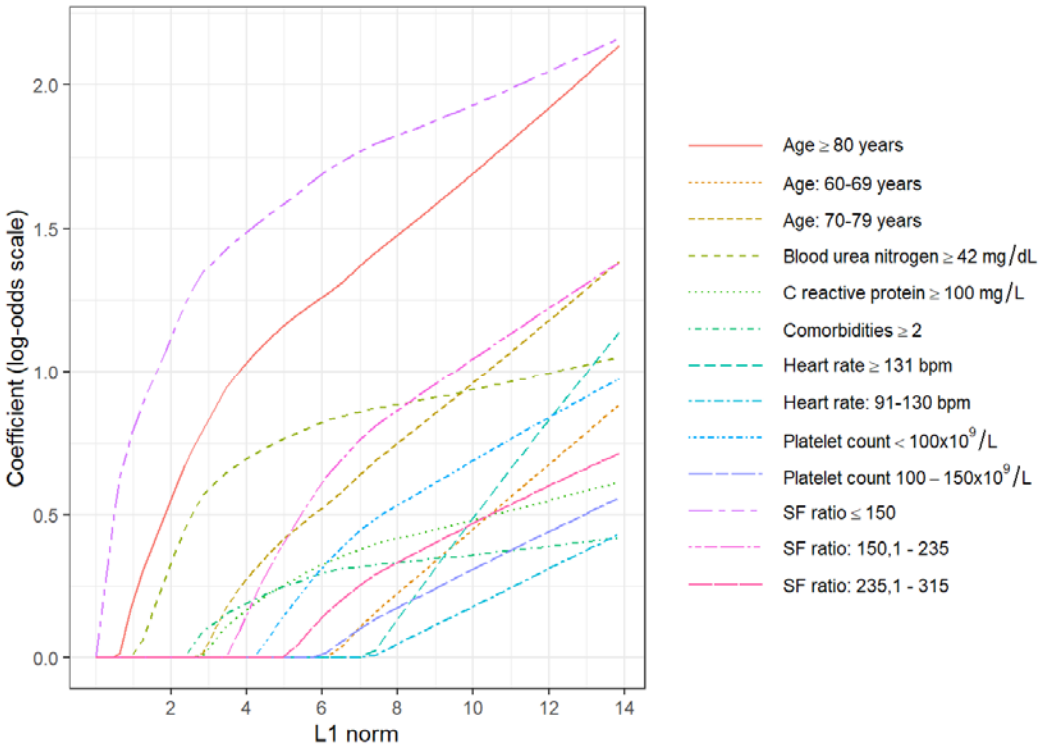

Figure S1. Least absolute shrinkage and selection operator logistic regression (LASSO) trace plot

(a) (b)

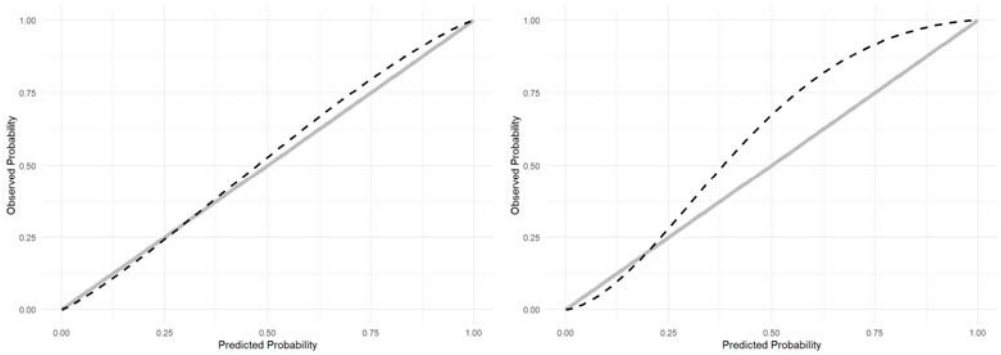

Figure S2. Calibration plot of ABC_2-SPH Score in (a) Brazilian and (b) Spanish external validation cohorts