COVID-19 mortality risk assessment: An international multi-center study

Dimitris Bertsimas1,2,*, Galit Lukin2, Luca Mingardi1,2,*, Omid Nohadani3, Agni Orfanoudaki2, Bartolomeo Stellato1,2,*, Holly Wiberg2,*, Sara Gonzalez-Garcia4, Carlos Luis Parra-Calderon5, Kenneth Robinson5, Michelle Schneider5, Barry Stein5, Alberto Estirado6, Lia a Beccara7, Rosario Canino7, Martina Dal Bello8, Federica Pezzetti7, Angelo Pan7, The Hellenic COVID-19 Study Group8

1 Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, 2 Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, 3 Benefits Science Technologies, Boston, Massachusetts, United States of America, 4 Institute of Biomedicine of Seville (IBIS), Virgen del Rocío University Hospital, CSIC, University of Seville, Seville, Spain, 5 Hartford HealthCare, Hartford, Connecticut, United States of America, 6 HM Hospitals, Madrid, Spain, 7 Azienda Socio-Sanitaria Territoriale di Cremona, Cremona, Italy, 8 Physics of Living Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America

* These authors contributed equally to this work.
¶ Membership of the Hellenic COVID-19 Study Group is listed in the Acknowledgments.
* dbertsim@mit.edu

Abstract

Timely identification of COVID-19 patients at high risk of mortality can significantly improve patient management and resource allocation within hospitals. This study seeks to develop and validate a data-driven personalized mortality risk calculator for hospitalized COVID-19 patients. De-identified data was obtained for 3,927 COVID-19 positive patients from six independent centers, comprising 33 different hospitals. Demographic, clinical, and laboratory variables were collected at hospital admission. The COVID-19 Mortality Risk (CMR) tool was developed using the XGBoost algorithm to predict mortality. Its discrimination performance was subsequently evaluated on three validation cohorts. The derivation cohort of 3,062 patients has an observed mortality rate of 26.84%. Increased age, decreased oxygen saturation (≤ 93%), elevated levels of C-reactive protein (≥ 130 mg/L), blood urea nitrogen (≥ 18 mg/dL), and blood creatinine (≥ 1.2 mg/dL) were identified as primary risk factors, validating clinical findings. The model obtains out-of-sample AUCs of 0.90 (95% CI, 0.87–0.94) on the derivation cohort. In the validation cohorts, the model obtains AUCs of 0.92 (95% CI, 0.88–0.95) on Seville patients, 0.87 (95% CI, 0.84–0.91) on Hellenic COVID-19 Study Group patients, and 0.81 (95% CI, 0.76–0.85) on Hartford Hospital patients. The CMR tool is available as an online application at covidanalytics.io/mortality_calculator and is currently in clinical use. The CMR model leverages machine learning to generate accurate mortality predictions using commonly available clinical features. This is the first risk score trained and validated on a cohort of COVID-19 patients from Europe and the United States.
commitment to using the data only for research purposes and not to identify any individual participant; 2) a commitment to securing the data using appropriate measures, and 3) a commitment to destroy or return the data after analyses are complete. Requests can be made to the research team of the corresponding institutions.

Researchers may contact covidanalytics_cmr@mit.edu, which will connect the request to designated representatives from each institution: Rosario Canino (ASST Cremona), Alberto Estirado (HM Hospitals), Carlos Luis Parra Calderón (Seville), Angelos Pefanis (Hellenic CSG), and Michelle Schneider (Hartford HealthCare). These authors have been selected by their respective institutions to address data requests. There is no single institutional body that manages the full database; the data has been curated from various hospital systems through the participating authors only.

Funding: HW is supported by the National Science Foundation Graduate Research Fellowship under Grant No. 174530. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation. This work was partially supported by a grant from c3.ai for COVID-19 related research; this organization played no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Benefits Science Technologies provided support in the form of salaries for ON, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Hartford HealthCare provided support in the form of salaries for KN, MS, and BS, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the ‘author contributions’ section. No further funding was provided for the study.

Competing interests: KN, MS, BS are employed by Hartford HealthCare. ON is employed by Benefits Science Technologies. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Introduction

The ongoing coronavirus disease pandemic (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has led to an alarming number of casualties across the world [1]. As the pandemic progresses globally, much remains unknown about the disease dynamics and risk factors. A better understanding of the clinical determinants of disease severity can improve patient management throughout the healthcare system. This task is challenging due to the rapid spread of the disease and the lack of detailed patient data.

