Background: Coronavirus disease 2019 (COVID-19) is a highly variable condition. Validated tools to assist in the early detection of patients at high risk of mortality can help guide medical decisions.

Objective: We sought to validate externally, as well as in patients from the second pandemic wave in Europe, our previously developed mortality prediction model for hospitalized COVID-19 patients.

Methods: Three validation cohorts were generated: 2 external with 185 and 730 patients from the first wave and 1 internal with 119 patients from the second wave. The probability of death was calculated for all subjects using our prediction model, which includes peripheral blood oxygen saturation/fraction of inspired oxygen ratio, neutrophil-to-lymphocyte ratio, lactate dehydrogenase, IL-6, and age. Discrimination and calibration were evaluated in the validation cohorts. The prediction model was updated by reestimating individual risk factor effects in the overall cohort (N = 1477).

Results: The mortality prediction model showed good performance in the external validation cohorts 1 and 2, and in the second wave validation cohort 3 (area under the receiver-operating characteristic curve, 0.94, 0.86, and 0.86, respectively), with excellent calibration (calibration slope, 0.86, 0.94, and 0.79; intercept, 0.05, 0.03, and 0.10, respectively). The updated model accurately predicted mortality in the overall cohort (area under the receiver-operating characteristic curve, 0.91), which included patients from both the first and second COVID-19 waves. The updated model was also useful to predict fatal outcome in patients without respiratory distress at the time of evaluation.

Conclusions: This is the first COVID-19 mortality prediction model validated in patients from the first and second pandemic waves. The COR+12 online calculator is freely available to facilitate its implementation (https://utrero-rico.shinyapps.io/COR12_Score/). (J Allergy Clin Immunol 2021;147:1652-61.)

Key words: COVID-19, IL-6, mortality risk, predictive model, second wave, external validation
Hospital Universitario 12 de Octubre (Madrid) have already been published.12

internal cohort from the second wave. Details on the development cohort from
external cohorts from the first wave of COVID-19 cases, and in a prospective
were not meeting the inclusion criteria or having a hematological malignancy
or death) within 40 days from hospital admission (40 days was the maximum
on the 5 variables included in the model, and (c) having an outcome (discharge
PCR, (b) having data at admission, or within the first 4 days of hospitalization,
being hospitalized with confirmed diagnosis of COVID-19 by real-time RT-
The institutional Clinical Research Ethics Committee approved the study pro-

Study design and population

The validation of this prediction model was performed in 2 retrospective
external cohorts from the first wave of COVID-19 cases, and in a prospective
internal cohort from the second wave. Details on the development cohort from
Hospital Universitario 12 de Octubre (Madrid) have already been published.12

The institutional Clinical Research Ethics Committee approved the study proto-
col (reference no. 20/167). The criteria for inclusion in the study were (a)
being hospitalized with confirmed diagnosis of COVID-19 by real-time RT-
PCR, (b) having data at admission, or within the first 4 days of hospitalization,
the 5 variables included in the model, and (c) having an outcome (discharge or
death) within 40 days from hospital admission (40 days was the maximum
time allowed for an outcome in the development cohort). Exclusion criteria
were not meeting the inclusion criteria or having a hematological malignancy
associated with increased lymphocyte count. A flowchart depicts the inclusion
of patients in the different cohorts (Fig 1).

The external validation cohort 1 was composed of 188 patients hospitalized
between March 10 and March 30, 2020, at the Hospital Universitario Ramón y
Cajal (Madrid) with IL-6 measurement. Three patients were not included
because of missing LDH measurements.

The external validation cohort 2 included 898 patients who attended the
emergency room (ER) at either Hospital Universitario Infanta Elena or
Hospital Universitario Fundación Jiménez Díaz (Madrid) from March 13 to
June 17, 2020, had confirmed COVID-19, and had an IL-6 measurement.
These 2 hospitals share a centralized laboratory and electronic clinical records
system and have therefore been considered as 1 cohort. One hundred
sixty-eight patients who did not meet the inclusion criteria were dismissed,
and the remaining 730 patients were included in the final analysis.

