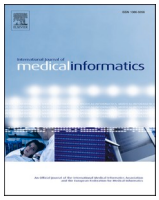
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

International Journal of Medical Informatics 173 (2023) 105039

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf



Clinical prediction rules for adverse evolution in patients with COVID-19 by the Omicron variant

Janire Portuondo-Jiménez [a,b,c], Irantzu Barrio [d,e,*], Pedro P. España [b,f], Julia García [g], Ane Villanueva [c,h,i,j], María Gascón [c,h,i,j], Lander Rodríguez [e], Nere Larrea [c,h,i,j], Susana García-Gutierrez [c,h,i,j], José M. Quintana [c,h,i,j], the COVID-Health Basque Country Research Group

[a] Osakidetza Basque Health Service, Sub-Directorate for Primary Care Coordination, Vitoria-Gasteiz, Spain
[b] Biocruces Bizkaia Health Research Institute, Barakaldo, Spain
[c] Network for Research on Chronicity, Primary Care, and Health Promotion (RICAPPS), Spain
[d] University of the Basque Country UPV/EHU, Department of Mathematics, Leioa, Spain
[e] Basque Center for Applied Mathematics, BCAM, Spain
[f] Osakidetza Basque Health Service, Galdakao-Usansolo University Hospital, Respiratory Unit, Galdakao, Spain
[g] Basque Government Department of Health, Office of Healthcare Planning, Organization and Evaluation, Basque Country, Spain
[h] Osakidetza Basque Health Service, Galdakao-Usansolo University Hospital, Research Unit, Galdakao, Spain
[i] Health Service Research Network on Chronic Diseases (REDISSEC), Bilbao, Spain
[j] Kronikgune Institute for Health Services Research, Barakaldo, Spain

ARTICLE INFO

Keywords:
SARS-CoV-2
COVID-19
Clinical decision rules
Outcome assessment
Health care

ABSTRACT

Objective: We identify factors related to SARS-CoV-2 infection linked to hospitalization, ICU admission, and mortality and develop clinical prediction rules.
Methods: Retrospective cohort study of 380,081 patients with SARS-CoV-2 infection from March 1, 2020 to January 9, 2022, including a subsample of 46,402 patients who attended Emergency Departments (EDs) having data on vital signs. For derivation and external validation of the prediction rule, two different periods were considered: before and after emergence of the Omicron variant, respectively. Data collected included socio-demographic data, COVID-19 vaccination status, baseline comorbidities and treatments, other background data and vital signs at triage at EDs. The predictive models for the EDs and the whole samples were developed using multivariate logistic regression models using Lasso penalization.
Results: In the multivariable models, common predictive factors of death among EDs patients were greater age; being male; having no vaccination, dementia; heart failure; liver and kidney disease; hemiplegia or paraplegia; coagulopathy; interstitial pulmonary disease; malignant tumors; use chronic systemic use of steroids, higher temperature, low O2 saturation and altered blood pressure-heart rate. The predictors of an adverse evolution were the same, with the exception of liver disease and the inclusion of cystic fibrosis. Similar predictors were found to be related to hospital admission, including liver disease, arterial hypertension, and basal prescription of immunosuppressants. Similarly, models for the whole sample, without vital signs, are presented.
Conclusions: We propose risk scales, based on basic information, easily-calculable, high-predictive that also function with the current Omicron variant and may help manage such patients in primary, emergency, and hospital care.

1. Introduction

The SARS-CoV-2 infection, which began in December 2019 [1] has now become a global pandemic of unpredictable consequences constituting a threat to public health [2], as well as causing thousands of deaths daily throughout the world [3]. The first wave hit health systems hard, generating great uncertainty as to the nature of the new disease and its prognosis [4]. In order to combat the disease, a variety of

* Corresponding author at: Departamento de Matemáticas, Universidad del País Vasco UPV/EHU, Leioa, Spain.
E-mail address: irantzu.barrio@ehu.eus (I. Barrio).

https://doi.org/10.1016/j.ijmedinf.2023.105039
Received 17 November 2022; Received in revised form 3 February 2023; Accepted 1 March 2023
Available online 7 March 2023

attempts have been made since the beginning of the pandemic to understand the pathophysiology of the infection [5]. Initially, most patients infected with SARS-CoV-2 were considered asymptomatic or had mild symptoms, and were therefore dealt with from Primary Care (PC) centers. However, in some cases, infection was associated with the "cytokine storm" syndrome [6,7], which, together with respiratory failure, was related to an increase in hospital and Intensive Care Unit (ICU) admissions and mortality [8].

Many aspects of COVID-19 remain unknown, given the changing nature of the infection and the similarities and differences between the characteristics of the different waves [9] and this has necessitated frequent re-appraisal of care planning [10]. Consequently, in order to provide crucial perspectives for care services and develop appropriate health policies, numerous predictive models have been developed [2], providing useful predictions for risk of clinical deterioration or ICU admission, with good discrimination [11,12], which are regularly being updated [13,14], to provide more information to clinicians about patients' health status and better risk stratification indications [15].

Currently, the prospect is that COVID-19 will not disappear in the short or medium term [14], despite the vaccination process implemented during 2021–2022. Moreover, constant study is required of the characteristics of the disease and the factors related to an adverse evolution, in order to enable rapid modification of treatments and reorganization of the health system if necessary [16].

In this paper, we seek to identify factors related to hospitalization, adverse evolution—defined as admission to an ICU or death—and mortality related to the infection with basic information, or adding vital signs, within COVID-19 patients from the general population, and evaluate their performance in the latest variant of SARS-CoV-2, Omicron.

2. Methods

All patients included in this retrospective cohort study were residents in our region who had a SARS-CoV-2 infection, laboratory-confirmed by a positive result on the reverse transcriptase-polymerase chain reaction assay for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) or a positive antigen test between March 1, 2020 and January 9, 2022. From March 1, 2020 to July 31, 2020, positive IgM or IgG antibody tests performed due to patients having symptoms suggestive of the disease or having had contact with a positive case were also included in the general population sample. The first positive from each patient was collected. Only patients aged over 18 years were included. From all of them, we select those who attended an ED by their COVID-19 infection and have information on vital signs being the Emergency Department (ED) sample. The study protocol was approved by the Ethics Committee of our area (reference PI2020123). All patient data was kept confidential.

Data collected include sociodemographic, baseline comorbidities (including those of the Charlson Comobidity Index [17] and based on ICD codes [18]), baseline treatments (based on the Anatomical, Therapeutic, Chemical [ATC] classification system) [19]; dates of hospital admission and discharge and whether patients were admitted to an intensive care unit (ICU); and vital status. From those attended at any ED, we recorded vital signs (body temperature, blood pressure, heart rate and O2 saturation). We defined heart rate-diastolic/systolic pressure as the combination of heart rate and diastolic-systolic pressure as presented in Online Table 1 [20–22].

The outcomes used in the study were as follows: 1.-Hospital admission due to COVID-19, defined if admission occurred within 15 days of the patient's testing positive, when the positive test preceded hospitalization, and up to 21 days after admission when the patient tested positive during hospitalization; 2.-Death during the three months following diagnosis or during a hospital admission as defined previously or three months from discharge; 3.-Adverse evolution, including death or ICU admission during a hospital admission related to a SARS-CoV-2 infection diagnosis as defined above. All patients were monitored to April 9, 2022. The period from March 1st, 2020 to December 13, 2021 was considered as a sample for model development (hereinafter referred to as the Derivation Data Set), while the period from December 14 to January 9, 2022, corresponding to the Omicron variant wave was used to validate the consistency of the results obtained (hereinafter referred to as the Omicron - Validation Data Set).

2.1. Statistical analysis

Models were developed for both the whole sample and the ED sample. The Derivation Data Set was randomly divided in equal halves for both samples. One half (50 %) was used for variable selection and estimation of parameters of the prediction model (train) and the other half (50 %) was used for internal validation (test) [19]. The Omicron Data Set was used for external validation, given that although it is also a database of individuals from our region, it includes different people who tested positive for a different variant to the previous ones. Patient characteristics were compared between the subsamples (train vs. test and train vs. Omicron) using Chi-square or Fisher's exact tests for categorical variables.

Given the large sample size (n train = 120,534 and n test = 120,533 in the general population sample and n train = 19,672 and n test = 19,672 in the ED sample), variable selection was performed by means of a multivariate Lasso logistic regression model (1.-Hospital admission; 2.-Death; and 3.-Adverse evolution) which employs penalized likelihood for parameter estimates and variable selection in the train subsample [19,20]. The final models were adjusted by means of a multilevel logistic regression considering that patients were nested in the IHOs. Odds ratios (ORs) and 99 % confidence intervals (CIs) were estimated. In addition, final models' variables' importance was measured by means of a Random Forest algorithm using the Boruta package, which gives a numerical estimate of the variable importance [26]. The discrimination ability of the model was measured by the area under the ROC curve (AUC) [21], and for calibration purposes calibration plots have been drawn and the Brier score has been calculated. In addition, given that we had an unbalanced sample (prevalence for COVID patients attending ER in train samples of 10 %, 15 % and 57 % for mortality, poor evolution and hospitalization, respectively) we calculated the Precision-Recall Curve and calculated the area under it (AUPRC) [27,28]. A significance level of 0.01 was considered.

