




Modelling the risk of hospital admission of lab confirmed SARS-CoV-2-infected patients in primary care: a population-based study

Janire Portuondo-Jimenez^{1,2,3} · Amaia Bilbao-González^{4,5,6} · Verónica Tíscar-González^{3,4,8} · Ignacio Garitano-Gutiérrez^{7,9} · Susana García-Gutiérrez^{5,6,10} · Almudena Martínez-Mejuto^{3,11} · Jaione Santiago-Garin^{7,8} · Silvia Arribas-García^{7,8} · Julia García-Asensio¹ · Johnny Chart-Pascual¹³ · Iñaki Zorrilla-Martínez^{12,13,14}  · Jose Maria Quintana-Lopez^{5,6,10} · the COVID-19-Osakidetza Working group

Received: 17 May 2021 / Accepted: 13 January 2022 / Published online: 10 February 2022
© The Author(s), under exclusive licence to Società Italiana di Medicina Interna (SIMI) 2022

Abstract

The objectives of this study are to develop a predictive model of hospital admission for COVID-19 to help in the activation of emergency services, early referrals from primary care, and the improvement of clinical decision-making in emergency room services. The method is the retrospective cohort study of 49,750 patients with microbiological confirmation of SARS-CoV-2 infection. The sample was randomly divided into two subsamples, for the purposes of derivation and validation of the prediction rule (60% and 40%, respectively). Data collected for this study included sociodemographic data, baseline comorbidities, baseline treatments, and other background data. Multilevel analyses with generalized estimated equations were used to develop the predictive model. Male sex and the gradual effect of age were the main risk factors for hospital admission. Regarding baseline comorbidities, coagulopathies, cancer, cardiovascular diseases, diabetes with organ damage, and liver disease were among the five most notable. Flu vaccination was a risk factor for hospital admission. Drugs that increased risk were chronic systemic steroids, immunosuppressants, angiotensin-converting enzyme inhibitors, and NSAIDs. The AUC of the risk score was 0.821 and 0.828 in the derivation and validation samples, respectively. Based on the risk score, five risk groups were derived with hospital admission ranging from 2.94 to 51.87%. In conclusion, we propose a classification system for people with COVID-19 with a higher risk of hospitalization, and indirectly with it a greater severity of the disease, easy to be completed both in primary care, as well as in emergency services and in hospital emergency room to help in clinical decision-making.

Registration: ClinicalTrials.gov Identifier: NCT04463706.

Keywords Coronavirus infections · Risk factors · Primary health care · Emergencies · Epidemiology

Introduction

In November 2019, a number of patients infected by a beta-coronavirus appeared in Wuhan (China), and in December, an outbreak of pneumonia associated with a new type of CoV was reported. On the 12th of February 2020, the World Health Organization (WHO) classified the illness as COVID-19 [1]. The first known case appeared in Spain on the 31st of January, and a few weeks later, the virus was

diagnosed for the first time in the Basque Country. In view of the rapid expansion of the virus [2], a State of Alarm was declared by the Spanish Government on the 13th of March, and a series of measures were introduced to restrict movement with the aim of containing the propagation of the virus [3].

Regarding risk factors, it is known that older aged patients, males, comorbidities such as arterial Hypertension (HTN) diabetes, cardiovascular, lung and cerebrovascular problems [4], immunosuppressant illnesses [5], high Body Mass Index (BMI) [6], or cancer [4] were all associated with poorer results or the seriousness of patients condition due to COVID-19. The WHO warned of higher mortality rates for patients who smoked [7] and patients who has smoked in the

✉ Iñaki Zorrilla-Martínez
inaki.zorrillamartinez@osakidetza.eus

Extended author information available on the last page of the article

past caused a higher prediction of admission into Intensive care unit (ICU), although not leading to a higher mortality rates [8]. Various studies indicate that the uses of different drugs were associated with this prognosis, among them, Angiotensin-converting enzyme inhibitors (ACE inhibitors) and Angiotensin II receptor blockers type I (ARBs) [5].

Regarding the pneumococcal polysaccharide vaccine and influenza vaccination, a certain degree of protection has been observed given that the Toll-Like receptor is of fundamental importance in the binding of respiratory single-stranded ARN virus such as SARS-CoV-2, thereby helping to explain the cross-protection [9].

To offer an individualized prediction with regard to the risk of the development of the disease, numerous publications have analyzed the clinical characteristics and offered predictive models for the prognostic of patients with COVID-19 [10–12]. Such models are of great use when taking clinical decisions and guarantee both a better and earlier handling of cases by Primary attention (PA) or Hospital Emergency Services leading to improved levels of recovery and reducing mortality rates. Therefore, patients requiring hospitalization are lower, helping to avoid the saturation of the health system allowing for a more efficient use of the system and reducing the consequences of late patient referrals.

The Basque Health Service-Osakidetza has one, uniform system for the collection, storage, and use of data obtained during the clinical assistance provided, and through the present study, we wish to identify the epidemiological characteristics of infected patients in the general population to establish patterns and systems of early referrals from Primary Attention to the different hospitals and the activation of Emergency units and the improvement of clinical decisions by the Emergency services.

Methodology

This is a retrospective study of a cohort of patients diagnosed with COVID-19 in the Basque Country based on data from the electronic database and health records of the Basque Health Service-Osakidetza. This Spanish region has a population of 2178 million people, the vast majority of whom are entitled to healthcare under Osakidetza. The Basque Health System is divided into 13 integrated healthcare organizations (IHOs), gathering all primary and hospital care resources in given areas under the same administrative management.

All patients included in this study were residents in the Basque Country with SARS-CoV-2 infection that was 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 IgM or IgG antibody test performed due to symptoms suggestive of the disease or having had contact

with a positive case, from February to September 25th, 2020. No patients were excluded. The study protocol was approved by the Ethics Committee of the Basque Country (reference PI2020059). All patient data were kept confidential.