Leveraging machine learning (ML) methods enables the rapid discovery of insights across large populations of heterogeneous patients. An algorithmic approach provides an objective evaluation and can often capture nonlinear interactions that are not obvious from pure observation of the population. Researchers have recognized the potential of these data-driven approaches across various facets of the effort to combat COVID-19 [2].

In this work, we present the COVID-19 Mortality Risk (CMR) tool, a novel ML model for predicting mortality in hospitalized COVID-19 patients. It enables physicians to better triage patient care in a resource-constrained system through a personalized mortality risk score. The CMR model synthesizes various clinical data elements from multiple European and US centers, including demographics, lab test results, symptoms, and comorbidities. We use the XGBoost algorithm [3], a leading ML method, to predict mortality probabilities. This score is able to capture nonlinearities in risk factors, resulting in strong predictive performance with an out-of-sample area under the receiver operating characteristic curve (AUC) of 0.90 (95% CI, 0.87–0.94). It also validates commonly accepted risk factors, such as age and oxygen saturation, while discerning novel insights.

The CMR tool leverages an international cohort from three hospital systems in Italy, Spain, and the United States. The model is subsequently validated on hospitalized patients in a consortium of six hospitals from Greece, Spain, and the United States. Each region presents a diverse set of patient profiles and mortality rates for the model. By considering severely ill populations from different countries and healthcare systems, the final dataset captures a wide array of features.

In recent months, ML scores have been proposed to predict COVID-19 mortality [4, 5] as well as disease severity [6]. Existing literature largely focuses on Chinese hospitals due to the disease’s emergence in Wuhan [5, 6]. However, it is instrumental to understand the clinical characteristics for more recent and diverse cases, considering that the virus strain may have mutated since surfacing in Wuhan [7]. Pourhomayoun et al. (2020) proposed a model based on a large international dataset, yet this model lacks comprehensive patient data and is thus limited in its ability to derive personalized insights [4]. In this work, we study patients in Europe and the US, offering a new lens into the clinical characteristics of this disease.

Methods

Study population

The study comprises 33 different hospitals, spanning across three countries in southern Europe as well as the US. The collaborating institutions were split into derivation and validation cohorts, as summarized in Table 1. The derivation cohort includes the healthcare systems of ASST Cremona (Northern Italy), HM Hospitals (Spain), and Hartford HealthCare affiliate hospitals (United States). The broad geographic spread of data sources offers a comprehensive sample of some of the most severely impacted regions in the world. To further validate the results, we partnered with Hospital Universitario Virgen del Rocío (Spain), the Hellenic COVID-19 Study Group (Hellenic CSG), a consortium of Greek hospitals, and Hartford
HealthCare’s main hospital (CT, USA). The study population consists of adult patients who were admitted to the hospital with confirmed SARS-CoV-2 infection by polymerase chain reaction testing of nasopharyngeal samples. The time horizon of admissions is displayed in Table 1.

All independent organizations and the Massachusetts Institute of Technology institutional review boards approved this protocol as minimal-risk research using data collected for standard clinical practice and waived the requirement for informed consent. The survey was anonymous and confidentiality of information was assured.

### Clinical features

Data is collected using the electronic health record (EHR) databases and COVID-19 specific registries of the collaborating hospitals. We compile 22 features, including patient demographic information, comorbidities, vitals upon admission, and laboratory test results. The full set of features is outlined in Table 2. The outcome of interest, mortality during the hospital admission, is derived from discharge records. Only the first recorded laboratory test results are considered, typically within 24 hours of admission. Comorbidities are identified using the International Classification of Diseases, 9th and 10th revision, codes of hospital discharges and...
are aggregated into four categories using the Clinical Classifications Software [8]. Missing values are imputed using k-nearest neighbors imputation [9] (S2 Text in S1 File). We exclude risk factors that are not consistently recorded in the derivation cohort, thereby omitting features whose values are more than 40% missing.