To develop the predictive risk scores for each of the outcomes (1.-Hospital admission; 2.-Death; and 3.-Adverse evolution), we first assigned a weight to each category of the predictor variable based on the estimated β parameters of the multilevel logistic regression model derived in the train subsample. Categories of predictive variables with $p > 0.01$ were assigned a weight of 0. We then added up the risk weights of all the patient's predictor variables, with higher scores indicating a greater likelihood of event. The predictive accuracy of the risk score was assessed using the AUC in train, test and Omicron samples. We categorized the risk score into five different levels of risk. The optimal thresholds in the continuous risk scores were estimated considering those cut-off points for which the maximal AUC for the categorized score was obtained in the train sample, following the methodology proposed by Barrio et al. 2017 [29] and using the BackAddFor algorithm proposed by Barrio et al. 2021 [30]. The performance of the risk classification was evaluated by means of the AUC, AUPRC, and by studying the probability of event occurrence in each of the risk categories. In addition, the true positive rate (TPR), true negative rate (TNR), F1-score, and the net benefit (NB), which considers the relative benefits and harms, were computed for each of the risk cut-off points [23–25,30,31]. The model, score and categorized score were all validated in the Omicron sample by means of the AUC and AUPRC. The graphical representation of the model's development pipeline is shown in Online Fig. 1. All statistical analyses were performed using R© version 4.1.2.

3. Results

During the study period, 380,081 people tested positive and 46,402 were also seen at EDs. Flowcharts describing patient evolution in each sample are shown in Online Fig. 2 and Fig. 1 respectively. Descriptive statistics for the main variables of the study are reported in Table 1 (ED sample) and Online Table 2 (whole sample), for train, test and Omicron samples, respectively. As can be seen, main outcomes prevalence were smaller in the Omicron sample ($p < 0.001$).

The variables identified in the multivariable model related to death on patients who attended an ED and has vital signs information were greater age; being male; no vaccination; baseline diseases such as heart failure, liver and kidney disease, dementia, hemiplegia or paraplegia, specific lung diseases such as interstitial pulmonary disease; coagulopathy and history of malignant tumors and from basal treatments, use of chronic systemic steroids. Among vital signs having a body temperature > 37, O2 saturation ≤ 94 and blood pressure-heart rate combination altered were also related to death. The AUCs for the categorized score were 0.90, 0.90 and 0.91, and the Brier scores were 0.06, 0.06 and 0.04, in train, test and Omicron samples, respectively (Table 2). For the whole sample, in addition to the previous, peripheral vascular disease, ischemic heart and cerebrovascular disease, diabetes, cystic fibrosis; and basal treatments, use of diuretics were also related to death. The AUCs for the categorized score were 0.95 and the Brier scores were 0.02, 0.02 and 0.004, in train, test and Omicron samples, respectively (Online Table 3).

The variables related to adverse evolution identified in the multivariable model of the ED sample were older age; being male; no vaccination; baseline diseases such as heart failure, kidney disease, dementia, hemiplegia or paraplegia, cystic fibrosis, interstitial pulmonary disease; coagulopathy and history of malignant tumors and from basal treatments, use of chronic systemic steroids. Among vital signs having a body temperature > 37, O2 saturation ≤ 94 and blood pressure-heart rate combination altered were also related to adverse evolution. The AUCs for the categorized score were 0.84, 0.83 and 0.88, and the Brier scores were 0.10, 0.10 and 0.05, in train, test and Omicron samples, respectively (Table 3). For the whole sample, in addition to the previous, peripheral vascular disease, ischemic heart and cerebrovascular disease, and basal treatments, use of diuretics were also related to death. The AUCs for the categorized score were 0.89, 0.89 and 0.91, and the Brier scores were 0.03, 0.03 and 0.005, in the train, test and Omicron samples, respectively (Online Table 4).

Finally, the variables related to hospital admission identified in the multivariable model of the ED sample were older age; being male; no vaccination; baseline diseases such as heart failure, liver disease, arterial hypertension, and history of malignant tumors. Among the basal treatments, the use of chronic systemic steroids and among vital signs having a body temperature > 37, O2 saturation ≤ 94 and blood pressure-heart rate combination altered were also related to hospital admission. The AUCs for the categorized score were 0.82, 0.82 and 0.83, and the Brier scores were 0.17, 0.17 and 0.15, in train, test and Omicron samples, respectively (Table 4). For the whole sample, in addition to the previous, baseline diseases such as ischemic heart and cerebrovascular disease, kidney disease, dyslipidemia, dementia, diabetes, inflammatory bowel disease, HIV, interstitial pulmonary disease; and history of malignant tumors. Among the basal treatments, the use of antidiabetics, bronchodilators, immunosuppressants, and diuretics were also related to hospital admission. The AUCs for the categorized score were 0.81, 0.82 and 0.84, and the Brier scores were 0.07, 0.07 and 0.02, in the train, test and Omicron samples, respectively (Online Table 5).

For all different models and cut points, we estimated the sensitivity, specificity, Net Benefit, F1-Score and Balanced Accuracy percentages (Table 5 and Online Table 6) while the risk/probability of event was represented for each outcome and risk category (Fig. 2 and Online Fig. 3). In addition, ROC and Precision-Recall Curves were plotted (Fig. 3 and Online Fig. 4) Finally, calibration plots were drawn for the derived models in both samples (Fig. 4 and Online Fig. 5).

4. Discussion

This study, which included a very large cohort of COVID-19-positive patients (380,081), recruited during almost two years of the pandemic, identified predictors of three different outcomes. It allows us to see a pattern of variables common to all three outcomes, including age, sex, cardio-cerebrovascular diseases, diabetes, kidney and liver disease, tumors, and some more serious specific lung diseases such as interstitial lung disease and cystic fibrosis. Additionally, we found a single treatment common to all three outcomes, namely the chronic systemic use of steroids and the protective effect of the COVID-19 vaccination. For patients with basic vital signs information, we pointed out the importance of the alteration of some vital signs.

Most of the above factors have been identified and summarized in

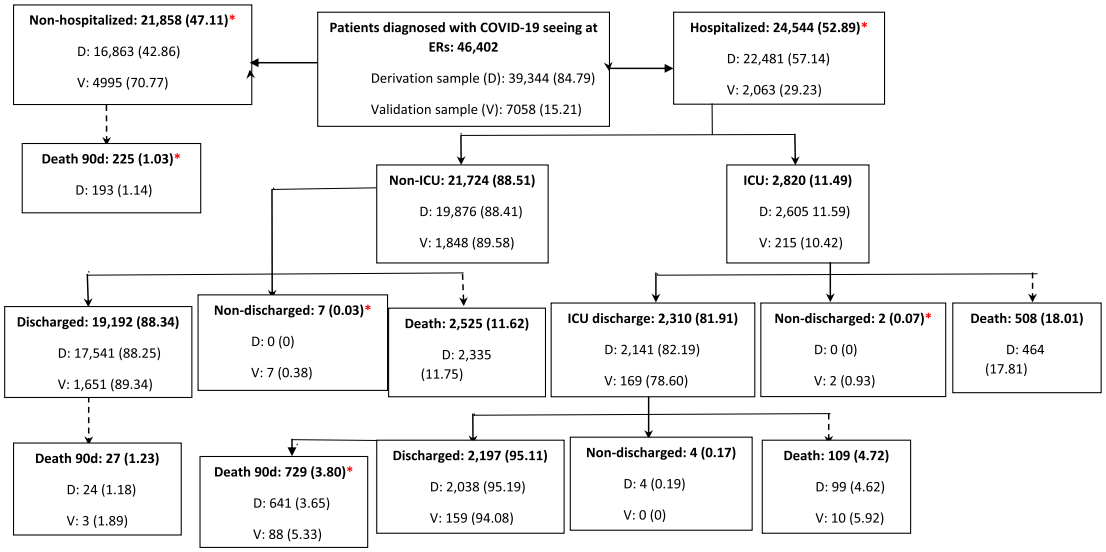


Fig. 1. Flow-chart of the evolution of 46,402 adult patients diagnosed with COVID-19 and seeing at Emergency Rooms (ERs) between 01/03/2020–13/12/2021 (Derivation data set (train + test samples), n = 39,344) and between 14/12/2021–09/01/2022 (Omicron validation sample, n = 7,058).*Statistically significant differences between derivation and validation samples.

Table 1
Descriptive statistics of COVID 19 patients who attended ERs (N = 46,402) in Train, Test and Omicron subsamples, together with chi square test p-values for independence among subsamples.