All data on patients under the care of Osakidetza are recorded in a unified electronic database. Analysts retrieved data from all positive cases detected during the study period including sociodemographic data (age, sex, and place of residence), smoking habits, BMI, hospital admission in the previous month to the COVID-19 diagnosis, baseline comorbidities [all those considered in Charlson's Comorbidity index [13] plus angina, arrhythmia, arterial hypertension, dyslipidemia, asthma, bronchiectasis, cystic fibrosis, interstitial lung disease, lymphoma, leukemia, coagulopathy, inflammatory bowel disease, gastrointestinal bleeding], flu vaccination, pneumococcal vaccine, baseline treatments [based on the Anatomical, Therapeutic, Chemical (ATC) classification system] [14], and other background data concerning care provided in hospital or primary care settings including dates of hospital admission. Comorbidities were identified based on the International Statistical Classification of Diseases and Related Health Problems (ICD) ICD-9 or 10 codes in patients' records at baseline [15].

We grouped comorbidities as follows: cardiovascular diseases (including myocardial infarction, angina, arrhythmia, congestive heart failure, and peripheral vascular disease); cerebrovascular disease, hemiplegia and/or paraplegia; arterial hypertension; dyslipidemia; dementia; respiratory disease [chronic obstructive pulmonary disease (COPD), bronchiectasis, chronic bronchial infection]; asthma; liver disease (mild liver, moderate or severe liver disease); diabetes (diabetes with/without organ damage), kidney disease, cancer(malignant tumor, metastatic solid tumor, lymphoma); rheumatic disease; peptic ulcer; inflammatory bowel disease; and coagulopathies.

Regarding baseline medication, we selected drugs based on ATC codes [14]. Baseline treatment was defined as any drugs prescribed before the patient was diagnosed with SARS-CoV-2 infection and had no end date. A summary information about the grouping of baseline treatments included in this study was included elsewhere [16].

Data identifying people living in a nursing home were obtained from the Basque Health Department. The outcomes of the study were hospital admission during the study period.

Additionally, we studied the validity of our electronic database by comparing the data obtained from it for subset variables with the information provided by a group of trained reviewers who retrieved the same information, item by item, from the electronic health records of the same patients. The results are provided elsewhere [16].

Statistical analysis

The sample was randomly divided into two subsamples, for derivation and validation purposes of the prediction rule (60% and 40% of the entire sample, respectively).

Descriptive statistics included frequency tables for categorical variables and means and standard deviations (SDs) for continuous variables. Patient characteristics were compared between the two subsamples (derivation vs. validation) using Chi-square or Fisher's exact tests for categorical variables, and Student's *t* test or nonparametric Wilcoxon tests for continuous variables.

Univariate logistic regression models were first built using the derivation sample to identify the significance of each potential risk factor for predicting hospital admission. In these models, hospital admission was used as the dependent variable and all candidate predictive variables (described previously) as the independent variables. Then, independent variables with a $p < 0.20$ in the univariate analyses were considered potential independent variables in the multivariate analysis, for which multilevel analysis with generalized estimated equations was performed considering the IHO. Potential interactions between variables were also examined. In the final models, only factors with $p < 0.05$ were retained. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated. The predictive accuracy of the model was determined by calculating area under the ROC curve (AUC) for discrimination [17], and by comparing predicted and observed hospital admission using the Hosmer and Lemeshow test for calibration [18].

To develop the predictive risk score, we first assigned a weight to each risk factor in relation to each β parameter based on the multilevel model. Then, we summed the weights of each of the risk factors of a patient, higher scores indicating a greater likelihood of hospital admission. The predictive accuracy of the hospital admission risk score was assessed using the AUC [17] in both derivation and validation samples (external validation). Furthermore, we sought to validate the risk score by K-fold cross-validation [19, 20], which uses part of the available data of the derivation sample to fit the model, and a different part to test it (internal validation) [19].

Based on the hospital admission risk score, we categorized the score into different levels of risk. The optimal thresholds on the continuous risk scores were determined with the *catpredi* function of the R package *CatPredi* using the genetic algorithm [21]. The performance of the risk classification was studied by comparing the hospital admission rate between categories and using the multilevel analyses with generalized estimated equations and AUCs, in both derivation and validation samples.

All effects were considered significant at $p < 0.05$. All statistical analyses were performed using SAS for Windows, version 9.4 (SAS Institute, Carey, NC), and R[®] version 4.0.4.

Results

A total of 49,750 COVID-19 positive patients were included in the analysis. The whole sample was divided in a derivation and validation samples for the purpose of this study. The main characteristics of both samples are described in Table 1. The univariable analysis results for risk of hospital admission in the derivation sample are shown in Table 2. As it can be seen, multiple background characteristics, comorbidities, and basal patient treatments were related to hospital admission.

The multivariable multilevel analysis for the prediction of the risk of admission in the derivation sample is shown in Table 3. Male showed higher risk of hospital admission. Regarding age, its gradual effect continues to be maintained with greater ease of admission up to 89 years of age, decreasing from 90 years of age or older. It should be noted that the age groups composed of those under 49 years of age did not exceed the OR of 8.2. Regarding comorbidities, the results observed in the univariable analysis are retained, highlighting coagulopathies, cancer, diabetes with organ damage, and liver disease among them. The fact of having been admitted in the previous month to the COVID-19 diagnosis implied a risk of hospital admission of an OR 4.23 (95% CI 3.24–5.52). On relation to flu vaccination in the previous season, it is also a risk factor of hospital admission with an OR of 1.18 (95% CI 1.08–1.29). However, the pneumococcal vaccine is no longer predictive of risk for admission. Related to the treatments, it should be noted that the drugs that increase the risk were the baseline prescription of chronic systemic steroids, immunosuppressants, angiotensin-converting enzyme (ACE) inhibitors, and NSAIDs. The AUC for this model was 0.821 (95% CI 0.815–0.828) (Fig. 1).