The model was further validated in a prospective internal cohort with
second wave COVID-19 patients (validation cohort 3). One hundred thirty-six
patients who attended the ER from August 24 to October 26 at the Hospital
Universitario 12 de Octubre (Madrid), who had confirmed COVID-19 and an
IL-6 measurement, were included in the second wave validation cohort 3.
Seventeen patients had missing values, and 119 patients were included in the
final analysis.

Data collection

Clinical electronic medical records were reviewed by researchers, and data
were manually collected at the Hospital 12 de Octubre and Hospital Ramón y
Cajal (validation cohorts 1 and 3). All data were revised by at least 2
independent researchers. At Hospital Infanta Elena and Hospital Fundación
Jiménez Díaz (validation cohort 2), clinical data were extracted by means of
big data/artificial intelligence processes from individual electronic medical
records and, then, reviewed and refined by 4 independent researchers.
Recorded data included demographic information, laboratory findings, length
of hospital stay, intensive care unit (ICU) admission, and outcome. SpO2/FiO2
was used to assess respiratory function. SpO2/FiO2 shows a good correlation
with the partial pressure of arterial oxygen (PaO2/FiO2) ratio (SpO2/FiO2 =
64 + 0.84 × PaO2/FiO2),13 and it was available for all the patients. Acute res-
piratory distress syndrome (ARDS) was classified according to Berlin criteria:
SpO2/FiO2 more than 315, no ARDS; 315 to 235, mild ARDS; 148 to 235,
moderate ARDS; less than 148, severe ARDS.14

Laboratory measurements

IL-6 was measured with the BD Cytometric Bead Array human IL-6 flex
set (BD Biosciences, San Jose, Calif) using a BD Canto II flow cytometer at
the Hospital 12 de Octubre and Hospital Ramón y Cajal. Results were
analyzed with FCAP Array software v3.0 (BD Biosciences). At Hospital
Fundación Jimenez Díaz, IL-6 was measured by Roche chemiluminescent
immunoassay. Only IL-6 measurements before tocilizumab therapy were
included in the study. Other laboratory parameters such as LDH, neutrophils,
and lymphocytes were measured as part of standard of care. All laboratory
measurements were taken on the same day.

The variables included in the prediction model (SpO2/FiO2, N/L ratio,
LDH, IL-6, and age) were taken at hospital admission or in the first 4 days
of the hospitalization. Most measurements were taken at the ER in the external
validation cohort 2 and the second wave validation cohort 3, whereas measure-
ments were taken with a median of 2 days from hospital admission in the
external validation cohort 1 (Table I).

Statistical analysis

Continuous numerical data were represented as median and interquartile
range and compared using the Mann-Whitney U test, or Kruskal-Wallis test
when relevant. Categorical variables were represented as N and percentage,
and compared using the chi-square test. The Fisher exact test was used
when appropriate. The potential of each variable in the model to be used indi-
vidually as a biomarker was evaluated using AUC analysis.

Transparent reporting of a multivariable prediction model for individual
prognosis or diagnosis (TRIPOD) guidelines for validation of multivariable
prediction models were followed.15 The logistic regression model in the dev-

0.0076 × (SpO2/FiO2) + 0.0547 × (N/L ratio) + 0.0046 × (LDH + 0.0043 × (IL-6 - 0.0682 × age))). This model was applied to the 3 validation
cohorts. The validity of the prediction model was assessed by evaluating
discrimination and calibration.16-18 Discrimination describes the ability
of the model to distinguish a patient who will survive from a patient who will

METHODS

Abbreviations used

ARDS: Acute respiratory distress syndrome
AUC: Area under the receiver-operating characteristic curve
COVID-19: Coronavirus disease 2019
ER: Emergency room
ICU: Intensive care unit
LDH: Lactate dehydrogenase
N/L: Neutrophil-to-lymphocyte
SpO2/FiO2: Peripheral blood oxygen saturation/fraction of inspired
oxygen ratio

VOLUME 147, NUMBER 5
J ALLERGY CLIN IMMUNOL
UTRERO-RICO ET AL 1653

after viral clearance. This hyperinflammatory dysregulation,
combined with other infection-derived complications such as
thrombosis,11 is responsible for the disease severity. However,
there are still many uncertainties about the pathophysiology of
COVID-19. Until we have a better understanding of this disease,
the availability of tools that allow risk stratification of infected pa-
cients can be useful for optimizing therapeutic management and
improving patient prognosis.