Modeling approach

We train a binary classification model in which the outcome is patient mortality: 1, if the patient was deceased, or 0, if discharged. Specifically, we use the XGBoost algorithm [3] for the training process, described further in the (S3 Text in S1 File). For comparison, we also present the predictive performance of other ML methods in the (S3 Table and S4 Text in S1 File). The derivation population is randomly divided into training (85%) and testing (15%) sets, ensuring that mortality prevalence was consistent between the two. We tune seven model parameters by maximizing the K-fold cross-validation AUC using the Optuna optimization framework [10] (S5 Text in S1 File). This technique provides a more accurate parameter search compared to grid search by efficiently pruning suboptimal parameter combinations and continuously refining the search space. We apply SHapley Additive exPlanations (SHAP) to generate importance plots for transparency of the model predictions and risk drivers [11] (S6 Text in S1 File). All statistical analysis is conducted using version 3.7 of the Python programming language.

Table 2. Summary statistics of all patient characteristics for the total sample, the survivor, and non-survivor cohorts.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Median (IQR)</th>
<th>Survivor (N = 2,302)</th>
<th>Non-Survivor (N = 760)</th>
<th>P-Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>68.0 (57.0–79.0)</td>
<td>64.0 (54.0–74.0)</td>
<td>80.0 (73.0–85.0)</td>
<td>&lt;1.0E-04</td>
</tr>
<tr>
<td>Female*</td>
<td>1207.0 (39.42%)</td>
<td>958.0 (41.62%)</td>
<td>249.0 (32.76%)</td>
<td>&lt;1.0E-04</td>
</tr>
<tr>
<td>Heart Rate (bpm)</td>
<td>90.0 (80.0–102.0)</td>
<td>91.0 (80.0–102.0)</td>
<td>88.0 (79.0–100.25)</td>
<td>7.7E-03</td>
</tr>
<tr>
<td>Oxygen Saturation (%)</td>
<td>94.0 (91.0–96.0)</td>
<td>94.4 (92.0–96.0)</td>
<td>90.55 (85.5–94.0)</td>
<td>&lt;1.0E-04</td>
</tr>
<tr>
<td>Temperature (°F)</td>
<td>98.6 (97.7–99.86)</td>
<td>98.42 (97.59–99.86)</td>
<td>98.79 (97.7–100.08)</td>
<td>9.3E-04</td>
</tr>
<tr>
<td>Alanine Aminotransferase (U/L)</td>
<td>27.0 (17.0–43.0)</td>
<td>27.0 (17.22–44.53)</td>
<td>26.0 (16.12–41.0)</td>
<td>7.9E-02</td>
</tr>
<tr>
<td>Aspartate Aminotransferase (U/L)</td>
<td>35.9 (25.3–54.5)</td>
<td>34.0 (24.55–50.2)</td>
<td>44.0 (30.0–68.0)</td>
<td>&lt;1.0E-04</td>
</tr>
<tr>
<td>Blood Glucose (mg/dL)</td>
<td>118.35 (105.0–142.0)</td>
<td>115.0 (103.4–134.0)</td>
<td>134.0 (113.0–170.55)</td>
<td>&lt;1.0E-04</td>
</tr>
<tr>
<td>Blood Urea Nitrogen (mg/dL)</td>
<td>17.0 (12.62–25.56)</td>
<td>15.0 (11.61–20.96)</td>
<td>29.0 (20.46–46.0)</td>
<td>&lt;1.0E-04</td>
</tr>
<tr>
<td>C-Reactive Protein (mg/L)</td>
<td>73.37 (28.88–146.43)</td>
<td>58.62 (22.74–117.83)</td>
<td>137.93 (69.81–214.13)</td>
<td>&lt;1.0E-04</td>
</tr>
<tr>
<td>Creatinine (mg/dL)</td>
<td>0.95 (0.77–1.22)</td>
<td>0.9 (0.74–1.08)</td>
<td>1.25 (0.95–1.75)</td>
<td>&lt;1.0E-04</td>
</tr>
<tr>
<td>Hemoglobin (g/dL)</td>
<td>13.9 (12.6–14.9)</td>
<td>13.9 (12.8–15.0)</td>
<td>13.4 (11.9–14.6)</td>
<td>&lt;1.0E-04</td>
</tr>
<tr>
<td>Mean Corpuscular Volume (fL)</td>
<td>88.0 (85.0–91.2)</td>
<td>87.7 (84.9–90.7)</td>
<td>89.4 (85.93–92.9)</td>
<td>&lt;1.0E-04</td>
</tr>
<tr>
<td>Platelets (10^3/μL)</td>
<td>202.0 (157.0–259.75)</td>
<td>205.0 (160.0–263.0)</td>
<td>187.0 (146.5–248.5)</td>
<td>2.0E-04</td>
</tr>
<tr>
<td>Potassium (mmol/L)</td>
<td>4.05 (3.7–4.4)</td>
<td>4.0 (3.7–4.4)</td>
<td>4.1 (3.7–4.6)</td>
<td>&lt;1.0E-04</td>
</tr>
<tr>
<td>Prothrombin Time (INR)</td>
<td>1.11 (1.02–1.25)</td>
<td>1.11 (1.02–1.23)</td>
<td>1.13 (1.02–1.31)</td>
<td>&lt;1.0E-04</td>
</tr>
<tr>
<td>Sodium (mmol/L)</td>
<td>137.1 (135.0–140.0)</td>
<td>137.0 (135.0–139.5)</td>
<td>138.0 (135.0–141.0)</td>
<td>&lt;1.0E-04</td>
</tr>
<tr>
<td>White Blood Cell Count (10^3/μL)</td>
<td>6.73 (5.13–9.09)</td>
<td>6.51 (5.05–8.59)</td>
<td>7.92 (5.37–11.0)</td>
<td>&lt;1.0E-04</td>
</tr>
<tr>
<td>Cardiac dysrhythmias*</td>
<td>201.0 (6.56%)</td>
<td>128.0 (5.56%)</td>
<td>73.0 (9.61%)</td>
<td>5.8E-04</td>
</tr>
<tr>
<td>Chronic kidney disease*</td>
<td>72.0 (2.35%)</td>
<td>40.0 (1.74%)</td>
<td>32.0 (4.21%)</td>
<td>1.5E-03</td>
</tr>
<tr>
<td>Heart disease*</td>
<td>125.0 (4.08%)</td>
<td>80.0 (3.48%)</td>
<td>45.0 (5.92%)</td>
<td>9.2E-03</td>
</tr>
<tr>
<td>Diabetes*</td>
<td>384.0 (12.54%)</td>
<td>263.0 (11.42%)</td>
<td>121.0 (15.92%)</td>
<td>2.5E-03</td>
</tr>
</tbody>
</table>