Based on the multivariable model, we assigned a weight to each category of each significant predictor and created a risk score, ranging from 0 to 97. The AUCs (95% CI) of the risk score were 0.821 (0.815–0.828) and 0.828 (0.821–0.835) in the derivation and validation sample, respectively. Similar AUC were obtained by K-fold cross-validation, 0.820 (95% CI 0.814–0.827) (Table 3).

From the previous risk score, we developed five risk groups of hospital admission where patients risk of admission ranged from 2.26 to 54.58% from the minor to the most severe risk category in the derivation sample (Table 4), with

Table 1 Comparative analysis between the derivation and validation samples

Variables	Validation (<i>N</i> =19,900) <i>n</i> (%)	Derivation (<i>N</i> =29,850) <i>n</i> (%)	<i>p</i>
Hospital admissions	3178 (15.97)	4763 (15.96)	0.9681
Sex			
Male	8898 (44.71)	13,513 (45.27)	0.2220
Female	11,002 (55.29)	16,337 (54.73)	
Age (years), mean (SD)	47.73 (23.19)	47.85 (23.13)	0.5275
BMI, mean (SD) ^a	26.57 (6.57)	26.59 (6.37)	0.5044
Nursing home residents	1746 (8.82)	2571 (8.65)	0.5305
Hospital admission in the previous month	263 (1.32)	348 (1.17)	0.1222
Comorbidities			
Cardiovascular	1981 (9.95)	2878 (9.64)	0.2489
Cerebrovascular	1570 (7.89)	2367 (7.93)	0.8707
Hemiplegia and/or paraplegia	284 (1.43)	391 (1.31)	0.2681
Dementia	837 (4.21)	1283 (4.30)	0.6182
Respiratory disease	3256 (16.36)	4960 (16.62)	0.4537
Liver disease			
No	19,023 (95.59)	28,621 (95.88)	0.2172
Mild	781 (3.92)	1107 (3.71)	
Moderate/severe	96 (0.48)	122 (0.41)	
Diabetes			
No	18,080 (90.85)	27,096 (90.77)	0.1082
Without organ damage	1475 (7.41)	2298 (7.70)	
With organ damage	345 (1.73)	456 (1.53)	
Kidney disease	954 (4.79)	1379 (4.62)	0.3679
Cancer	1711 (8.60)	2682 (8.98)	0.1362
Rheumatic disease	405 (2.04)	623 (2.09)	0.6900
Peptic ulcer	453 (2.28)	678 (2.27)	0.9706
Inflammatory bowel disease	439 (2.21)	658 (2.20)	0.9901
Arterial hypertension	4776 (24.00)	7112 (23.83)	0.6553
Dyslipidemia	4564 (22.93)	6785 (22.73)	0.5946
Coagulopathies	104 (0.52)	174 (0.58)	0.3767
Gastrointestinal bleeding	216 (1.09)	340 (1.14)	0.5774
Asthma	2070 (10.40)	3171 (10.62)	0.4313
Flu vaccine	3505 (17.61)	5364 (17.97)	0.3084
Pneumococcal vaccine	2630 (13.22)	3990 (13.37)	0.6277
Baseline treatments			
Antidiabetic	1299 (6.53)	2011 (6.74)	0.3586
Cardiovascular	647 (3.25)	893 (2.99)	0.1014
Antihypertensive	136 (0.68)	278 (0.93)	0.0029
Diuretics	1349 (6.78)	1984 (6.65)	0.5630
Beta-blockers	1119 (5.62)	1658 (5.55)	0.7438
Calcium channel blockers	744 (3.74)	1162 (3.89)	0.3803
Renin-angiotensin	2873 (14.44)	4343 (14.55)	0.7277
Lipid lowering agent/statins	2523 (12.68)	3651 (12.23)	0.1383
NSAIDs	3316 (16.66)	5000 (16.75)	0.7987
Oral anticoagulants	887 (4.46)	1343 (4.50)	0.8250
Antiplatelets	1353 (6.80)	2014 (6.75)	0.8213
Heparin	92 (0.46)	118 (0.40)	0.2588
Bronchodilators	1608 (8.08)	2437 (8.16)	0.7377
Chronic Azithromycin	45 (0.23)	64 (0.21)	0.7841
Immunosuppressors	269 (1.35)	403 (1.35)	0.9873
Chronic systemic steroids	739 (3.71)	1093 (3.66)	0.7632

SD Standard deviation, *BMI* Body Mass Index

^aBMI data was only available for *n* = 16,237 patients (*n* = 6547 validation and *n* = 9690 derivation)

an AUC of 0.814 (95% CI 0.808–0.821). Those risk groups were validated in the validation sample, reaching an AUC of 0.820 (95% CI 0.812–0.827) with a risk of admission ranging from 2.09 to 54.31% from minor to severe categories (Table 4).

We also present different scenarios for the specificity and sensibility of our risk score for different cut-off points of the previous risk groups (Table 5) with a sensitivity of 85.7 and specificity of 61.75 (85.53 and 62.35, respectively, in the validation sample) for a risk score ≥ 26 points.

Additionally, other relevant variables as smoking habits and BMI have been considered, as well. Nevertheless, smoking habits was finally excluded from the previous final model, because it was only present in 39,193 records and was not considered reliable enough. The BMI variable presented a similar situation, since it was available just in 16,237 records though when included in our previous multivariable model was statistically significantly associated with hospital admission, with OR increasing from 1.19 ($25 \leq \text{BMI} < 30$) to 1.25 ($30 \leq \text{BMI} < 35$), 1.39 ($35 \leq \text{BMI} < 40$), and 2.36 ($35 \leq \text{BMI} < 40$) compared to $\text{BMI} < 25$.