Variables	Train N (%)	Test N (%)	Omicron N (%)	p_values train-test	P_values train-omicron
TOTAL	19,672	19,672	7058		
Sociodemographic variables					
Gender				0.4024	<0.0001
Female	9,483 (48.21)	9,567 (48.63)	3,812 (54.01)		
Male	10,189 (51.79)	10,105 (51.37)	3,246 (45.99)		
Age (years) categorized				0.2573	<0.0001
18–39	3,810 (19.37)	3,947 (20.06)	1,953 (27.67)		
40–49	3,094 (15.73)	3,133 (15.93)	1,549 (21.95)		
50–59	3,419 (17.38)	3,372 (17.14)	1,295 (18.35)		
60–69	3,112 (15.82)	3,099 (15.75)	829 (11.75)		
70–79	2,758 (14.02)	2,794 (14.20)	585 (8.29)		
80–89	2,529 (12.86)	2,456 (12.48)	594 (8.42)		
>=90	950 (4.83)	871 (4.43)	253 (3.58)		
Vaccines				0.9415	<0.0001
No vaccine	17,313 (88.01)	17,297 (87.93)	1,488 (21.08)		
1 dose	673 (3.42)	670 (3.41)	321 (4.55)		
2–3 doses	1,686 (8.57)	1705 (8.67)	5,249 (74.37)		
Comorbidities					
Peripheral vascular disease	1,275 (6.48)	1,243 (6.32)	348 (4.93)	0.5231	<0.0001
Cerebrovascular disease	2,330 (11.84)	2,367 (12.03)	683 (9.68)	0.5756	<0.0001
Dementia	950 (4.83)	1,014 (5.15)	216 (3.06)	0.1447	<0.0001
Rheumatic disease	681 (3.46)	682 (3.47)	210 (2.98)	1.0000	0.0556
Peptic ulcer	742 (3.77)	754 (3.83)	234 (3.32)	0.7718	0.0860
Liver disease				0.8799	0.2336
Mild	1,261 (6.41)	1,280 (6.51)	430 (6.09)		
Moderate/Severe	168 (0.85)	162 (0.82)	48 (0.68)		
Diabetes				0.5555	<0.0001
Yes, without organ damage	2,416 (12.28)	2,441 (12.41)	622 (8.81)		
Yes, with organ damage	639 (3.25)	603 (3.07)	158 (2.24)		
Hemiplegia/ Paraplegia	333 (1.69)	356 (1.81)	91 (1.29)	0.3978	0.0231
Kidney	2,421 (12.31)	2,359 (11.99)	727 (10.30)	0.3465	<0.0001
HIV	39 (0.20)	42 (0.21)	15 (0.21)	0.8240	0.9405
Inflammatory bowel disease	1,158 (5.89)	1,151 (5.85)	549 (7.78)	0.8976	<0.0001
Arterial hypertension	7,461 (37.93)	7,398 (37.61)	1,975 (27.98)	0.5191	<0.0001
Dyslipidemia	6,780 (34.47)	6,880 (34.97)	1,872 (26.52)	0.2945	<0.0001
Lymphoma	962 (4.89)	960 (4.88)	498 (7.06)	0.9813	<0.0001
Leukemia	60 (0.31)	70 (0.36)	25 (0.35)	0.4291	0.6124
Coagulopathy	158 (0.80)	158 (0.80)	41 (0.58)	1.0000	0.0746
				0.2378	0.1356
Gastrointestinal bleeding	335 (1.70)	367 (1.87)	101 (1.43)		
Asthma	2,905 (14.77)	2,943 (14.96)	1,181 (16.73)	0.6000	<0.0001
Cystic fibrosis	441 (2.24)	466 (2.37)	117 (1.66)	0.4201	0.0038
Interstitial lung disease	117 (0.59)	118 (0.60)	55 (0.78)	1.0000	0.1150
Tumor	1,443 (7.34)	1,398 (7.11)	525 (7.44)	0.3914	0.7965
Respiratory disease	4,205 (21.38)	4,215 (21.43)	1,628 (23.07)	0.9119	0.0034
Heart disease	1,606 (8.16)	1,603 (8.15)	461 (6.53)	0.9706	<0.0001
Heart failure	1,957 (9.95)	1,853 (9.42)	533 (7.55)	0.0791	<0.0001
Basic treatments					
Antidiabetics	2,515 (12.78)	2,492 (12.67)	657 (9.31)	0.7393	<0.0001
Cardiovascular	1,077 (5.47)	1,019 (5.18)	330 (4.68)	0.2007	<0.0108
Antihypertensive	325 (1.65)	324 (1.65)	101 (1.43)	1.0000	<0.2236
Diuretics	2,317 (11.78)	2,249 (11.43)	619 (8.77)	0.2916	<0.0001
Beta-blockers	1,965 (9.99)	1,890 (9.61)	542 (7.68)	0.2095	<0.0001
Calcium channel blockers	1,331 (6.77)	1,316 (6.69)	344 (4.87)	0.7781	<0.0001
RAAS inhibitors	4,839 (24.60)	4,842 (24.61)	1,285 (18.21)	0.9813	<0.0001
Lipid lowering drugs/statins	4,390 (22.32)	4,316 (21.94)	1,158 (16.41)	0.3753	<0.0001
NSAIDs	4,091 (20.80)	4,129 (20.99)	1,908 (27.03)	0.6464	<0.0001
Direct oral anticoagulants	4,587 (23.32)	4,477 (22.76)	1,411 (19.99)	0.1919	<0.0001
Antiplatelets	2,229 (11.33)	2,239 (11.38)	654 (9.27)	0.8863	<0.0001
Heparin	908 (4.62)	929 (4.72)	441 (6.25)	0.6327	<0.0001
Broncodilators	2,753 (13.99)	2,750 (	1,086 (15.39)	0.9768	0.0045
Immunosuppressants	385 (1.96)	358 (1.82)	152 (2.15)	0.3356	0.3371
Chronic systemic steroids	1,047 (5.32)	1,028 (5.23)	485 (6.87)	0.6847	<0.0001
Constants					
Temperature				0.2316	<0.0001
<37	13,255 (67)	13,148 (67)	5,031 (71)		
37–38	4,479 (22.77)	4,486 (22.80)	1,302 (18.45)		
>38	1,328 (6.75)	1,431 (7.27)	419 (5.94)		
Missing	610 (3.10)	607 (3.09)	306 (4.34)		
Saturation				0.8170	<0.0001
<91	1,502 (7.64)	1,455 (7.40)	263 (3.73)		
91–94	3,418 (17.37)	3,448 (17.53)	614 (8.70)		
>94	13,845 (70)	13,852 (70)	5,565 (79)		
Missing	907 (4.61)	917 (4.66)	616 (8.73)		
Diastolic systolic frequency				0.3823	<0.0001
Normal	11,550 (59)	11,411 (58)	3,941 (56)		
Medium	6,233 (31.68)	6,395 (32.51)	2,315 (32.80)		
High	1,114 (5.66)	1,101 (5.60)	411 (5.82)		

(*continued on next page*)

Table 1 (*continued*)

Variables	Train N (%)	Test N (%)	Omicron N (%)	p_values train-test	P_values train-omicron
Missing	775 (3.94)	765 (3.89)	391 (5.54)		
Output variables					
Hospitalization	11,268 (57.28)	11,213 (57.00)	2,063 (29.23)	0.5822	<0.0001
COVID related death	1,925 (9.79)	1,831 (9.31)	367 (5.20)	0.1106	<0.0001
Adverse evolution	2,934 (14.91)	2,840 (14.44)	525 (7.44)	0.1852	<0.0001

previous studies [32,33] Among the predictors of these three outcomes, we find a number of chronic pathologies identified by different studies [34–37] such as cardiovascular disease (CVD) and cerebrovascular disease (CVD), as well as diabetes, kidney and liver disease. A history of tumors has also been identified as a predictor [34–38].

In the case of CVD, the exact pathophysiology underlying the pre-existing role and poor outcome has yet to be determined [39–40]. SARS-CoV-2 is believed to infect the heart, vascular tissues, and circulating cells via ACE2 (angiotensin-converting enzyme 2), the host cell receptor for the viral spike protein [41]. However, these patients are at higher risk due to concurrent underlying conditions such as advanced age, hypertension, cardiovascular disorders such as arrhythmia, diabetes, etc. These patients are also at risk of developing cardioembolic events, secondary to viral and bacterial infections or new

Table 2
Multivariable predictive model of death within COVID 19 patients who attended ERs (N = 46,402).