*Count (proportion) is reported for binary variables.

* P-value reports significance of a two-sided T-test between the survivor and non-survivor populations.

https://doi.org/10.1371/journal.pone.0243262.t002
Performance evaluation

All predictive models are evaluated based on their ability to discriminate between outcomes for each population. We report results for the training and testing sets of the derivation cohort, as well as for each independent institution in the validation cohort, with the corresponding confidence intervals (CI). The AUC, accuracy, specificity, precision, and negative predictive value are computed for all patient subpopulations across different thresholds. Receiver operating characteristic (ROC) curves were created for each of the cohorts.

Results

Patient characteristics

The CMR model is created using a derivation population of 3,062 patients, of which 1,441 are from ASST Cremona, 1,390 from HM Hospitals, and 231 from Hartford Affiliates. The validation population consists of 865 patients: 219 patients from Seville, 323 from the Hellenic CSG, and 323 from Hartford Hospital. The clinical characteristics of the derivation population are outlined in Table 2. The average observed mortality rate in this population is 26.84%. In comparison to survivors, non-survivors tend to be older (median age 80 vs. 64) and more commonly men (67.2% vs. 58.4% of cohort). Moreover, the prevalence of comorbidities such as cardiac dysrhythmias, chronic kidney disease, and diabetes is higher in the non-survivor population (9.61%, 4.21% and 15.92% versus 5.56%, 1.74%, and 11.42%, respectively). The clinical characteristics for each participating study site are reported in the (S1 and S2 Tables in S1 File).

Performance metrics

The final mortality model exhibits an out-of-sample AUC of 0.90 (95% CI, 0.87–0.94) on the derivation testing set; see Table 3. The AUC for the Seville cohort is slightly higher at 0.92 (95% CI, 0.88–0.95). For the other two validation centers, there is a decrease in AUC. In the Hellenic CSG cohort, the model performs 0.87 (95% CI, 0.84–0.91) and in the Hartford Hospital population 0.81 (95% CI, 0.76–0.85). The corresponding ROC curves are included in the (S1 Fig in S1 File).