Discussion

The present is a population-based study with a cohort of 49,750 consecutive patients in Basque Country with confirmed microbiological infection with SARS-CoV-2. In line with the other studies, we find that males have a higher risk of requiring hospitalization and suffer worse development of the disease [22]. As far as the age of the patients, this factor has a gradual impact concerning the risk of requiring hospitalization. Several studies [22] indicate how those over 50 have a higher risk of needing intensive care treatment or death, ranging from 60 [23] or even 68 in Italy [24]. Regarding comorbidities in our cohort, the coagulopathies are the pathologies that offer the greatest risk of hospitalization and the main factor in the final outcome [25]. Numerous studies present cardiovascular illness producing a greater risk of a bad prognosis [4], as well as HTN or diabetes [6], which also figure in or study as one of the top five factors.

Cancer patients have a higher risk of suffering complications due to COVID-19, although we have to consider that most Cancer patients are over 65, suffer one or more comorbidities, and their treatment frequently causes immunosuppression [26]. This bad prognosis in line with the results from different studies with cohorts in USA [27] and the UK [22], and one of the five comorbidities of greater risk when calculating hospitalization.

Regarding the BMI, the results obtained highlight the increased gradient with weight. In a study based on a cohort of 306 patients [28], the main finding concerned the high frequency of obese patients requiring Intensive Care treatment,

with a marked difference with patients treated in previous years; 47.5% registered BMI of $\geq 30 \text{ kg/m}^2$; 13.7% with an BMI 35–39.9 kg/m^2 and 14.5% recording an BMI $\geq 40 \text{ kg/m}^2$.

As far as the use of smoking habits, a meta-analysis conducted by Carlos A. Jiménez-Ruiz [29], was not able to prove that smokers had a higher chance of becoming infected but did show that a smoker or ex-smoker had twice as high a risk of undergoing a worsening condition or adverse outcome (OR 1.96, IC del 95%, 1.36–2.83). Equally smoking habits use whether a current or ex-smoker is shown to be a risk factor for those more critical, requiring orotracheal intubation, Intensive Care treatment or death (OR 1.79, IC del 95%, 1.19–2.70). However, the data obtained by us are not statistically significant nor reliable enough.

With respect to the flu vaccine in the Basque Country, it is targeted primarily at those aged 65 or over, patients with chronic illness or pregnant women and it is recommended to those who work in the Health Service. [30]. This distribution of the vaccine may help to explain the link between those who have been vaccinated and those under greater risk of hospital treatment for COVID-19. Being therefore a statistical significance association, but not a casual one, mainly due to the distribution of and degree of compliance with the vaccine in the older population and those most at risk in our environment.

Regarding treatment, higher risk has been observed in patients who use ACE inhibitors and NSAIDs. This could be due to a confounding factor with underlying medical comorbidities, increasing the risk and linked to higher rates of hospitalization and of death by COVID, which are treated more frequently with these medicines [5]. Nevertheless, based on the evidence available to date, there are no compendiums which suggest the need to modify the use of ACE inhibitors, and therefore, they should be introduced or maintained in accordance with the current guides, independently of SARS-CoV-2. This conclusion, however, should be constantly reviewed.

In so far as the limitations of the study, it should be pointed out that the population studied included only patients in the Basque Country diagnosed positive with COVID-19 in the general population. However, we hope that this model is more widely used in other geographic regions. Second, the data collected through our system and used for extraction and analysis provided a solid and rapid analysis of a large cohort. Nevertheless, due to the retrospective design, not all the variables to be considered were obtained from all the patients, or in a reliable way, as was the case with BMI or smoking habits use, and therefore, these variables cannot be evaluated appropriately. Moreover, in some of the diagnosed cases during the first wave, the health crisis meant that the information gathered was incomplete and lacking in the necessary detail for a more

Table 2 Univariate analysis for prediction of hospital admissions in derivation sample ($N=29,850$)

Variables	OR (95% CI)	<i>p</i>
Sex		
Male (vs female)	1.59 (1.49–1.69)	< 0.0001
Age (years) categorized (vs < 20 years)		
20–29	1.98 (1.39–2.83)	0.0002
30–39	4.54 (3.29–6.27)	< 0.0001
40–49	8.01 (5.88–10.91)	< 0.0001
50–59	12.55 (9.25–17.02)	< 0.0001
60–69	27.89 (20.57–37.80)	< 0.0001
70–79	55.26 (40.69–75.03)	< 0.0001
80–89	56.50 (41.62–76.70)	< 0.0001
≥ 90	35.39 (25.79–48.58)	< 0.0001
Categorized BMI ^a (vs BMI < 25)		
25 ≤ BMI < 30	2.56 (2.29–2.86)	< 0.0001
30 ≤ BMI < 35	2.52 (2.22–2.88)	< 0.0001
35 ≤ BMI < 40	2.04 (1.68–2.47)	< 0.0001
BMI ≥ 40	2.61 (2.00–3.41)	< 0.0001
Nursing home residents	1.85 (1.68–2.03)	< 0.0001
Hospital admission in the previous month	10.74 (8.58–13.43)	< 0.0001
Comorbidities		
Cardiovascular	5.93 (5.57–6.44)	< 0.0001
Cerebrovascular	3.77 (3.44–4.12)	< 0.0001
Hemiplegia and/or paraplegia	2.80 (2.26–3.46)	< 0.0001
Dementia	2.99 (2.66–3.37)	< 0.0001
Respiratory disease	2.14 (1.99–2.30)	< 0.0001
Liver disease (vs No)		
Mild	3.68 (3.25–4.17)	< 0.0001
Moderate/severe	5.90 (4.13–8.43)	< 0.0001
Diabetes (vs No)		
Without organ damage	4.11 (3.75–4.50)	< 0.0001
With organ damage	8.62 (7.14–10.41)	< 0.0001
Kidney disease	6.16 (5.51–6.88)	< 0.0001
Cancer	2.30 (2.10–2.52)	< 0.0001
Rheumatic disease	3.91 (3.32–4.60)	< 0.0001
Peptic ulcer	3.13 (2.67–3.67)	< 0.0001
Arterial hypertension	4.98 (4.67–5.32)	< 0.0001
Dyslipidemia	3.70 (3.47–3.95)	< 0.0001
Coagulopathies	4.14 (3.06–5.59)	< 0.0001
Gastrointestinal bleeding	4.11 (3.31–5.10)	< 0.0001
Flu vaccine	4.25 (3.97–4.55)	< 0.0001
Pneumococcal vaccine	5.02 (4.67–5.41)	< 0.0001
Baseline treatments		
Antidiabetic	4.06 (3.69–4.46)	< 0.0001
Cardiovascular	4.05 (3.54–4.65)	< 0.0001
Antihypertensive	5.40 (4.26–6.84)	< 0.0001
Diuretics	4.98 (4.53–5.48)	< 0.0001
Beta-blockers	4.11 (3.70–4.55)	< 0.0001
Calcium channel blockers	4.40 (3.90–4.96)	< 0.0001
Renin–angiotensin	4.07 (3.79–3.38)	< 0.0001
Lipid lowering agent/statins	4.37 (4.05–4.71)	< 0.0001
NSAIDs	0.88 (0.81–0.96)	0.0027
Oral anticoagulants	5.20 (4.64–5.81)	< 0.0001