During the first wave of COVID-19 cases in Europe, we
developed a prediction model that estimates the probability of
death in hospitalized patients with COVID-19, based on 5
parameters taken at, or soon after, hospital admission: peripheral
blood oxygen saturation/fraction of inspired oxygen (SpO2/FiO2)
ratio, neutrophil-to-lymphocyte (N/L) ratio, lactate dehydroge-
nase (LDH), IL-6, and age.12 This model was rigorously devel-
oped by logistic regression, showing high accuracy in the
prediction of patient outcome (area under the receiver-operating
characteristic curve [AUC], 0.94), and it was translated into the
freely available COR + 12 online calculator (https://utrero-rico.
shinyapps.io/COR12_Score/). We have now externally validated
this prediction model in 2 independent cohorts from 2 different
tertiary hospitals (validation cohorts 1 and 2). In addition, we
have internally validated the model in a prospective patient cohort
recruited during the second wave (validation cohort 3). Finally,
the model has been updated with the overall sum of patients
from the development and validation cohorts.

The model was rigorously developed by logistic regression, showing high accuracy in the prediction of patient outcome (area under the receiver-operating characteristic curve [AUC], 0.94), and it was translated into the freely available COR + 12 online calculator (https://utrero-rico.shinyapps.io/COR12_Score/). We have now externally validated this prediction model in 2 independent cohorts from 2 different tertiary hospitals (validation cohorts 1 and 2). In addition, we have internally validated the model in a prospective patient cohort recruited during the second wave (validation cohort 3). Finally, the model has been updated with the overall sum of patients from the development and validation cohorts.
Discrimination of the model was evaluated by AUC analysis in the validation cohorts. The improved discrimination of the model, in comparison with each individual parameter, was evaluated using De Long test. Calibration of a model describes the agreement between the predicted and the observed mortality. Calibration was assessed (a) by the intercept alpha ($\alpha$), which compares the mean of all predicted risks and the mean observed risk, and in a perfect prediction would be 0, and (b) by the slope beta ($\beta$), which in good predictions is close to 1. Calibration results were represented as calibration plots and a circular barplot. The C-statistic does not appear in the text because for binary outcomes it corresponds to the AUC of the model.

The prediction model was updated with the compiled data from all the cohorts, by reestimating the intercept and effect of each individual risk factor. The updated model was probability of death = \frac{1}{1 + \exp (8.3777 + 0.0071 \times (\text{SpO}_2/\text{FiO}_2) - 0.0326 \times (\text{N/L ratio}) - 0.0046 \times \text{LDH} - 0.0027 \times \text{IL-6} - 0.0852 \times \text{age})). K-fold cross-validation ($k = 10$) for a generalized linear model was used to validate the updated model. The updated model excluding IL-6 was probability of death = \frac{1}{1 + \exp (7.7260 + 0.0070 \times (\text{SpO}_2/\text{FiO}_2) - 0.0364 \times (\text{N/L ratio}) - 0.0044 \times \text{LDH} - 0.0798 \times \text{age})).

Youden index was used for model cutoff selection, and with this cutoff sensitivity, specificity, negative predictive value, and positive predictive value.
were calculated. Time-to-event curves were plotted by the Kaplan-Meier method, and differences were compared with the log-rank test to analyze the ability of the score for stratification across risk categories. Hazard ratio was calculated using Cox regression method. Throughout the analysis, only patients with available data were compared and the cohorts’ size is specified in figures and tables.

\[ P \text{ less than } .05 \] was considered statistically significant.

Data sets can be made available upon formal request to the corresponding author. All the analysis was performed with R v4.0.3.