A different threshold is selected for each cohort to enforce a minimum sensitivity of 80%. Given the implications of these predictions, we report conservative risk estimates in order to ensure that all critically ill patients are accounted for. This comes at the expense of specificity, i.e., it increases the number of patients whom we may incorrectly flag as high risk of mortality. For the fixed sensitivity requirement, we achieve a classification accuracy of 0.85 (95% CI, 0.81–0.89) in the testing set with specificity of 0.87 (95% CI, 0.83–0.92); see Table 3.

The model generalizes better in the Seville cohort with an accuracy of 0.87 (95% CI, 0.82–0.91) and specificity of 0.87 (95% CI, 0.83–0.92). The necessary threshold for a sensitivity of 80% is lower for the Hellenic CSG compared to the other populations. This is due to the low baseline incidence of mortality in this sample when compared to the derivation and other

| Table 3. AUC performance (%) and threshold-based metrics for training, testing, and validation population. |
|---|---|---|---|---|---|---|---|
| Cohort | N | AUC | Threshold | Accuracy | Specificity | Precision | Negative predictive value |
| Training Set | 2755 | 94.7 (93.87,95.54) | 38.44 (36.62,40.25) | 89.62 (88.48,90.76) | 92.76 (91.79,93.73) | 78.51 (76.98,80.04) | 93.39 (92.46,94.32) |
| Testing Set | 307 | 90.19 (86.86,93.52) | 28.3 (23.26,33.34) | 85.02 (81.02,89.01) | 86.58 (82.77,90.39) | 66.3 (61.02,71.59) | 93.02 (90.17,95.87) |
| Hellenic CSG | 323 | 87.45 (83.83,91.06) | 20.23 (15.85,24.61) | 74.92 (70.27,79.65) | 74.23 (69.46,79.0) | 25.74 (20.97,30.51) | 97.3 (95.53,99.07) |
| Seville | 219 | 91.62 (87.95,95.29) | 33.21 (26.98,39.45) | 86.76 (82.27,91.25) | 87.43 (83.04,91.82) | 48.94 (42.32,55.56) | 97.09 (94.87,99.32) |
| Hartford | 323 | 80.66 (76.36,84.97) | 29.74 (24.75,34.72) | 61.3 (55.99,66.61) | 58.12 (52.74,63.5) | 24.18 (19.51,28.85) | 94.71 (92.26,97.15) |

https://doi.org/10.1371/journal.pone.0243262.t003
validation cohorts. The model achieves lower performance in this set of patients, with an accuracy of 0.75 (95% CI, 0.7–0.8) and specificity of 0.74 (95% CI, 0.69–0.79). For Hartford Hospital, the accuracy of CMR is 0.61 (95% CI, 0.56–0.67) with a specificity of 0.58 (95% CI, 0.53–0.64).

**Model results**

Through the SHAP framework, we identified the most important drivers of mortality risk and the interplay between individual features. For a particular patient, SHAP values indicate the feature contributions towards the risk. The patient risk normalized between 0 and 1 is the sum of the SHAP values of all the features (S6 Text in S1 File). Fig 1a displays the risk contributions of the 10 most important features. For example, higher values of age (red) yield higher SHAP values, suggesting that older patients are at higher risk. In contrast, the SHAP value increases with lower values (blue) of Oxygen Saturation, suggesting an inverse relationship with this feature.

When BUN is below 20 mg/dL, the mortality risk decreases, particularly for ages below 55 years. On the other hand, BUN values greater than 25 mg/dL for older patients increase the risk (Fig 1b). A C-reactive protein (CRP) between 50 and 130 mg/L does not affect the risk, independent of age. As CRP goes below 50 mg/L, the mortality risk decreases. For a CRP above 160 mg/L, the elevated risk does not change and is higher for older patients (Fig 1c). An oxygen saturation below 93% increases the mortality risk rapidly and this trend is accelerated by growing age (Fig 1d). A blood creatinine level greater than 1.2 mg/dL increases the risk moderately, specifically for older patients. Levels above 3 mg/dL rapidly escalate the mortality risk (Fig 1e). Fig 1f illustrates that while a blood glucose less than 130 mg/dL lowers the risk, it can increase the risk for levels above 180 mg/dL, in particular for older patients. An aspartate aminotransferase (AST) level above 65 U/L increases the risk, while a level below 25 U/L sharply decreases the risk, independent of age (Fig 1g). A platelet count in $10^3/\mu$L affects the risks in 4 distinct ranges: (i) below 50 the risk is elevated, (ii) between 50 and 180 the risk is marginally increased (more for older patients), (iii) between 180 and 330 the risk is slightly decreased, and (iv) above 330 the risk is sizably decreased (Fig 1h). Fig 1i shows that a mean corpuscular volume (MCV) between 90 and 94 fL increases the risk moderately, while other values have only small effects. Lastly, an increased risk is observed when white blood cell (WBC) count is above 10 in $10^3/\mu$L, in particular for older patients (Fig 1j).