Table 2 (continued)

Variables	OR (95% CI)	<i>p</i>
Antiplatelets	4.06 (3.69–4.46)	< 0.0001
Heparin	5.51 (3.83–7.91)	< 0.0001
Bronchodilators	2.33 (2.12–2.56)	< 0.0001
Chronic azithromycin	6.01 (3.67–9.82)	< 0.0001
Immunosuppressors	3.26 (2.66–4.00)	< 0.0001
Chronic systemic steroids	3.51 (3.09–3.98)	< 0.0001

OR Odds Ratio, CI Confidence Interval, BMI Body Mass Index

^aBMI data were only available for *n* = 9690 patients

in-depth analysis of specific aspects. Among the strengths of the study, it is to be noted that the multicenter population design developed in a public health system with universal access for the entire population with a network of Primary Care, Hospital Care and Emergencies fully connected with both a clinical history and universal laboratory network. Therefore, the size and the wide geographic coverage offer a statistical power to confirm the hypothesis and represent strong points of the study, together with the potential to apply it at the first line of medical attention. In addition, from a practical point of view, the proposed model of prediction of hospitalization would allow the two lower risk categories with a high degree of negative prediction value (95.79%) to be sent home and maybe even category 3. Those patients in categories 3–4 could be dealt with bearing in mind other data and evaluate deeply those in category 5 who may require hospitalization. To date, the criteria used to admit patients to a hospital were the

continuously updated by the WHO [31] and the guidelines from the Spanish Ministry of Health [32], which specifies the criteria to be followed for the evaluation of the severity of patients in the services of hospital emergencies for subsequent admission or home monitoring by primary care teams, with age, comorbidities, fever, and respiratory failure; the factors to be taken into account mainly. Nevertheless, in the first few weeks of the pandemic, patients admitted were those more severe which may condition our conclusion, though was limited in time.

As conclusion, we propose a risk scale of hospitalization in COVID-19 patients, which indirectly concerns to people with a higher risk of poor evolution, to facilitate clinical decision-making by both primary care professionals and emergency services. Our scale can be easily and quickly completed either in primary attention units or in hospital or out-of-hospital emergencies, which would help in clinical decision-making.

Table 3 Multivariable analysis for prediction of hospitalization admissions in the derivation sample using multilevel analysis

Variables	OR (95% CI)	<i>p</i>	β Coefficient	Weight
Sex				
Male	1.77 (1.65–1.90)	<0.0001	0.5708	5.5
Female	Ref	–	–	0
Age (years) categorized				
≤ 19	Ref	–	–	0
20–29	2.19 (1.53–3.13)	<0.0001	0.7821	7.5
30–39	4.82 (3.47–6.70)	<0.0001	1.5734	15
40–49	8.20 (5.99–11.23)	<0.0001	2.1040	20
50–59	11.42 (8.36–15.60)	<0.0001	2.4358	23.5
60–69	19.47 (14.21–26.67)	<0.0001	2.9687	28.5
70–79	28.90 (20.93–39.92)	<0.0001	3.3640	32.5
80–89	26.59 (19.15–36.92)	<0.0001	3.2806	31.5
≥ 90	19.11 (13.58–26.90)	<0.0001	2.9502	28.5
Hospital admissions 1 month	4.23 (3.24–5.52)	<0.0001	1.4415	14
Comorbidities				
Cardiovascular	1.46 (1.32–1.62)	<0.0001	0.3801	3.5
Cerebrovascular	1.23 (1.10–1.37)	0.0002	0.2071	2
Respiratory	1.45 (1.33–1.59)	<0.0001	0.3730	3.5
Liver disease	1.53 (1.33–1.75)	<0.0001	0.4229	4
Diabetes				
No	Ref	–	–	0
Without organ damage	1.28 (1.15–1.42)	<0.0001	0.2434	2.5
With organ damage	1.54 (1.24–1.92)	<0.0001	0.4345	4
Kidney disease	1.45 (1.26–1.65)	<0.0001	0.3692	3.5
Cancer	1.54 (1.38–1.72)	<0.0001	0.4342	4
Rheumatic disease	1.34 (1.10–1.65)	0.0043	0.2954	3
Arterial hypertension	1.16 (1.05–1.29)	0.0042	0.1514	1.5
Dyslipidemia	1.15 (1.06–1.24)	0.0008	0.1372	1.5
Coagulopathies	1.82 (1.26–2.62)	0.0014	0.5971	5.5
Flu vaccine	1.18 (1.08–1.29)	0.0003	0.1656	1.5
Baseline treatments				
Renin–angiotensin	1.13 (1.02–1.25)	0.0172	0.1243	1
NSAIDs	1.11 (1.005–1.22)	0.0399	0.1036	1
Immunosuppressors	1.31 (1.01–1.71)	0.0427	0.2731	2.5
Chronic systemic steroids	1.36 (1.15–1.61)	0.0003	0.3096	3
Valid N	29,557			
AUC (IC 95%)	0.821 (0.815–0.828)			
Hosmer–Lemeshow, <i>p</i>	<0.0001			
Risk score, range				0–97
AUC (IC 95%)				
Derivation sample	0.821 (0.815–0.828)			
K-fold cross-validation	0.820 (0.814–0.827)			
Validation sample	0.828 (0.821–0.835)			