### RESULTS

#### Patient demographic and clinical characteristics

The demographic, clinical, and laboratory characteristics of the patients in the external validation cohorts 1 and 2, and in the second wave validation cohort 3, are presented in Table I. Most patients in the 3 validation cohorts were men, in their sixties, and were hospitalized for a median of 8 days. Median age was higher in the validation cohorts compared with the development cohort, and it was highest in the external validation cohort 2 (median age, 69 years). Patient severity was also greater in the validation cohorts compared with the development cohort, as measured by the rate of ICU admission, which ranged from 12.6% to 18.9%, and the mortality rate, which ranged from 10.7% to 23.8%. The external validation cohort 1 exhibited the highest degree of disease severity, with 18.9% of patients admitted to ICU and 23.8% patients dying during hospitalization.

#### Analysis of prediction model variables according to disease outcome

In the development cohort, the 5 variables with the highest association with mortality and the highest contribution to the predictive capacity of the model were included in the final prediction model. These variables (SpO2/FiO2, N/L ratio, LDH, IL-6, and age) differed statistically between the development and validation cohorts (Table I). Nonetheless, in the validation cohorts, all the variables were significantly increased in patients who died compared with those who survived, except for SpO2/FiO2 in the second wave validation cohort 3, which was relatively high in nonsurvivors (Table II). Overall, patients who died were significantly older and had significantly lower SpO2/FiO2 and higher levels of N/L ratio, LDH, and IL-6 than patients who survived.

#### Validation of the mortality prediction model

We first evaluated the generalizability of our prediction model in 2 external cohorts with a large number of patients from the first wave of COVID-19 in Spain. Validation of the model demonstrated a good predictive performance in both, the external validation cohort 1 (\( N = 185 \)) (AUC = 0.93; 95% CI, 0.88-0.99) and the external validation cohort 2 (\( N = 730 \)) (AUC = 0.86; 95% CI, 0.81-0.91) (Fig 2). Notably, the model showed a superior predictive capacity with the highest AUC, when compared with each individual risk factor, except for SpO2/FiO2 in the second wave validation cohort 3, which was relatively high in nonsurvivors (Table II). Overall, patients who died were significantly older and had significantly lower SpO2/FiO2 and higher levels of N/L ratio, LDH, and IL-6 than patients who survived.

### Table II. Prediction model variables measured in survivor and nonsurvivor patients

<table>
<thead>
<tr>
<th>Variables</th>
<th>Survivors</th>
<th>IQR</th>
<th>Nonsurvivors</th>
<th>IQR</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Development cohort (H12O)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpO2/FiO2</td>
<td>352</td>
<td>269-457</td>
<td>107</td>
<td>101-209</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>N/L Ratio</td>
<td>4.0</td>
<td>2.3-7.8</td>
<td>17.3</td>
<td>8.8-30.3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>LDH (U/L)</td>
<td>341</td>
<td>269-443</td>
<td>537</td>
<td>391-715</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>IL-6 (pg/mL)</td>
<td>17</td>
<td>5-44</td>
<td>86</td>
<td>20-225</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Age (y)</td>
<td>52</td>
<td>44-59</td>
<td>65</td>
<td>56-71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>External validation cohort 1 (RyC)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpO2/FiO2</td>
<td>296</td>
<td>204-346</td>
<td>106</td>
<td>100-190</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>N/L Ratio</td>
<td>5.6</td>
<td>3.4-10.3</td>
<td>8.5</td>
<td>5.0-12.3</td>
<td>.099</td>
</tr>
<tr>
<td>LDH (U/L)</td>
<td>354</td>
<td>296-425</td>
<td>512</td>
<td>395-622</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>IL-6 (pg/mL)</td>
<td>28</td>
<td>10-54</td>
<td>55</td>
<td>26-87</td>
<td>.0004</td>
</tr>
<tr>
<td>Age (y)</td>
<td>59</td>
<td>50-69</td>
<td>73</td>
<td>69-79</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>External validation cohort 2 (FJD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpO2/FiO2</td>
<td>350</td>
<td>266-448</td>
<td>129</td>
<td>103-341</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>N/L Ratio</td>
<td>4.8</td>
<td>2.7-8.9</td>
<td>9.1</td>
<td>6.1-16.8</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>LDH (U/L)</td>
<td>277</td>
<td>220-361</td>
<td>371</td>
<td>249-440</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>IL-6 (pg/mL)</td>
<td>28</td>
<td>8-63</td>
<td>45</td>
<td>14-143</td>
<td>.0004</td>
</tr>
<tr>
<td>Age (y)</td>
<td>67</td>
<td>55-80</td>
<td>84</td>
<td>77-90</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Second wave validation cohort 3 (H12O)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/L Ratio</td>
<td>6.3</td>
<td>3.9-10.0</td>
<td>14.8</td>
<td>6.7-22.7</td>
<td>.007</td>
</tr>
<tr>
<td>LDH (U/L)</td>
<td>389</td>
<td>317-457</td>
<td>440</td>
<td>382-681</td>
<td>.021</td>
</tr>
<tr>
<td>IL-6 (pg/mL)</td>
<td>30</td>
<td>16-57</td>
<td>78</td>
<td>33-195</td>
<td>.005</td>
</tr>
<tr>
<td>Age (y)</td>
<td>58</td>
<td>49-71</td>
<td>78</td>
<td>69-85</td>
<td>.016</td>
</tr>
</tbody>
</table>