Variables	Beta (99 % CI)	OR (99 % CI)	p	Importance	Score
Sociodemographic variables					
Gender					
Female	Ref	Ref	–	12.60	–
Male	0.46 (0.3–0.62)	1.58 (1.35–1.85)	<0.0001		2
Age (years) categorized				112.45	
18–39	Ref	Ref	–		–
40–49	2.09 (0.51–3.68)	8.11 (1.66–39.65)	<0.001		8
50–59	2.77 (1.24–4.3)	15.96 (3.47–73.42)	<0.0001		10
60–69	3.79 (2.29–5.3)	44.39 (9.87–199.61)	<0.0001		14
70–79	4.58 (3.09–6.08)	97.95 (21.91–437.92)	<0.0001		17
80–89	5.46 (3.96–6.96)	235.23 (52.68–1050.43)	<0.0001		20
>=90	6.16 (4.66–7.67)	474.42 (105.41–2135.23)	<0.0001		23
Vaccines				6.71	
2–3 doses	Ref	Ref	–		–
1 dose	0.22 (−0.27–0.71)	1.24 (0.76–2.03)	0.26		0
No dose	0.42 (0.17–0.66)	1.52 (1.19–1.94)	<0.0001		2
Comorbidities					
Dementia	0.93 (0.71–1.15)	2.53 (2.04–3.14)	<0.0001	44.04	3
Liver disease				–	
No	Ref	Ref	–		–
Mild	0.27 (0.02–0.53)	1.31 (1.02–1.7)	<0.01		1
Moderate/Severe	1.13 (0.59–1.67)	3.1 (1.8–5.33)	<0.0001		4
Hemiplegia/Paraplegia	0.75 (0.35–1.15)	2.12 (1.42–3.15)	<0.0001	6.53	3
Kidney	0.43 (0.26–0.6)	1.54 (1.29–1.83)	<0.0001	8.22	2
Coagulopathy	0.92 (0.35–1.49)	2.51 (1.42–4.44)	<0.0001	5.16	3
Interstitial lung disease	0.65 (0.04–1.26)	1.91 (1.04–3.51)	<0.01	10.67	2
Tumor	0.77 (0.56–0.98)	2.16 (1.75–2.67)	<0.0001	15.56	3
Heart failure	0.5 (0.32–0.67)	1.64 (1.37–1.96)	<0.0001	24.39	2
Basic treatments					
Chronic systemic steroids	0.76 (0.51–1.01)	2.14 (1.67–2.75)	<0.0001	8.09	3
Constants at ER					
Temperature				10.89	
<37	Ref	Ref	–		–
37–38	0.28 (0.1–0.46)	1.32 (1.11–1.58)	<0.0001		1
>38	0.52 (0.24–0.79)	1.68 (1.27–2.21)	<0.0001		2
Missing	0.54 (0.14–0.93)	1.71 (1.15–2.55)	<0.001		2
SpO2				40.31	
>94	Ref	Ref	–		–
90–94	0.57 (0.39–0.75)	1.77 (1.48–2.12)	<0.0001		2
<90	1.53 (1.33–1.73)	4.62 (3.77–5.66)	<0.0001		6
Missing	0.56 (0.13–0.99)	1.75 (1.14–2.69)	<0.001		2
Heart rate and blood pressure				13.89	
Normal	Ref	Ref	–		–
Medium	0.36 (0.2–0.53)	1.44 (1.22–1.7)	<0.0001		1
High	0.42 (0.09–0.75)	1.52 (1.1–2.11)	<0.001		2
Missing	0.13 (−0.31–0.57)	1.13 (0.73–1.76)	0.46		0
	Train (99 %CI)		Test (99 %CI)		Omicron (99 %CI)
AUC	0.9126 (0.9056–0.9195)		0.9109 (0.9035–0.9183)		0.9285 (0.9160–0.9411)
AUC score continuous	0.9112 (0.9042–0.9182)		0.9103 (0.9029–0.9177)		0.9278 (0.9154–0.9403)
AUC score categorical (16,21,25,29)	0.9044 (0.8966–0.9122)		0.9029 (0.8950–0.9108)		0.9144 (0.8970–0.9317)
AUPRC	0.5222 (0.4928–0.5514)		0.4845 (0.4545–0.5146)		0.3995 (0.3358–0.4668)
AUPRC_score continuous	0.5184 (0.4891–0.5477)		0.4818 (0.4519–0.512)		0.3984 (0.3348–0.4657)
AUPRC score categorical (16,21,25,29)	0.4676 (0.4384–0.4969)		0.4395 (0.4098–0.4695)		0.3328 (0.2727–0.3989)

SpO2: Oxygen saturation.
Risk score range: 0–62. Cut-off points of categorical scale at 16, 21, 25, 29.

Table 3
Multivariable predictive model of adverse evolution within COVID 19 patients who attended ER (N = 46,402).

Variables	Beta (99 % CI)	OR (99 % CI)	p	Importance	Score
Sociodemographic variables					
Gender				15.92	
Female	Ref	Ref	–		–
Male	0.56 (0.43–0.69)	1.75 (1.54–1.98)	<0.0001		2
Age (years) categorized				83.53	
18–39	Ref	Ref	–		–
40–49	0.73 (0.36–1.1)	2.08 (1.44–3)	<0.0001		3
50–59	1.18 (0.84–1.52)	3.26 (2.33–4.57)	<0.0001		4
60–69	1.63 (1.3–1.96)	5.12 (3.68–7.11)	<0.0001		6
70–79	1.83 (1.5–2.16)	6.22 (4.47–8.64)	<0.0001		6
80–89	2.15 (1.82–2.49)	8.62 (6.17–12.06)	<0.0001		7
>=90	2.86 (2.49–3.23)	17.39 (12.01–25.18)	<0.0001		10
Vaccines				5.46	
2–3 doses	Ref	Ref	–		–
1 dose	0.14 (−0.27–0.55)	1.15 (0.76–1.74)	0.38		0
No dose	0.48 (0.26–0.7)	1.61 (1.3–2.01)	<0.0001		2
Comorbidities					
Dementia	0.77 (0.55–0.98)	2.15 (1.73–2.67)	<0.0001	29.94	3
Hemiplegia/Paraplegia	0.52 (0.16–0.88)	1.68 (1.17–2.41)	<0.001	6.16	2
Kidney	0.42 (0.26–0.57)	1.52 (1.3–1.77)	<0.0001	15.76	1
Coagulopathy	0.95 (0.44–1.46)	2.58 (1.55–4.29)	<0.0001	12.57	3
Cystic fibrosis	0.42 (0.12–0.73)	1.53 (1.12–2.08)	<0.001	16.76	1
Interstitial lung disease	0.59 (0.02–1.16)	1.8 (1.02–3.2)	<0.01	13.68	2
Tumor	0.58 (0.39–0.77)	1.78 (1.47–2.16)	<0.0001	15.69	2
Heart failure	0.38 (0.21–0.55)	1.46 (1.23–1.73)	<0.0001	22.73	1
Basic treatments					
Chronic systemic steroids	0.53 (0.3–0.75)	1.69 (1.36–2.11)	<0.0001	8.93	2
Constants at ER					
Temperature				19.72	
<37	Ref	Ref	–		–
37–38	0.37 (0.22–0.51)	1.44 (1.25–1.66)	<0.0001		1
>38	0.75 (0.54–0.96)	2.11 (1.71–2.6)	<0.0001		3
Missing	0.74 (0.43–1.05)	2.1 (1.54–2.87)	<0.0001		3
SpO2				102.49	
>94	Ref	Ref	–		–
90–94	0.92 (0.78–1.07)	<0.0001	<0.0001		3
<90	2.09 (1.92–2.27)	8.11 (6.8–9.66)	<0.0001		7
Missing	0.52 (0.17–0.87)	1.68 (1.18–2.38)	<0.001		2
Heart rate and blood pressure				16.22	
Normal	Ref	Ref	–		–
Medium	0.29 (0.16–0.43)	1.34 (1.17–1.53)	<0.0001		1
High	0.42 (0.16–0.69)	1.53 (1.17–1.98)	<0.0001		1
Missing	0.04 (−0.32–0.4)	1.04 (0.72–1.5)	0.78		0
	Train (99 %CI)		Test (99 %CI)		Omicron (99 %CI)
AUC	0.8540 (0.8450–0.8629)		0.8443 (0.8347–0.8539)		0.8951 (0.8824–0.9078)
AUC_score continuous	0.8473 (0.8381–0.8565)		0.8380 (0.8283–0.8478)		0.8858 (0.8681–0.9035)
AUC score categorical (7,10,13,17)	0.8402 (0.8308–0.8496)		0.8289 (0.8188–0.8390)		0.8763 (0.8568–0.8959)
AUPRC	0.5107 (0.487–0.5345)		0.4725 (0.4485–0.4967)		0.4215 (0.3672–0.4777)
AUPRC_score continuous	0.4997 (0.476–0.5235)		0.4636 (0.4396–0.4877)		0.3987 (0.3452–0.4548)
AUPRC score categorical (7,10,13,17)	0.4674 (0.4438–0.4912)		0.436 (0.4122–0.4601)		0.3683 (0.3159–0.4239)

SpO2: Oxygen saturation.
Risk score range: 0–42. Cut-off points of categorical scale at 7, 10, 13, 17.

cerebrovascular events secondary to thrombotic microangiopathy, hypercoagulability leading to macro and microthrombus formation in the vessels, hypoxic injury and blood–brain barrier disruption [40]. Likewise, acute cardiac injury is a common extrapulmonary manifestation of COVID-19 with possible chronic consequences [41] and is more prevalent amongst patients with advanced age, a functionally impaired immune system or high levels of ACE2, or patients with CVD predisposed to COVID-19 [39].

Possible pathogenetic links between diabetes mellitus and COVID-19 include effects on glucose homeostasis, inflammation, altered immune status, and activation of the renin-angiotensin-aldosterone system (RAAS) [42].

In the case of patients with renal disease, most cases of fatality were related to end-stage renal disease (ESRD). This could be partly explained by immune system dysfunction and high frequency of underlying comorbidities such as hypertension, CVD, and diabetes in ESRD patients. Generally, chronic kidney disease (CKD) is associated with an increased risk of pneumonia and a high pneumonia-related mortality rate. Moreover, the results of two recent meta-analyses reveal a significant association between preexisting CKD and severe COVID-19. CKD has been associated with inflammatory status and impaired immune system, as well as a result of over-expression of ACE2 receptor in the tubular cells of patients with CKD [43].

Any explanations of the relationship between patients with liver disease and adverse evolution of COVID-19 infection remain controversial. Some studies have shown that patients with a pre-existing hepatic disease have an increased risk of severe COVID-19 infection and higher mortality, which might be correlated with low platelets and lymphocytes in those patients. This may be due to cirrhosis-associated immune dysfunction. Additionally, it has been postulated that liver impairment in COVID-19 patients could also be drug-related and induced when treating COVID-19 infection [44].

With regard to cancer patients, some analyses of clinical outcomes in different cancer types indicate that the case fatality rate is higher in lung or hematological cancer than other solid cancers. In any case, the occurrence of severe events and death in cancer patients with COVID-19

Table 4
Multivariable predictive Lasso model of hospital admission within COVID 19 patients who attended ER (N = 46,402).