**Discussion**

The CMR calculator predicts mortality with high accuracy using clinical measurements collected early within a patient’s hospital admission. An early risk assessment of patient mortality allows physicians to triage patients and prioritize resources in a highly congested system. It uses commonly available laboratory results and does not require imaging results or advanced testing. The presented tool can be particularly useful in lower acuity facilities or remote hospitals with constrained diagnostic capabilities.

Age is the most important determinant of mortality in the model: older patients have higher mortality risk, which has been observed in retrospective patient analysis [12] and subsequently reflected in public health guidance [13]. Predicted mortality also increases for patients with low oxygen saturation, corroborating findings that link hypoxemia to mortality [14], as well as the observed prevalence of shortness of breath in severe patients [15]. This measurement additionally serves as signal of respiratory distress, and respiratory failure has been found clinically as one of the major mortality causes of COVID-19 [16]. This can also appear in cases of silent hypoxia where shortness of breath is not observed [17].
Fig 1. SHAP importance plots for final model. The top 10 features are displayed in panel (a), ordered by decreasing significance. For a given feature, the corresponding row indicates the SHAP values as the feature ranges from its lowest (blue) to highest (red) value. Panel (b)-(j) display the individual feature plots and the impact of each feature on the mortality risk (colors indicate the age here) with gray areas indicating reference ranges.

https://doi.org/10.1371/journal.pone.0243262.g001
Our study finds that elevated BUN, CRP, creatinine, glucose, AST, and platelet counts are highly significant laboratory features. Several of these biomarkers have been identified in other retrospective analyses of mortality outcomes of COVID-19 [16, 18]. Prior work has also uncovered the critical role of these biomarkers in identifying severe cases of patients with community acquired pneumonia [19]. The PSI, CURB-65, and SCAP scores are also based on similar risk factors such as glucose levels ≥ 250 mg/dL and BUN > 19 mg/dL [20, 21]. Moreover, CRP levels have been recognized to characterize severity for H1N1 patients [22].

CRP is a widely available inflammatory marker which has been independently observed as a biomarker of COVID-19 severity [6, 23]. Our findings show that CRP values outside the reference ranges do not necessarily increase the risk of mortality. In fact, CRP has a negative effect on mortality until approximately 50 mg/L, it has a negligible effect between approximately 50mg/L and 130 mg/L, and it significantly increases the mortality risk above 130mg/L. Elevated BUN and creatinine levels are both indicative of impaired kidney function, which has been associated with poor prognosis [24]. The individual feature plots indicate a clear transition from low to high risk when BUN exceeds approximately 18 mg/dL and creatinine exceeds approximately 1.2 mg/dL. These values are slightly lower than reference ranges for these values, providing data-driven validation of the ranges [25] targeted for COVID-19. The increase in mortality risk for patients with elevated glucose levels is consistent with the reports in other studies of diabetes as a risk factor [12, 26]. Elevated AST levels have been observed due to liver dysfunction in severe COVID-19 cases [27]. Finally, low platelets are associated with increased risk, which match findings of thrombocytopenia in critical COVID-19 patients [28].

We recognize that the derivation populations may differ from other populations based both on hospital conditions and inherent demographic differences. An external validation using Seville, Greece, and US populations allows us to assess the broader clinical utility of our findings. The CMR model performs well on these patients, with the strongest performance observed in Seville. Seville consists of a South European population similar to the majority of the derivation cohort. However, it did not face the same capacity challenges as ASST Cremona and HM Hospitals during the study period. Greece had a significantly lower disease spread, resulting in a lower mortality rate compared to the derivation population. Nevertheless, the model yields comparable results in this cohort to the other European hospitals. Hartford has the weakest validation performance, which may suggest inherent differences between Europe and the US in disease dynamics, treatment protocols, or underlying population susceptibility. This attests to the need for training and validation on a diverse set of populations.