OR Odds Ratio, CI Confidence Interval, AUC Area under the ROC curve

Ref. Reference group

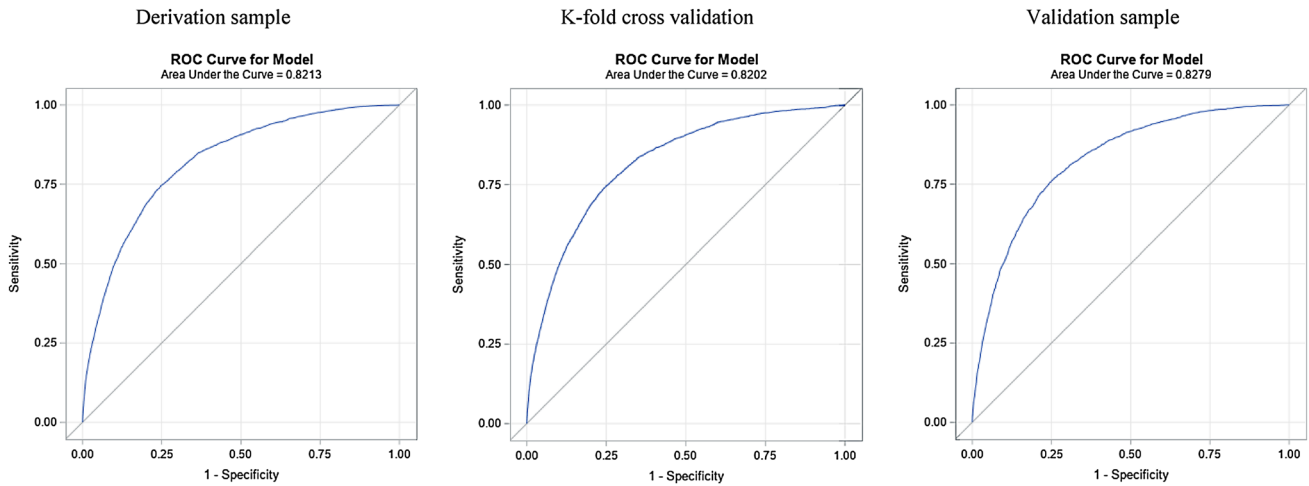


Fig. 1 ROC curves for the risk score in both derivation and validations samples, and k-fold cross-validation

Table 4 Risk groups of hospitalization in the derivation and validation samples

	Derivation sample				Validation sample			
	<i>N</i>	Hospitalization, <i>n</i> (%)	OR (95% CI)	<i>p</i>	<i>N</i>	Hospitalization, <i>n</i> (%)	OR (95% CI)	<i>p</i>
5 risk groups								
Score < 17	8581	194 (2.26)	Ref	–	5788	121 (2.09)	Ref	–
17 ≤ Score < 26	7588	487 (6.42)	2.94 (2.48–3.49)	<0.0001	5095	339 (6.65)	3.33 (2.70–4.12)	<0.0001
26 ≤ Score < 35	6309	1043 (16.53)	8.50 (7.25–9.96)	<0.0001	4194	675 (16.09)	8.89 (7.28–10.86)	<0.0001
35 ≤ Score < 43	4277	1352 (31.61)	19.90 (17–23.30)	<0.0001	2652	866 (32.65)	22.49 (18.43–27.45)	<0.0001
Score ≥ 43	3091	1687 (54.58)	51.87 (44.19–60.89)	<0.0001	2167	1177 (54.31)	55.07 (45.06–67.31)	<0.0001
AUC (95% CI)			0.814 (0.808–0.821)				0.820 (0.812–0.827)	

OR Odds ratio, CI Confidence interval, AUC Area under the ROC curve, Ref. Reference group

Table 5 Sensitivity, specificity, and positive and negative predictive values according to different cut-off points in both derivation and validation samples

	Derivation sample				Validation sample			
	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Based on 5 risk groups								
Score ≥ 17	95.93	33.44	21.49	97.74	96.19	33.90	21.67	97.91
Score ≥ 26	85.70	61.75	29.85	95.79	85.53	62.35	30.16	95.77
Score ≥ 35	63.80	82.74	41.25	92.33	64.29	83.40	42.39	92.47
Score ≥ 43	35.42	94.40	54.58	88.50	37.04	94.08	54.31	88.71

Acknowledgements We are grateful for the support of the Basque Health Service-Osakidetza and the Department of Health of the Basque Government. We also gratefully acknowledge the patients who participated in the study. Susana García-Gutiérrez, Miren Orive, Nerea Gonzalez, Iratxe Lafuente, Ane Anton, Ane Villanueva, Josune Martin,

Cristina Muñoz, Maria Jose Legarreta, Urko Aguirre and Jose María Quintana (Research Unit) Hospital de Galdakao-Usansolo; Pedro Pablo España Yandiola, Ane Uranga, Mikel Egurrola, Amaia Aramburu, Amaia Artaraz, Leire Chasco, Olaia Bronte, Patricia García, Ana Jodar, Virginia Fernandez, and Cristobal Esteban (Respiratory