Variables	Beta (99 % CI)	OR (99 % CI)	p	Importance	Score
Sociodemographic variables					
Gender				29.46	
Female	Ref	Ref	–		
Male	0.4 (0.31–0.5)	1.5 (1.37–1.64)	<0.0001		2
Age (years) categorized				177.63	
18–39	Ref	Ref	–		
40–49	0.72 (0.57–0.87)	2.05 (1.77–2.38)	<0.0001		4
50–59	1.18 (1.03–1.32)	3.25 (2.8–3.76)	<0.0001		6
60–69	1.65 (1.49–1.81)	5.22 (4.43–6.13)	<0.0001		9
70–79	2.09 (1.91–2.28)	8.12 (6.73–9.8)	<0.0001		11
80–89	2.35 (2.13–2.56)	10.45 (8.44–12.92)	<0.0001		13
>=90	2.26 (1.97–2.56)	9.62 (7.15–12.95)	<0.0001		12
Vaccines				40.92	
2–3 doses	Ref	Ref	–		
1 dose	0.02 (−0.27–0.31)	1.02 (0.76–1.36)	0.86		0
No dose	0.88 (0.71–1.04)	2.4 (2.03–2.84)	<0.0001		5
Comorbidities					
Liver disease				23.02	
No	Ref	Ref	–		
Mild	0.29 (0.09–0.49)	1.34 (1.1–1.64)	<0.001		2
Moderate/Severe	1.29 (0.61–1.96)	3.62 (1.84–7.12)	<0.0001		7
Arterial hypertension	0.31 (0.19–0.42)	1.36 (1.21–1.52)	<0.0001	61.30	2
Cystic fibrosis	0.52 (0.13–0.91)	1.68 (1.14–2.48)	<0.001	18.64	3
Tumor	0.24 (0.05–0.43)	1.27 (1.05–1.54)	<0.01	19.62	1
Heart failure	0.52 (0.31–0.73)	1.68 (1.37–2.07)	<0.0001	39.61	3
Basic treatments					
Immunosuppresants	0.78 (0.39–1.16)	2.18 (1.48–3.2)	<0.0001	32.91	4
Chronic systemic steroids	0.33 (0.09–0.56)	1.39 (1.1–1.75)	<0.001	24.10	2
Constants at ER					
Temperature				32.33	
<37	Ref	Ref	–		
37–38	0.46 (0.34–0.57)	1.58 (1.41–1.77)	<0.0001		2
>38	1.02 (0.82–1.23)	2.79 (2.27–3.42)	<0.0001		6
Missing	0.08 (−0.19–0.35)	1.08 (0.83–1.42)	0.45		0
SpO2				165.66	
>94	Ref	Ref	–		
90–94	1.61 (1.46–1.76)	5 (4.29–5.81)	<0.0001		9
<90	2.4 (2.07–2.73)	11.02 (7.9–15.38)	<0.0001		13
Missing	0.35 (0.1–0.59)	1.41 (1.11–1.81)	<0.001		2
Heart rate and blood pressure				29.39	
Normal	Ref	Ref	–		
Medium	0.18 (0.08–0.29)	1.2 (1.08–1.33)	<0.0001		1
High	0.39 (0.17–0.6)	1.47 (1.19–1.83)	<0.0001		2
Missing	−0.17 (−0.45–0.12)	0.84 (0.64–1.12)	0.12		0
	Train (99 %CI)		Test (99 %CI)		
AUC	0.8347 (0.8273–0.8420)		0.8327 (0.8253–0.8401)		0.8438 (0.8302–0.8575)
AUC_score continuous	0.8251 (0.8175–0.8327)		0.8227 (0.8152–0.8303)		0.8376 (0.8237–0.8516)
AUC_ score categorical (9,14,19,24)	0.8204 (0.8127–0.8281)		0.8178 (0.8100–0.8255)		0.8292 (0.8147–0.8437)
AUPRC	0.864 (0.8554–0.8721)		0.8641 (0.8555–0.8722)		0.7326 (0.7068–0.7569)
AUPRC_score continuous	0.8563 (0.8475–0.8646)		0.8562 (0.8475–0.8645)		0.7205 (0.6943–0.7452)
AUPRC score categorical (9,14,19,24)	0.8486 (0.8397–0.8571)		0.8454 (0.8364–0.854)		0.7031 (0.6765–0.7283)

SpO2: Oxygen saturation.
Risk score range: 0–62. Cut-off points of categorical scale at 9, 14, 19, 2.

appears to be primarily accentuated by age, sex, and coexisting comorbidities [36].

As for less prevalent diseases such as ILD and cystic fibrosis, fewer studies have been conducted in this field. However, patients with ILD are more susceptible to COVID-19 and experience more severe evolution as compared to those without ILD, and clinicians should therefore be aware of the increased risk of COVID-19 in their ILD patients and manage or educate them appropriately during the COVID-19 pandemic [45].

With regard to treatment, chronic or recurrent use of systemic steroids prior to SARS-CoV-2 infection is a major risk factor for poor outcome and worse survival in asthmatics [46], and clinicians treating patients should therefore follow current guidelines carefully [44], achieve asthma control and reduce the need for chronic or recurrent systemic steroid therapy [46]. However, there are studies showing that patients undergoing biologic therapy for severe allergic and eosinophilic asthma do not have an increased risk of SARS-CoV-2 infection or severity. We therefore believe that the fact that use of chronic systemic corticosteroids is related to these results may be linked to a greater alteration in these patients' immunity [46].

On relation to COVID-19 vaccination we show that having no vaccination increases the risk of all outcomes, as in other studies outlining the importance of the vaccination in preventing adverse evolution [47].

Dementia appears as a potential risk factor in many studies. There are many possible explanations for this observed increase in risk. Changes in health care delivery may disproportionately affect older adults with ADRD [48]. Patients with dementia have higher vulnerability, which may be due to living conditions in nursing homes, need for intensive caregiver assistance, and to the inability to self-isolate and manage preventative health measures. As hypotheses, the presence of chronic inflammatory conditions or defective immune responses in patients with dementia may increase their vulnerability to infection or reduce their ability to mount effective responses to infection [49].

Most previous studies have also shown that age and sex (male) are

Table 5
True positive rate (TPR), True negative rate (TNR) and Net Benefit (NB) according to different cutoff points in train, test and Omicron samples, reported in percentages for COVID 19 patients who attended ER (N = 46,402).

	TRAIN Sample					TEST Sample					OMICRON Sample				
	TPR	TNR	NB	F1	BA	TPR	TNR	NB	F1	BA	TPR	TNR	NB	F1	BA
Death Model based on 5 risk groups															
Score ≥ 16	99.06	48.95	9.01	29.59	74.01	98.85	49.73	8.48	28.71	74.29	97.28	68.88	4.52	25.45	83.08
Score ≥ 21	94.96	70.20	7.73	40.44	82.58	94.70	70.50	7.24	39.28	82.60	88.56	83.19	3.62	35.77	85.88
Score ≥ 25	82.08	83.99	5.48	49.79	83.04	82.80	84.58	5.22	49.72	83.69	71.93	91.54	2.08	44.11	81.74
Score ≥ 29	54.49	93.64	2.52	51.13	74.07	54.61	93.67	2.38	50.51	74.14	39.51	96.82	0.62	40.00	68.17
Adverse Evolution Model based on 5 risk groups															
Score ≥ 7	98.47	28.37	12.47	32.44	63.42	97.92	28.39	11.74	31.47	63.16	96.38	53.83	5.96	25.01	75.11
Score ≥ 10	91.55	55.12	10.25	40.91	73.34	90.99	55.38	9.63	39.96	73.19	84.38	77.59	4.57	36.43	80.99
Score ≥ 13	75.39	77.93	7.13	50.05	76.66	73.27	77.99	6.55	48.25	75.63	62.48	90.59	2.77	44.69	76.54
Score ≥ 17	41.68	93.43	2.58	46.54	67.56	41.41	93.38	2.44	45.84	67.40	28.00	97.34	0.55	34.75	62.67
Hospital Admission Model based on 5 risk groups															
Score ≥ 9	95,47	28,62	44.28	76.77	62.05	94,94	29,4	43.65	76.50	62.17	87,78	55,78	19.76	59.54	71.78
Score ≥ 14	83,15	62,15	34.67	78.67	72.65	82,4	62,42	33.92	78.20	72.41	72,42	81,14	14.44	66.41	76.78
Score ≥ 19	65,15	83,16	24.13	73.32	74.16	64,35	83,98	23.91	72.95	74.17	49,88	93,57	8.86	60.30	71.73
Score ≥ 24	44,12	93,38	13.16	59.20	68.75	43,76	94,08	13.63	59.05	68.92	31,17	97,72	4.48	45.60	64.45