We observe that the thresholds needed for obtaining 80% sensitivity differ across the external validation cohorts. When applying the CMR tool to a new hospital, the threshold should be calibrated to the severity of this population. A sample of historical patients at the hospital can be used to validate the model. Using the risk predictions and true outcomes of this sample, various risk thresholds can be evaluated for sensitivity and specificity. Clinicians can determine the relevant threshold for their hospital’s needs. For example, highly constrained systems may employ a higher threshold (lower sensitivity) due to capacity limits, whereas other centers may use this tool as an initial screening tool where sensitivity is required to be very high.

Risk models are most useful when they are readily available for healthcare clinicians. For this reason, a dynamic online application has been created as the interface of the CMR model for use by clinical providers. Fig 2 provides a visualization of the application that is available at covidanalytics.io/mortality_calculator. After entering a patient’s clinical features, the model returns a predicted mortality risk. It additionally produces a SHAP plot to elucidate the major factors contributing to an individual patient’s risk score. Features in blue decrease risk from the population baseline, whereas features in red increase risk. The contribution is proportional to the width of the feature’s bar. In the example, we see that the patient’s age and oxygen
Fig 2. Visualization of the calculator interface. Using the SHAP package, personalized interpretations of the predicted score are provided to the user.

https://doi.org/10.1371/journal.pone.0243262.g002
saturation levels increase his risk assessment, but his temperature and glucose lower his risk. The CMR tool is currently undergoing prospective validation at two of the collaborating institutions in the study: the application is in use in the emergency room of ASST Cremona to prioritize hospitalizations on higher risk patients, and the model also interoperates with the EHR of the Virgen del Rocio University Hospital in Seville, Spain.

**Limitations**

Limited hospital capacity can impose potential biases in the training population. Only severe patients were able to be treated, particularly in Europe, and some hospitals were forced to turn away patients deemed too critically ill during the peak of the virus. Thus, hospital admissions data may exclude patients on both ends of the acuity spectrum. Additionally, the scarcity of hospital resources may have led patients to receive insufficient care, increasing mortality risk due to lack of treatment. While this warrants further investigation, initial validation results suggest that the CMR tool generalizes well to less congested systems in Greece and the United States.

The differences related to Hartford Hospital might also be related to the timing of the virus. The virus affected Europe before the US. This provided an opportunity to learn from the experience in Europe, which may have resulted in different or more effective treatment decisions as well as governmental policies in the US. This is an opportunity for further study through validation on additional US cohorts.

Our clinical features are limited by the data that was commonly available across all sites in the derivation population. We expect that a more comprehensive set of clinical features such as D-Dimer and IL-6 levels, Body Mass Index, radiographic diagnosis, symptoms, and time elapsed between the disease and treatment onset will yield more accurate results. A broader set of comorbidities, including hypertension, cancer, chronic obstructive pulmonary disease, and others could be included when available. Recent reports on racial disparities and socio-economic determinants of COVID-19 severity [29, 30] could be addressed through the incorporation of additional demographic data and external data sources.

Additionally, there is significant variability in treatment protocols across countries and individual organizations. In future work, we hope to expand the set of captured clinical features and incorporate treatments to disentangle some of the observed heterogeneity in outcomes and clinical characteristics.

**Conclusions**

This international study provides a mortality risk calculator of high accuracy for hospitalized patients with confirmed COVID-19. The CMR model validates several reported risk factors and offers insights through a user-friendly interface. Validation on external data shows strong generalization to unseen populations in both Europe and the United States and offers promise for adoption by clinicians as a support tool.

**Supporting information**

S1 File. (PDF)

S2 File. (PDF)
Acknowledgments

We would like to thank the HM Hospitals for creating and providing us access to the anonymized clinical data set “COVID Data Saves Lives”. It constitutes a valuable database that is available to the broader research community. In order to gain access to this resource, the reader can contact the following address: coviddatasavelives@hmhospitaless.com. We would also like to thank Sophie Testa for the help with the laboratory data collection from ASST Cremona. Moreover, we would like to acknowledge the work of Jeff Mather, Qun Yu, and Lizabeth Roper from the Hartford HealthCare system for their work on data collection. Our team recognizes the contribution and support of José Miguel Cisneros-Herreros, as well as the extensive work of Jesús Moreno-Conde to collect the data from the Hospital Universitario Virgen del Rocío. We would like to thank Álvaro Fernandez Galiana for helping us to get access to the data sources from Spain and reviewing the manuscript. We would also like to thank Aggelos Stefos, Sarah P. Georgiadou, and Anastasia Michail from the Department of Medicine and Research Laboratory of Internal Medicine, National Expertise Center of Greece in Autoimmune Liver Diseases, General University Hospital of Larissa, Larissa, Greece for their help in data collection.