Service) Hospital de Galdakao-Usansolo, Naia Mas(ICU) Hospital de Galdakao-Usansolo, Esther Pulido (Emergency Service) Hospital de Galdakao-Usansolo, Itxaso Bengoetxea (Hospital at Home Service), Galdakao-Barrualde Health Organization; Antonio Escobar Martínez, Amaia Bilbao and Iñigo Gorostiza, (Research Unit) Hospital Universitario Basurto; Iñaki Arriaga, (Respiratory Service) Hospital Universitario Basurto; José Joaquín Portu Zapiarain (Internal Medicine) and Naiara Parraza (Research Unit), Instituto Bioaraba; Milagros Iriberrri, Rafael Zalacain , Luis Alberto Ruiz and Leyre Serrano (Respiratory Service) Hospital Universitario de Cruces; Adriana Couto and Oier Ateka (Internal Medicine Service) Hospital Universitario Donostia; Arantza Cano, Hospital Santa Marina (Respiratory Service); Maria Olatz Ibarra, Hospital de Urduliz (Pharmacy Service); Eduardo Millan, Mayte Bacigalupe, Jon Letona and Andoni Arcelay, Central Services Osakidetza; Iñaki Berraondo, Department of Health, Basque Government.

Author contributions JP-J: oversaw the conception and methodological design of the study as well as development, analysis, data interpretation, and article writing, and approval of the final version of the article. AB-G: oversaw analysis and data interpretation and contributed to writing the article and approval of the final version of the article. VT-G: took part in design and data interpretation and contributed to writing the article and approval of the final version of the article. IG-G: took part in data interpretation and contributed to writing the article and approval of the final version of the article. SG-G: oversaw analysis and data interpretation and contributed to writing the article and approval of the final version of the article. AM-M: took part in design and data interpretation and contributed to writing the article and approval of the final version of the article. JS-G: took part in design and contributed to writing the article and approval of the final version of the article. SA-G: took part in design and contributed to writing the article and approval of the final version of the article. JA-G: took part in data extraction and interpretation and contributed to writing the article and approval of the final version of the article. JC-P: took part in design and contributed to writing the article and approval of the final version of the article. IZ-M: took part in design and data interpretation and contributed to writing the article and approval of the final version of the article. The COVID-19-Osakidetza Working group approval of the final version of the article. Jose Maria Quintana-Lopez oversaw analysis, data interpretation, and article writing and approval of the final version of the article.

Funding This work was supported in part by grants from the Instituto de Salud Carlos III and the European Regional Development Fund COVID20/00459; the health outcomes group from Galdakao-Barrualde Health Organization; the Kronikune 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.

Data availability Data are available on request from the authors.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

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


References

1. World Health Organization (WHO) (2020) WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 Feb 2020 [Internet]. Available from: <https://www.who.int/director-general/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020>
2. Velavan TP, Meyer CG (2020) The COVID-19 epidemic. *Trop Med Int Health* 25(3):278–280
3. España. Ministerio de la Presidencia. Real Decreto 463/2020, de 14 de marzo, por el que se declara el estado de alarma para la gestión de la situación de crisis sanitaria ocasionada por el COVID-19. *Boletín Oficial del Estado* [Internet]. 2020 [cited 2022 Feb 5];67(I):25390–400. Available from: <https://www.boe.es/buscar/doc.php?id=BOE-A-2020-3692>
4. Caramelo F, Ferreira N, Oliveiros B (2020) ESTimation of risk factors for COVID-19 mortality—preliminary results. *MedRxiv*. <https://doi.org/10.1101/2020.02.24.20027268>
5. Jehi L, Ji X, Milinovich A, Erzurum S, Merlino A, Gordon S et al (2020) Development and validation of a model for individualized prediction of hospitalization risk in 4,536 patients with COVID-19. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0237419> (**Orueta JF, Editor**)
6. Fernández García L, Puentes Gutiérrez AB, García BM (2020) Relationship between obesity, diabetes and ICU admission in COVID-19 patients. *Med Clí (English Edition)* 155(7):314–315
7. World Health Organization (WHO) Coronavirus disease (COVID-19): Tobacco [Internet]. [Cited 26 Jan 2021]. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19-tobacco>
8. Zhao Z, Chen A, Hou W, Graham JM, Li H, Richman PS et al (2020) Prediction model and risk scores of ICU admission and mortality in COVID-19. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0236618>
9. Goff PH, Hayashi T, Martínez-Gil L et al (2015) Synthetic toll-like receptor 4 (TLR4) and TLR7 ligands as influenza virus vaccine adjuvants induce rapid, sustained, and broadly protective responses. *J Virol* 89(6):3221–3235. <https://doi.org/10.1128/JVI.03337-14>
10. Chih-cheng L, Tzu-Ping S, Wen-Chien K, Hung-Jen T, Po-Ren H (2020) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *Int J Antimicrob Agents* 55(3):105924
11. Deng Y, Liu W, Liu K, Fang Y-Y, Shang J, Zhou L et al (2020) Clinical characteristics of fatal and recovered cases of coronavirus disease 2019 (COVID-19) in Wuhan, China. *Chin Med J*. <https://doi.org/10.1097/CM9.0000000000000824>
12. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395(10223):497–506
13. Charlson ME, Sax FL, MacKenzie CR, Fields SD, Braham RL, Douglas RG (1986) Assessing illness severity: does clinical judgment work? *J Chronic Dis* 39(6):439–452
14. WHO/WHO Collaborating Centre for Drug Statistics and Methodology [Internet] (2016) WHO. World Health Organization; [Cited 10 Apr 2021]. Available from: <http://www.who.int/medicines/regulation/medicines-safety/about/collab-centres-norwegian/en/>
15. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC et al (2005) Coding algorithms for defining comorbidities