\( FJD, \) Hospital Universitario Fundación Jiménez Díaz; \( H12O, \) Hospital Universitario 12 de Octubre; \( IQR, \) interquartile range; \( RyC, \) Hospital Universitario Ramón y Cajal.
patients with low and high risk of death in both external cohorts ($P < .0001$; Fig 3). We next assessed the calibration of the model, that is, the agreement between the observed and the model-predicted mortality in the external validation cohorts. As shown in the calibration plots, the model performed remarkably well at predicting patient mortality (Fig 4). The slopes in the external validation cohorts 1 and 2 were close to 1 (0.86 and 0.94, respectively), and the intercepts were close to 0 (0.05 and 0.03, respectively). We have depicted in a circular barplot how the observed mortality increased steadily as the predicted mortality increased (Fig 4). Altogether, these results indicate a good performance of the model and a minimal difference between predicted and observed values, and, therefore, they validate the model externally.

We considered it relevant to further validate the utility of this prediction model with patients from the second wave of COVID-19 cases in Spain. We built a prospective internal validation cohort with patients admitted to the same hospital as the development cohort, but approximately 6 months apart (validation cohort 3, $N = 119$). The mortality prediction model also maintained a good performance during the second wave ($AUC = 0.86; 95\% CI, 0.75-0.97$) (Fig 2, D). The probability of dying was calculated for all patients on admission, and it was significantly higher in nonsurvivors than in patients who survived, either with intensive care requirement ($P < .01$) or without it ($P < .0001$) (Fig 3, D). Kaplan-Meier analysis also showed a very significant difference in survival for patients below or above the model threshold during the second wave ($P < .001$) (Fig 3, H). The calibration curve of the model performance on the second wave validation cohort 3 also did reasonably well, with a slope of 0.79 and an intercept of 0.10. These results validate the accuracy of our mortality prediction model in patients during the second wave.
The prediction model accurately identified patients at high risk of dying in the validation cohorts. A-D, The probability of dying predicted by the model was significantly higher in nonsurvivors (red), than in survivors who required intensive care (blue) and than in survivors who did not require intensive care (gray), in all cohorts. Dashed lines indicate the model’s optimal cutoff for mortality (0.07). E-H, Using the model’s optimal cutoff, Kaplan-Meier analysis showed a very different survival between the groups with low and high risk of death (P < .001, for all cohorts). FJD, Hospital Universitario Fundación Jiménez Díaz; H12O, Hospital Universitario 12 de Octubre; RyC, Hospital Universitario Ramón y Cajal. Color shades represent the 95% CI. Time is indicated in days.
Update of the mortality prediction model