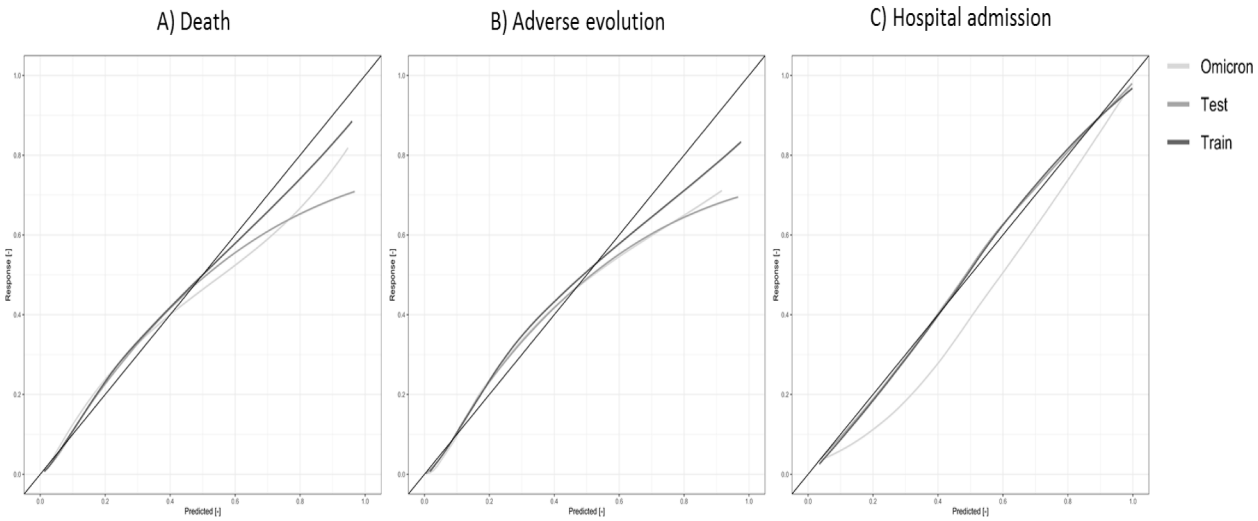


Fig. 2. Bar chart representing the probability of risk for each of the outcomes (Death, Adverse Evolution and Hospital Admission) and the five risk groups in the Omicron sample COVID 19 patients who attended ER (N = 46,402).

significant risk factors for adverse evolution [34,37,50]. A higher proportion of men than women have died, which could be partly explained by the greater effect of age among men [51]. Furthermore, it has been hypothesized that age-related decline and dysregulation of the immune function, i.e., immunosenescence and inflammation, may play an important role in contributing to increased vulnerability to severe COVID-19 outcomes in older adults [52]. As for sex, immunological differences suggest that women mount a rapid and aggressive innate immune response, and angiotensin-converting enzyme 2 (ACE2) is involved in disease pathogenesis in cardiovascular disease and COVID-19, either to serve as a protective mechanism by deactivating the RAS or as the receptor for viral entry, respectively [53]. Furthermore, circulating sex hormones in men and women could influence susceptibility to COVID-19 infection, as demonstrated in a previous study, since they modulate adaptive and innate immunity responses [51].

Finally, in cases where simple basic vital signs data is available, we show that alteration of any of those constants is related to adverse evolution, improving the predictive ability of our models, findings already described in other studies [54].

Not too many studies have focused on developing predictive models for the general population [55]. Most of them center in hospital admitted patients, which may imply a bias, and their adverse outcomes, and most have as potential predictors laboratory data or combined laboratory data with other clinical data, which requires to perform previous lab test, which is not our case.

Amongst the strengths of this study are the enormous sample size, which includes all epidemics and patients in our region up to the beginning of last year, the inclusion of three outcomes, and the external validation of the models in the wave of the more recent and less severe Omicron variant. In developing all predictive models, we followed the standards of the TRIPOD guidelines [56] as well as other requirements to ensure fairness and equity of our models in terms of equal outcomes, allocation and performance of our models [57–59]. The three models are based on variables that are easy to obtain in any setting, easy to calculate and provide a quick prediction of the patient's risk. Those different prediction models will be also available in short in an easy-to-use software. As a practical proposal, patients with low scores (very low or low classes for death or adverse evolution) can safely stay at home, while those in high or very high classes should be seen at a hospital level and more intensive care should be considered. In the case of patients in the moderate class, their particular casuistry in terms of age, baseline comorbidities, and clinical presentation should be individually analyzed. In order to facilitate decision making in practice, we have developed a very easy to use shiny application, which incorporates the models developed and allows to identify the risk based on the categorized score of each patient. In any case, the clinical judgment for each

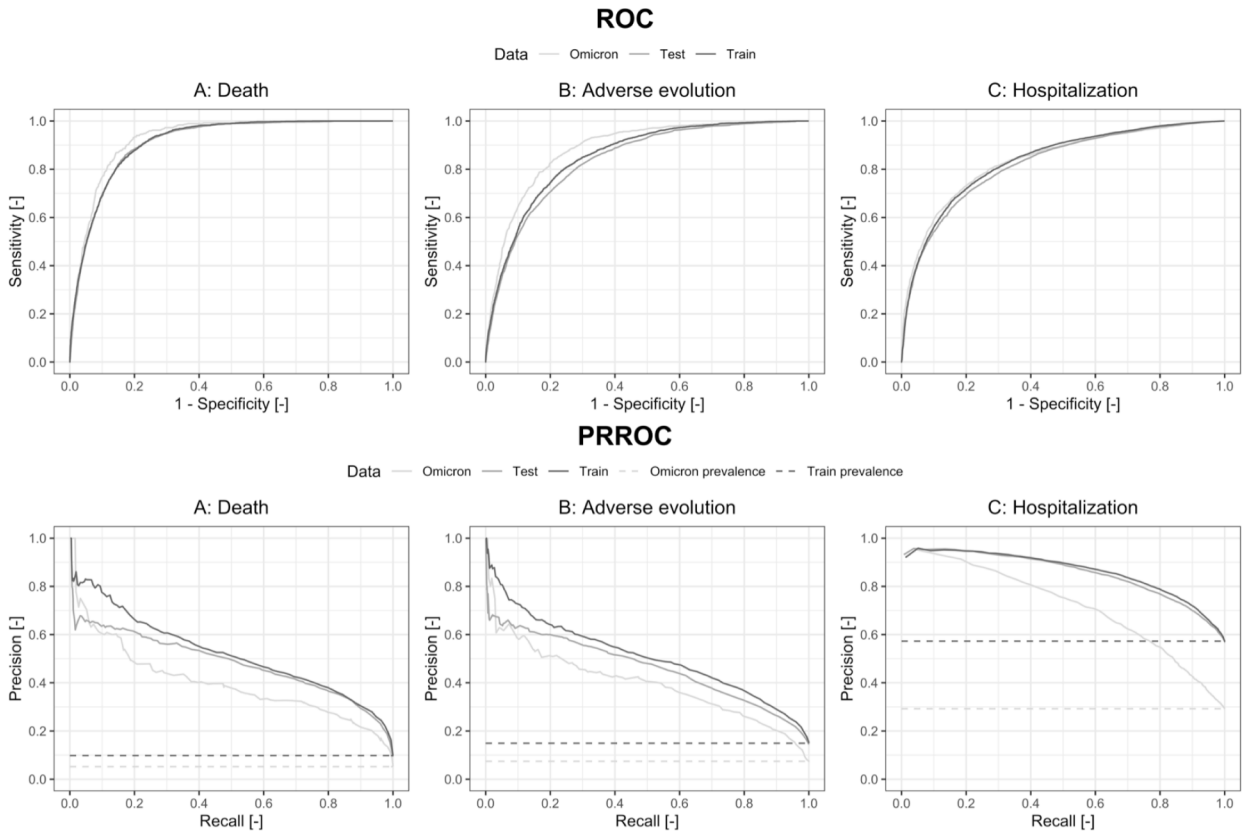


Fig. 3. ROC and Precision-Recall Curves for the continuous score for COVID 19 patients who attended ERs adjusted in the Train, Test and Omicron samples, for each of the outcomes: A) Death, B) Adverse Evolution and C) Hospital Admission.

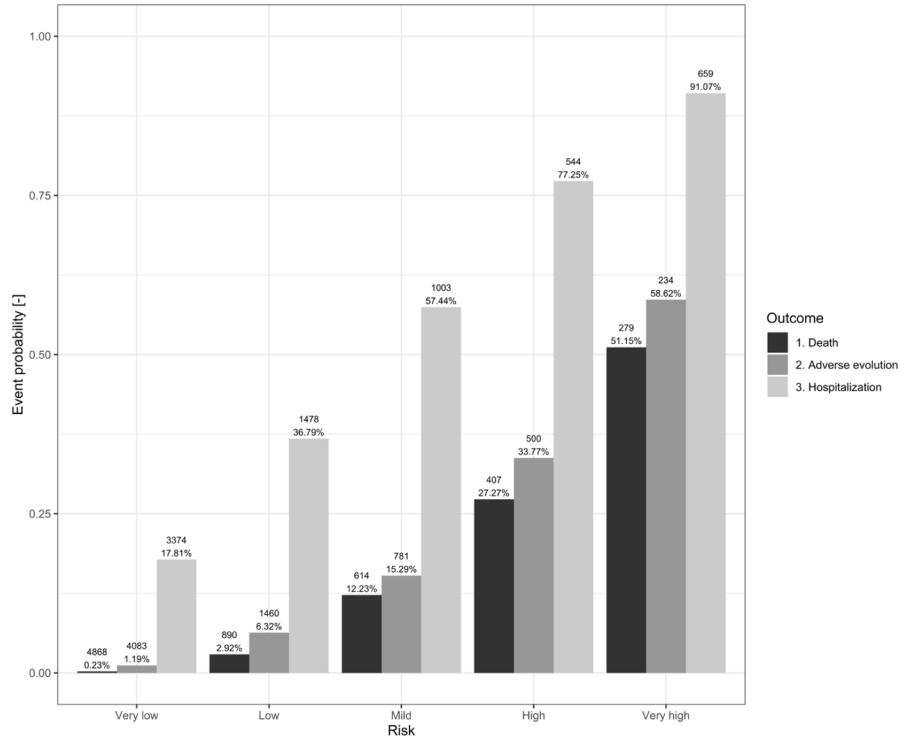


Fig. 4. Calibration plots for the prediction model for COVID 19 patients who attended ERs adjusted in the Train, Test and Omicron samples, for each of the outcomes: A) Death, B) Adverse Evolution and C) Hospital Admission.

individual patient should prevail. Regarding the limitations, our data is limited to baseline diseases and treatments plus sociodemographic data, without subsequent clinical follow-up information on those admitted. It was decided to proceed in this way in order to select the basic information we believed to be most reliable and easiest to obtain in any setting. Calibration plots show an overestimation of the probability of event in the Omicron sample which makes sense in part because the prevalence of the outcomes with this variant is statistically lower. Nonetheless, the AUC of all models is very high, even in the case of hospitalized patients, and is replicated in the Omicron sample, and good (small) Brier scores were obtained.