- in ICD-9-CM and ICD-10 administrative data. *Med Care* 43(11):1130–1139
16. España PP, Bilbao A, García-Gutiérrez S, Lafuente I, Anton-Ladislao A, Villanueva A et al (2021) Predictors of mortality of COVID-19 in the general population and nursing homes. *Intern Emerg Med*. <https://doi.org/10.1007/s11739-020-02594-8>
 17. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982 Apr;143(1):29–36.
 18. Hosmer DW, Lemeshow S. *Applied logistic regression*. 1989. 1–37
 19. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. The Business of Giving. Springer, New York; 2009
 20. Steyerberg EW. *Clinical prediction models*. Springer, New York; 2010
 21. Barrio I, Arostegui I, Rodríguez-Álvarez M-X, Quintana J-M (2017) A new approach to categorising continuous variables in prediction models: proposal and validation. *Stat Methods Med Res* 26(6):2586–2602
 22. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L et al (2020) Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO clinical characterisation protocol: prospective observational cohort study. *BMJ*. <https://doi.org/10.1136/bmj.m1985>
 23. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H et al (2020) Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med* 8(5):475–481
 24. Colombi D, Bodini FC, Petrini M, Maffi G, Morelli N, Milanese G et al (2020) Well-aerated lung on admitting chest CT to predict adverse outcome in COVID-19 pneumonia. *Radiology* 296(2):E86–96. <https://doi.org/10.1148/radiol.2020201433>
 25. Sáenz Morales OA, Rubio AM, Yomayusa N, Gamba N, Garay Fernández M (2020) Coagulopatía en la infección por el virus SARS-CoV-2 (COVID-19): de los mecanismos fisiopatológicos al diagnóstico y tratamiento. *Acta Colombiana de Cuidado Intensivo*
 26. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* [Internet]. 2020 Mar 28 [cited 2020 May 9];395(10229):1054–62. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0140673620305663>
 27. Singh K, Valley TS, Tang S, Li BY, Kamran F, Sjoding MW et al (2020) Validating a widely implemented deterioration index model among hospitalized COVID-19 patients. medRxiv. <https://doi.org/10.1101/2020.04.24.20079012>
 28. Simonnet A, Chetboun M, Poissy J, Raverdy V, Noulette J, Duhamel A et al (2020) High prevalence of obesity in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. *Obesity* 28(7):1195–1199
 29. Jiménez-Ruiz CA, López-Padilla D, Alonso-Arroyo A, Alexandre-Benavent R, Solano-Reina S, de Granda-Orive JI (2021) COVID-19 y tabaquismo: revisión sistemática y metaanálisis de la evidencia. *Arch Bronconeumol* 57:21–34
 30. Osasun Saila—Departamento de Salud. Campaña de vacunación antigripal 2019 [Internet]. Available from: <https://www.euskadi.eus/gobierno-vasco/-/noticia/2019/osakidetza-presenta-la-campa-na-de-vacunacion-antigripal-2019/>
 31. WHO (2020) Manejo clínico de la infección respiratoria aguda grave (IRAG) en caso de sospecha de COVID-19. Mar 13 [Cited 2 Sep 2021]; Available from: <https://apps.who.int/iris/bitstream/handle/10665/331660/WHO-2019-nCoV-clinical-2020.4-spa.pdf?sequence=1&isAllowed=y>
 32. Ministerio de Sanidad—Gobierno de España (2020) Manejo en urgencias del COVID-19. Jun 26 [Cited 2 Sep 2021]; Available from: https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/Manejo_urgencias_pacientes_con_COVID-19.pdf

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Janire Portuondo-Jimenez^{1,2,3} · Amaia Bilbao-González^{4,5,6} · Verónica Tíscar-González^{3,4,8} · Ignacio Garitano-Gutiérrez^{7,9} · Susana García-Gutiérrez^{5,6,10} · Almudena Martínez-Mejuto^{3,11} · Jaione Santiago-Garin^{7,8} · Silvia Arribas-García^{7,8} · Julia García-Asensio¹ · Johnny Chart-Pascual¹³ · Iñaki Zorrilla-Martínez^{12,13,14}  · Jose Maria Quintana-Lopez^{5,6,10} · the COVID-19-Osakidetza Working group

¹ Basque Government Department of Health, Office of Healthcare Planning, Organisation and Evaluation, Basque Country, Spain

² Osakidetza Basque Health Service, Sub-Directorate for Primary Care Coordination, Vitoria-Gasteiz, Spain

³ Biocruces Bizkaia Health Research Institute, Clinical Nursing and Community Health Group, Barakaldo, Spain

⁴ Osakidetza Basque Health Service, Basurto University Hospital, Research Unit, Bilbao, Spain

⁵ Health Service Research Network on Chronic Diseases (REDISSEC), Bilbao, Spain

⁶ Kronikgune Institute for Health Services Research, Barakaldo, Spain

⁷ Bioaraba Health Research Institute, Clinical Nursing and Community Health Group, Vitoria-Gasteiz, Spain

⁸ Osakidetza Basque Health Service, University School of Nursing, University of the Basque Country, Vitoria-Gasteiz, Spain

⁹ Osakidetza Basque Health Service, Araba University Hospital, Vitoria-Gasteiz, Spain

¹⁰ Osakidetza Basque Health Service, Galdakao-Usansolo University Hospital, Research Unit, Galdakao, Spain

¹¹ Osakidetza Basque Health Service, Cruces University Hospital, Barakaldo, Spain

¹² Bioaraba Health Research Institute, Clinical trials, Vitoria-Gasteiz, Spain

¹³ Osakidetza Basque Health Service, Araba IHO, Psychiatry Service, Vitoria-Gasteiz, Spain

¹⁴ Faculty of Medicine, University of the Basque Country, Vitoria-Gasteiz, Spain