The Hellenic COVID-19 Study Group

The Hellenic COVID-19 Study Group is comprised of the following contributors: Karolina Akinosoglou (University Hospital of Patra), Anastasia Antoniadou (Attikon GH), Katerina Argyraki (Sotiria Thoracic Diseases Hospital of Athens), George N. Dalekos (Department of Medicine and Research Laboratory of Internal Medicine, National Expertise Center of Greece in Autoimmune Liver Diseases, General University Hospital of Larissa), Mina Gaga (Sotiria Thoracic Diseases Hospital of Athens), Nikolaos K. Gatselis (Department of Medicine and Research Laboratory of Internal Medicine, National Expertise Center of Greece in Autoimmune Liver Diseases, General University Hospital of Larissa), Charalambos Gogos (University Hospital of Patra), Ioannis Kalomenidis (Evangelismos Hospital, National and Kapodistrian University of Athens), Eleftheria Kranidioti (Evangelismos Hospital), Lykourgos Kolilekas (Sotiria Thoracic Diseases Hospital of Athens), Giota Lourida (Sotiria Thoracic Diseases Hospital of Athens), Evangelia Margellou (Evangelismos Hospital), Georgios Ntaios (Department of Medicine and Research Laboratory of Internal Medicine, National Expertise Center of Greece in Autoimmune Liver Diseases, General University Hospital of Larissa), Periklis Panagopoulos (University Hospital of Alexandroupolis), Angelos Pefanis (Sotiria Thoracic Diseases Hospital of Athens), Vasilis Petrakis (University Hospital of Alexandroupolis), Christos Psarrakis (Attikon GH), Vissaria Sakka (Evangelismos Hospital), Konstantinos Thomas (Attikon GH), Eleftherios Zervas (Sotiria Thoracic Diseases Hospital of Athens).

Author Contributions

Conceptualization: Dimitris Bertsimas, Luca Mingardi, Agni Orfanoudaki, Bartolomeo Stellato.

Data curation: Luca Mingardi, Agni Orfanoudaki, Bartolomeo Stellato, Holly Wiberg, Sara Gonzalez-Garcia, Carlos Luis Parra-Calderón, Kenneth Robinson, Michelle Schneider, Barry Stein, Alberto Estirado, Lia a Beccara, Rosario Canino, Martina Dal Bello, Federica Pezzetti, Angelo Pan.
Formal analysis: Luca Mingardi, Omid Nohadani, Agni Orfanoudaki, Bartolomeo Stellato, Holly Wiberg.

Funding acquisition: Dimitris Bertsimas.

Investigation: Dimitris Bertsimas, Luca Mingardi, Omid Nohadani, Agni Orfanoudaki, Bartolomeo Stellato, Holly Wiberg.

Methodology: Dimitris Bertsimas, Luca Mingardi, Agni Orfanoudaki, Bartolomeo Stellato, Holly Wiberg.

Project administration: Dimitris Bertsimas.

Resources: Dimitris Bertsimas.

Software: Galit Lukin, Luca Mingardi, Agni Orfanoudaki, Bartolomeo Stellato, Holly Wiberg.

Supervision: Dimitris Bertsimas.

Validation: Luca Mingardi, Agni Orfanoudaki, Bartolomeo Stellato, Holly Wiberg.

Visualization: Galit Lukin, Luca Mingardi, Agni Orfanoudaki, Bartolomeo Stellato, Holly Wiberg.

Writing – original draft: Luca Mingardi, Omid Nohadani, Agni Orfanoudaki, Bartolomeo Stellato, Holly Wiberg.

Writing – review & editing: Dimitris Bertsimas, Galit Lukin, Luca Mingardi, Omid Nohadani, Agni Orfanoudaki, Bartolomeo Stellato, Holly Wiberg, Sara Gonzalez-Garcia, Carlos Luis Parra-Calderón, Kenneth Robinson, Michelle Schneider, Barry Stein, Alberto Estirado, Lia a Beccara, Rosario Canino, Martina Dal Bello, Federica Pezzetti, Angelo Pan.

References