These analyses provide very useful practical tools both in the field of primary care and in emergency and hospital settings for making decisions on follow-up and treatment of these patients, including during the current Omicron wave. This may allow better clinical follow-up and case management.

5. Authors' contribution

Drs Jose M. Quintana, Janire Portuondo-Jiménez, Pedro P. España and Irantzu Barrio had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Quintana, Portuondo-Jiménez, García-Gutierrez and Irantzu Barrio.

Acquisition, analysis, or interpretation of data: Julia García, María José Legarreta, Ane Villanueva, María Gascón, Lander Rodríguez, Nere Larrea, Irantzu Barrio Quintana and the COVID-Health Basque Country Research Group.

Drafting of the manuscript: Quintana, Portuondo-Jiménez, España, García-Gutierrez and Irantzu Barrio.

Critical revision of the manuscript for important intellectual content: Quintana, Portuondo-Jiménez, España, García-Gutierrez, Julia García, María José Legarreta, Ane Villanueva, María Gascón, Lander Rodríguez, Nere Larrea, Irantzu Barrio, and the COVID-Health Basque Country Research Group.

Statistical analysis: Rodríguez, Quintana, Portuondo-Jiménez, Barrio.

Obtained funding: Portuondo-Jiménez, Garcia-Gutierrez, España, Quintana and Barrio.

Administrative, technical, or material support: Julia García, Legarreta, Villanueva, Gascón, Rodríguez, Larrea, Barrio, and the COVID-Health Basque Country Research Group group.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful for the support of the Basque health service, Osakidetza, and the Department of Health of the Basque Government, and to Tim Nicholson for English language editing. We also gratefully acknowledge the patients who participated in the study. Open access funding provided by the University of the Basque Country.

Funding

This work was supported in part by the health outcomes group from Galdakao-Barrualde Health Organization; the Kronikgune Institute for Health Service Research; and the thematic network–REDISSEC (Red de Investigación en Servicios de Salud en Enfermedades Crónicas)–of the Instituto de Salud Carlos III. The work of IB was financially supported in part by grants from the Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco [IT1456-22] and by the Ministry of Science and Innovation through BCAM Severo Ochoa accreditation [CEX2021-001142-S/MICIN/AEI/10.13039/501100011033] and through project [PID2020-115882RB-I00/AEI/10.13039/501100011033] funded by Agencia Estatal de Investigación and acronym "S3M1P4R" and also by the Basque Government through the BERC 2022–2025 program and the BMTF ''Mathematical Modeling Applied to Health'' Project.

Ethical approval

The study protocol was approved by the Ethics Committee of the Basque Country (reference PI2020059).

Registration: ClinicalTrials.gov Identifier: NCT04463706.

COVID-Health Basque Country Research Group: Janire Portuondo, Endika Munitiz, Julia Garcia (Basque Government Department of Health); Verónica Tiscar, Amaia Bilbao (Basurto University Hospital); Susana García-Gutierrez, Jose M. Quintana, Maria J. Legarreta, Ane Villanueva, María Gascón, Nere Larrea, Iratxe Lafuente, Cristóbal Esteban, Amaia Aramburu, Pedro Pablo España, Ane Uranga (Galdakao-Usansolo University Hospital); Iñaki Zorrilla (Bioaraba Health Research Institute); Irantzu Barrio (UPV/EHU); Dae-Jin Lee, Abelardo-Enrique Monsalve-Cobis, Lander Rodríguez (Basque Center for Applied Mathematics, BCAM).

Summary table.

What was already known on the topic.

- The SARS-CoV-2 infection severity is changing depending on variants.
- Different predictive models of adverse evolution have been developed but need to be updated.
- COVID-19 will not disappear in the short or medium term.

What this study added to our knowledge.

- We present models developed in a whole large sample of our area during six waves of the pandemic.
- Developed predictive models are based on variables easy to obtain in any setting, easy to calculate and provide a quick prediction tool of the patient's risk.
- Those tools can be used in the field of primary care and in emergency and hospital settings for making decisions on follow-up and treatment of these patients

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2023.105039.

References

[1] N. Zhu, et al., A novel coronavirus from patients with pneumonia in China, 2019, N. Engl. J. Med. 382 (8) (2020) 727–733, https://doi.org/10.1056/NEJMOA2001017/SUPPL_FILE/NEJMOA2001017_DISCLOSURES.PDF.

[2] R. McCabe, et al., Adapting hospital capacity to meet changing demands during the COVID-19 pandemic, BMC Med. 18 (1) (2020) 1–12, https://doi.org/10.1186/S12916-020-01781-W/FIGURES/3.

[3] WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data, https://covid19.who.int/(accessed Feb. 01, 2022).

[4] B. Sen-Crowe, M. Sutherland, M. McKenney, A. Elkbuli, A closer look into global hospital beds capacity and resource shortages during the COVID-19 pandemic, J. Surg. Res. 260 (2021) 56, https://doi.org/10.1016/J.JSS.2020.11.062.

[5] A. Parasher, COVID-19: current understanding of its pathophysiology, clinical presentation and treatment, Postgrad. Med. J. 97 (1147) (2021) 312–320, https://doi.org/10.1136/POSTGRADMEDJ-2020-138577.

[6] B.M. Henry, M.H.S. de Oliveira, S. Benoit, M. Plebani, G. Lippi, Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis, Clin.