Because we observed significant differences in the variables included in the model between the cohorts, we revised the model with the patients from the entire 4 cohorts (N = 1477) with the aim of increasing its generalizability. In the updated model, the 5 variables had little collinearity (demonstrated by variation inflation factor < 1.4 for each of the 5 parameters included in the model) and significantly contributed to the model’s prediction capacity (data not shown). Age increased its weight in the updated model, whereas SpO2/FiO2, N/L ratio, and IL-6 reduced it (see the 2 model equations in the Methods section). The classifying accuracy of the updated model was very robust (AUC = 0.91; 95% CI, ...

FIG 4. Calibration analysis depicting the predicted vs observed mortality. A-C, Calibration curves were close to the diagonal dotted line, which represents ideal calibration in which predicted and observed risks are identical. Intercept (Fig 4, A) and slope (Fig 4, B) are shown for each validation cohort. D, Patients were grouped in 5 brackets of increasing probability of death (0%-20%, 21%-40%, 41%-60%, 61%-80%, and 81%-100%). The circular barplot shows how the observed mortality increased steadily as the predicted mortality increased. Of note, in the 2 external validation cohorts, mortality in the highest risk bracket was lower than predicted. Observed mortality rate is represented in concentric circles. FJD, Hospital Universitario Fundación Jiménez Díaz; H12O, Hospital Universitario 12 de Octubre; RyC, Hospital Universitario Ramón y Cajal.
0.87-0.94) (Fig 5, A), and it was not affected by patients’ sex (see Fig E1 in this article’s Online Repository at www.jacionline.org). Cross-validation of the updated model showed a substantial agreement between the predicted and observed mortality (accuracy of 0.91 [0.89-0.93] and kappa coefficient 0.51 [0.42-0.60]). An updated model including only SpO2/FiO2, N/L ratio, LDH, and age was also developed, to be used in settings without IL-6 availability (see model equation in the Methods section). The classifying accuracy of the 4-variable updated model was also robust (AUC = 0.89; 95% CI, 0.86-0.92); however, it was significantly lower than that of the updated model including IL-6 (P < .05).

The Youden index–based cutoff generated for the updated model containing 5 variables was 0.107, with a 0.85 sensitivity, 0.81 specificity, 0.37 positive predictive value, and 0.98 negative predictive value. Kaplan-Meier analysis showed very clear differences in survival for patients below or above this model threshold (P < .0001; Fig 5, B). Of note, the probability of survival in the low-risk group started to decrease after the third week of hospitalization, suggesting that the model decreased its accuracy in patients with long hospital stays.

Clinical significance of the updated prediction model

Respiratory function is a key factor evaluated in patients with COVID-19 to assess their disease severity. As an example of the clinical utility of this prediction model, we analyzed its utility at predicting fatal outcome in patients without ARDS at the time of evaluation. In the overall cohort, 832 patients had no ARDS on admission, of whom 37 (4.4%) finally died. The cutoff value established in the prediction model classified 755 patients as low risk and 77 patients as high risk. In the low-risk group, 17 (2.3%) patients died, whereas in the high-risk group, 20 (26%) patients passed away. Kaplan-Meier analysis showed that patients classified as high risk by the model survived significantly less than patients classified as low risk (P < .0001) (Fig 6). Moreover, patients without ARDS classified as high risk had 8.76 times more risk of dying than low-risk patients (hazard ratio, 8.76; 95% CI, 5.74-13.36; P < .0001). These results indicate that the model can be helpful at identifying patients at high risk of dying, who a priori may not be considered as potentially severe.

DISCUSSION

Here, we validate our previously published mortality prediction model for hospitalized COVID-19 patients in 2 independent external validation cohorts and in a second wave validation cohort (Fig 1). From a methodological point of view, this work reinforces the utility of BigData-oriented frameworks to structure and extract patients’ information from electronic clinical records. Artificial intelligence–based data mining has already been successfully used to obtain results regarding COVID-19 diagnosis and outcome prediction.
The present prediction model, which includes 5 easily collected clinical and laboratory biomarkers (namely SpO₂/FiO₂, N/L ratio, LDH, IL-6, and age), performed with high accuracy in the 2 external validation cohorts, with AUC of 0.93 and 0.86 and well-fitted calibration curves (Figs 2-4). The data for the external validation were collected retrospectively, from patients hospitalized with COVID-19 in 2 unrelated hospitals during the first wave of the pandemic in Spain.