Chem. Lab. Med. 58 (7) (2020) 1021–1028, https://doi.org/10.1515/CCLM-2020-0369/PDF.
[7] S.A. Vardhana, J.D. Wolchok, The many faces of the anti-COVID immune response, J. Exp. Med. 217(6) (2020), doi: 10.1084/JEM.20200678/151725.
[8] S. Bülow Anderberg et al., Increased levels of plasma cytokines and correlations to organ failure and 30-day mortality in critically ill Covid-19 patients, Cytokine 138 (2021) 155389, doi: 10.1016/J.CYTO.2020.155389.
[9] W. van Damme et al., The COVID-19 pandemic: diverse contexts; different epidemics - how and why? BMJ Glob. Health 5(7) (2020), doi: 10.1136/bmjgh-2020-003098.
[10] S. Iftimie et al., First and second waves of coronavirus disease-19: a comparative study in hospitalized patients in Reus, Spain, PLoS ONE 16(3) (2021), doi: 10.1371/JOURNAL.PONE.0248029.
[11] O. Noy, et al., A machine learning model for predicting deterioration of COVID-19 inpatients, Sci. Rep. 12 (1) (2022) Dec, https://doi.org/10.1038/S41598-022-05822-7.
[12] A. Vaid, et al., Machine learning to predict mortality and critical events in a cohort of patients with COVID-19 in New York City: model development and validation, J. Med. Internet Res. 22 (11) (2020), https://doi.org/10.2196/24018.
[13] M. Sperrin, B. Mcmillan, Prediction models for covid-19 outcomes, BMJ 371 (2020), https://doi.org/10.1136/BMJ.M3777.
[14] L. Wynants, et al., Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal, BMJ (2020), https://doi.org/10.1136/bmj.m1328.
[15] L. Famiglini, A. Campagner, A. Carobene, F. Cabitza, A robust and parsimonious machine learning method to predict ICU admission of COVID-19 patients, Med. Biol. Eng. Comput. 1 (2022) 1–13, https://doi.org/10.1007/S11517-022-02543-X/FIGURES/7.
[16] J. Montomoli, et al., Machine learning using the extreme gradient boosting (XGBoost) algorithm predicts 5-day delta of SOFA score at ICU admission in COVID-19 patients, J. Intensive Med. 1 (2) (2021) 110, https://doi.org/10.1016/J.JOINTM.2021.09.002.
[17] M.E. Charlson, F.L. Sax, C.R. MacKenzie, S.D. Fields, R.L. Braham, R.G. Douglas, Assessing illness severity: does clinical judgment work? J. Chronic Dis. 39 (6) (1986) 439–452, https://doi.org/10.1016/0021-9681(86)90111-6.
[18] H. Quan, et al., Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data, Med. Care 43 (11) (2005) 1130–1139, https://doi.org/10.1097/01.MLR.0000182534.19832.83.
[19] Guidelines for ATC classification and DDD assignment 2013, Accessed: Oct. 11, 2022, [Online], Available: www.whocc.no.
[20] A.Y.A. Amer, et al., Vital signs prediction for COVID-19 patients in ICU, Sensors (Basel) 21 (23) (2021), https://doi.org/10.3390/S21238131.
[21] C. Areia, et al., The impact of wearable continuous vital sign monitoring on deterioration detection and clinical outcomes in hospitalised patients: a systematic review and meta-analysis, Crit Care 25 (1) (2021) 1–17, https://doi.org/10.1186/S13054-021-03766-4/TABLES/4.
[22] I.J. Brekke, L.H. Puntervoll, P.B. Pedersen, J. Kellett, M. Brabrand, The value of vital sign trends in predicting and monitoring clinical deterioration: a systematic review, PLoS ONE 14(1) (2019), doi: 10.1371/JOURNAL.PONE.0210875.
[23] E.W. Steyerberg, Validation of prediction models, in: Clinical Prediction Models, Springer, New York, 2009, pp. 299–310.
[24] R. Tibshirani, Regression shrinkage and selection via the Lasso, J. Roy. Stat. Soc.: Ser. B (Methodol.) 58 (1) (1996) 267–288, https://doi.org/10.1111/J.2517-6161.1996.TB02080.X.
[25] J.A. Hanley, B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology 143 (1) (1982) 29–36, https://doi.org/10.1148/radiology.143.1.7063747.
[26] M.B. Kursa, W.R. Rudnicki, Feature selection with the Boruta package, J. Stat. Softw. 36 (11) (2010) 1–13, https://doi.org/10.18637/JSS.V036.I11.
[27] K. Boyd, K.H. Eng, C. David Page, area under the precision-recall curve: point estimates and confidence intervals, Accessed: Feb. 02, 2023, [Online], Available: http://link.springer.com/chapter/10.1007 %.
[28] T. Saito, M. Rehmsmeier, The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets, PLoS ONE 10 (3) (2015), e0118432, https://doi.org/10.1371/JOURNAL.PONE.0118432.
[29] I. Barrio, I. Arostegui, M.-X. Rodríguez-Álvarez, J.-M. Quintana, A new approach to categorising continuous variables in prediction models: proposal and validation, Stat. Methods Med. Res. 26 (6) (2017) 2586–2602, https://doi.org/10.1177/0962280215601873.
[30] I. Barrio, J. Roca-Pardiñas, I. Arostegui, Selecting the number of categories of the lymph node ratio in cancer research: a bootstrap-based hypothesis test, Stat. Methods Med. Res. 30 (3) (2021) 926–940, https://doi.org/10.1177/0962280220965631.
[31] A.J. Vickers, E.B. Elkin, Decision curve analysis: a novel method for evaluating prediction models, Med. Decis. Making 26 (6) (2006) 565–574, https://doi.org/10.1177/0272989X06295361.
[32] V. Bellou, I. Tzoulaki, M. van Smeden, K.G.M. Moons, E. Evangelou, L. Belbasis, Prognostic factors for adverse outcomes in patients with COVID-19: a field-wide systematic review and meta-analysis, Eur. Respir. J. 59 (2) (2021), https://doi.org/10.1183/13993003.02964-2020.
[33] B. Gallo Marin, et al., Predictors of COVID-19 severity: a literature review, Rev. Med. Virol. 31 (1) (2021) 1–10, https://doi.org/10.1002/RMV.2146.
[34] A.B. Docherty, et al., Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO clinical characterisation protocol: prospective observational cohort study, BMJ 369 (March) (2020) 1–12, https://doi.org/10.1136/bmj.m1985.
[35] W. Guan, et al., Clinical characteristics of coronavirus disease 2019 in China, N. Engl. J. Med. 382 (18) (2020) 1708–1720, https://doi.org/10.1056/nejmoa2002032.
[36] H. Zhang, 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, medRxiv (2020) 1–23, https://doi.org/10.1101/2020.04.28.20082222.
[37] J. Portuondo-Jimenez, et al., Modelling the risk of hospital admission of lab confirmed SARS-CoV-2-infected patients in primary care: a population-based study, Int. Emerg. Med. (2022), https://doi.org/10.1007/s11739-022-02931-z.
[38] K. Singh et al., Validating a widely implemented deterioration index model among hospitalized COVID-19 patients, medRxiv Preprint, 2020 Apr 29 [revised 2020 Jun 20]., pp. 1–13, 2020, doi: 10.1101/2020.04.24.20079012.
[39] K.J. Clerkin, et al., COVID-19 and cardiovascular disease, Circulation 141 (2020) 1648–1655, https://doi.org/10.1161/CIRCULATIONAHA.120.046941.
[40] U. Patel, P. Malik, D. Shah, A. Patel, M. Dhamoon, V. Jani, Pre-existing cerebrovascular disease and poor outcomes of COVID-19 hospitalized patients: a meta-analysis, J. Neurol. 268 (1) (2021) 240–247, https://doi.org/10.1007/S00415-020-10141-W/FIGURES/2.
[41] M.K. Chung, et al., COVID-19 and cardiovascular disease, Circ. Res. 128 (8) (2021) 1214–1236, https://doi.org/10.1161/CIRCRESAHA.121.317997.
[42] S. Lim, J.H. Bae, H.S. Kwon, M.A. Nauck, COVID-19 and diabetes mellitus: from pathophysiology to clinical management, Nat. Rev. Endocrinol. 17(1) (2020) 11–30, doi: 10.1038/s41574-020-00435-4.
[43] H. Rastad, et al., The risk factors associated with COVID-19-Related death among patients with end-stage renal disease, BMC Nephrol. 22 (1) (2021) 1–8, https://doi.org/10.1186/S12882-020-02221-W/TABLES/4.
[44] M.I. Metawea, W.I. Yousif, I. Moheb, COVID 19 and liver: an A-Z literature review, Dig. Liver Dis. 53 (2) (2021) 146–152, https://doi.org/10.1016/J.DLD.2020.09.010.
[45] H. Lee, et al., Interstitial lung disease increases susceptibility to and severity of COVID-19, Eur. Respir. J. 58 (6) (2021), https://doi.org/10.1183/13993003.04125-2020.
[46] Y. Adir, M. Humbert, W. Saliba, COVID-19 risk and outcomes in adult asthmatic patients treated with biologics or systemic corticosteroids: nationwide real-world evidence, J. Allergy Clin. Immunol. 148 (2) (2021) 361–367.e13, https://doi.org/10.1016/J.JACI.2021.06.006.
[47] T. Nyberg, et al., Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: a cohort study, Lancet 399 (10332) (2022) 1303–1312, https://doi.org/10.1016/S0140-6736(22)00462-7.
[48] L. Gilstrap, W. Zhou, M. Alsan, A. Nanda, J.S. Skinner, Trends in mortality rates among Medicare enrollees with Alzheimer disease and related dementias before and during the early phase of the COVID-19 pandemic, J. Am. Med. Assoc. Neurol. (2022), https://doi.org/10.1001/JAMANEUROL.2022.0010.
[49] A.C. Tahira, S. Verjovski-Almeida, S.T. Ferreira, Dementia is an age-independent risk factor for severity and death in COVID-19 inpatients, Alzheimer's Dementia 17 (11) (2021) 1818–1831, https://doi.org/10.1002/ALZ.12352.
[50] J.M. Casas-Rojo, et al., Clinical characteristics of patients hospitalized with COVID-19 in Spain: results from the SEMI-COVID-19 Registry, Rev. Clin. Esp. 220 (8) (2020) 480–494, https://doi.org/10.1016/j.rce.2020.07.003.
[51] S. Jaillon, K. Berthenet, C. Garlanda, Sexual dimorphism in innate immunity, Clin. Rev. Allergy Immunol. 56 (3) (2019) 308–321, https://doi.org/10.1007/s12016-017-8648-x.
[52] Y. Chen, et al., Aging in COVID-19: vulnerability, immunity and intervention, Ageing Res. Rev. 65 (2021), https://doi.org/10.1016/J.ARR.2020.101205.
[53] A. Viveiros, et al., Sex differences in COVID-19: candidate pathways, genetics of ACE2, and sex hormones, Am. J. Physiol. Heart Circ. Physiol. 320 (1) (2021) H296–H304, https://doi.org/10.1152/AJPHEART.00755.2020/ASSET/IMAGES/LARGE/AJ-AHRT200091F004.JPEG.
[54] A. Izcovich, et al., Prognostic factors for severity and mortality in patients infected with COVID-19: a systematic review, PLoS ONE 15 (2020), https://doi.org/10.1371/JOURNAL.PONE.0241955.
[55] K.M. Kuo, P.C. Talley, C.S. Chang, The accuracy of machine learning approaches using non-image data for the prediction of COVID-19: a meta-analysis, Int. J. Med. Inform. 164 (2022), https://doi.org/10.1016/J.IJMEDINF.2022.104791.
[56] G.S. Collins, J.B. Reitsma, D.G. Altman, K.G.M. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement, Ann. Intern. Med. 162 (1) (2015) 55–63, https://doi.org/10.7326/M14-0697.
[57] I. Scott, S. Carter, E. Coiera, Clinician checklist for assessing suitability of machine learning applications in healthcare, BMJ Health Care Inform. 28 (1) (2021), e100251, https://doi.org/10.1136/BMJHCI-2020-100251.
[58] A. Rajkomar, M. Hardt, M.D. Howell, G. Corrado, M.H. Chin, Ensuring fairness in machine learning to advance health equity, Ann. Intern. Med. 169 (12) (2018) 866–872, https://doi.org/10.7326/M18-1990.
[59] F. Cabitza, A. Campagner, The need to separate the wheat from the chaff in medical informatics: introducing a comprehensive checklist for the (self)-assessment of medical AI studies, Int. J. Med. Inform. 153 (2021), 104510, https://doi.org/10.1016/J.IJMEDINF.2021.104510.