This external validation is important because it showed that, despite significant differences between the development and the validation cohorts, the prediction model worked with high accuracy in the latter. There were differences in the biomarkers included in the model between the cohorts (Table I). The model was developed in a younger and less severe cohort than the validation cohorts, as shown by higher SpO₂/FiO₂ and lower ICU admission and death rates. In addition, levels of IL-6 were measured by cytometric bead array in the development cohort and the external validation cohort 1, whereas they were measured by chemiluminescence in the external validation cohort 2. It was relevant to prove that different laboratory techniques could be used without altering the model performance, especially regarding IL-6, whose measurement is not as standardized as that of other inflammation markers. Finally, the measurement of the biomarkers included in the predictive model took place in the first days of hospitalization in the development cohort and the external validation cohort 1, whereas it occurred at ER in the external validation cohort 2. It was relevant to prove that different laboratory techniques could be used without altering the model performance, especially regarding IL-6, whose measurement is not as standardized as that of other inflammation markers. Finally, the measurement of the biomarkers included in the predictive model took place in the first days of hospitalization in the development cohort and the external validation cohort 1, whereas it occurred at ER in the external validation cohort 2. The model performed similarly for inpatients and ER patients who were subsequently hospitalized. These considerations demonstrate that this prediction model can work in different clinical settings.

To improve the predictive performance in any clinical setting, we have updated the model by reestimating the weights of the 5 biomarkers in an overall cohort including all patients from the 4 cohorts (N = 1477). The accuracy of the updated model was excellent (AUC, 0.91), and the cutoff allowed for the correct separation of low- and high-risk patients (Fig 5). Of note, the survival curves of these 2 patient groups were remarkably distinct since the beginning of hospitalization. In addition, this updated model can be useful to discriminate between patients with likely fatal outcome and those who are likely to survive, among patients without respiratory distress at the time of evaluation (Fig 6).

A limitation of this study is, however, related to the accuracy of the updated model in the low mortality risk group in the long-term (Fig 5, B). Of the initially 1083 patients classified as low risk, 26 (2.4%) died. Three weeks after evaluation, only 51 low-risk patients were still hospitalized, of whom 8 (15.7%) died. This shows that the probability of survival decreases in patients classified as low risk if they have not been discharged after 3 weeks. This should be taken into account for patients who have a prolonged hospital stay and suggests that it could be helpful to periodically reassess their mortality risk. Additional limitations to this study...
are the need for further external validation of the updated model and the need for impact studies in different populations and health care systems to be conducted.

Conclusions
Our mortality prediction model has been validated in 3 large validation cohorts, 2 external and 1 from the second wave. Moreover, the model has been updated, showing good discrimination and excellent calibration. This suggests that the updated model is likely to be generalizable to other populations and clinical settings, and that its predictive performance should be accurate when applied to its target population, patients with COVID-19 who are attended at ER and require hospitalization, and patients with COVID-19 who have recently been hospitalized. This model, and the COR+12 online calculator, could potentially assist in efficient classification of patients with COVID-19, and contribute to guide medical decisions.

We thank all patients, nurses, and medical colleagues who contributed to this study.

Clinical implications: This updated model has been developed with patients from first and second COVID-19 waves and provides the probability of dying, even in patients without respiratory distress at the time of evaluation.

REFERENCES
FIG E1. Classification performance of the updated model by sex. The capacity of the updated model to predict mortality was similar in men (AUC, 0.91; 95% CI, 0.87-0.94) than in women (AUC, 0.87; 95% CI, 0.82-0.93), with no statistical difference in accuracy between sex (De Long test, \( P = .27 \)).