Evaluating risk stratification scoring systems to predict mortality in patients with COVID-19

Kelly Chu, Batool Alharahsheh, Naveen Garg, Payal Guha

ABSTRACT

Background The COVID-19 pandemic has necessitated efficient and accurate triaging of patients for more effective allocation of resources and treatment.

Objectives The objectives are to investigate parameters and risk stratification tools that can be applied to predict mortality within 90 days of hospital admission in patients with COVID-19.

Methods A literature search of original studies assessing systems and parameters predicting mortality of patients with COVID-19 was conducted using MEDLINE and EMBASE.

Results 589 titles were screened, and 76 studies were found investigating the prognostic ability of 16 existing scoring systems (area under the receiver operating curve (AUROC) range: 0.550–0.966), 38 newly developed COVID-19-specific prognostic systems (AUROC range: 0.640–0.9940), 15 artificial intelligence (AI) models (AUROC range: 0.840–0.955) and 16 studies on novel blood parameters and imaging.

Discussion Current scoring systems generally underestimate mortality, with the highest AUROC values found for APACHE II and the lowest for SMART-COP. Systems featuring heavier weighting on respiratory parameters were more predictive than those assessing other systems. Cardiac biomarkers and CT chest scans were the most commonly studied novel parameters and were independently associated with mortality, suggesting potential for implementation into model development. All types of AI modelling systems showed high abilities to predict mortality, although none had notably higher AUROC values than COVID-19-specific prediction models. All models were found to have bias, including lack of prospective studies, small sample sizes, single-centre data collection and lack of external validation.

Conclusion The single parameters established within this review would be useful to look at in future prognostic models in terms of the predictive capacity their combined effect may harness.

INTRODUCTION

The SARS-CoV-2 outbreak has put enormous strain on healthcare systems around the world. According to the WHO, as of 12 January 2021, there have been more than 91 million cases of COVID-19 reported worldwide, with almost 2 million deaths.1 There is an urgent need for a simple, accurate system to rapidly identify patients who are at the highest risk of death.

Traditionally, scoring systems are used in healthcare to stratify risk, predict outcomes and appropriately manage patients.2 For example, the CRB-65 scoring system is efficiently used to assess the mortality risk of pneumonia in primary care to determine the need for management escalation.3

Risk stratification methods have been effectively used in previous viral outbreaks such as the Ebola epidemic in 2014 to reduce casualties.4 With COVID-19 being a novel disease, no pre-existing risk stratification methods were available, so traditional scoring systems were adapted in the early stages of the pandemic. As the pandemic progressed, COVID-19-specific tools were developed by studying patients’ characteristics relating strongly to mortality and incorporating them into scoring systems.

Although artificial intelligence (AI) algorithm development varies depending on the number of possible outcomes, it is an ideal way of stratifying patients.5 It uses dynamic data and continual updating of its algorithm to increase the accuracy of its predictions.

This review aims to provide a summary of the literature available on risk stratification tools, including prediction models and single parameters used to predict the mortality of patients with COVID-19 to aid clinical decision-making. This review also aims to evaluate the applications of AI in mortality prediction models.

This study hopes to fill in the gaps in the current literature reviewing human and AI scoring tools. In addition, new studies investigating parameters associated with SARS-CoV-2 mortality are being published; therefore, constant evaluation of risk stratification tools is imperative in a rapidly evolving pandemic.
The following search concepts were combined using Boolean operators: COVID-19 (TI, AB, KW) AND Risk stratification (TI, AB, KW) AND Mortality (TI, AB, KW) AB, abstract; KW, keywords; TI, title, the '/' indicated a different variation.

METHODS

A comprehensive search of MEDLINE and EMBASE between 1 January 2019 and 5 January 2021 was conducted to retrieve studies related to mortality risk prediction of patients with COVID-19. The search was done using the keywords and relevant MeSH terms displayed in Table 1.

Inclusion criteria were the following: (1) primary studies carried out on adult patients who are COVID-19-positive; (2) reporting of a model for predicting mortality with a reported area under the receiving operator curve (AUROC) value; and (3) routine blood or imaging parameters with mortality as the main outcome of interest. The established definition of AUROC applied to the context of a COVID-19 mortality prediction model was used; the accuracy of the model was used to discriminate the mortality risk levels in patients with COVID-19.

Exclusion criteria were non-English studies, sample size <100 patients and non-peer-reviewed publications. Any disagreements during screening were resolved by consensus. Mortality, for this review, is defined as death within 90 days of hospital admission due to COVID-19.

A data extraction form was generated to synthesise the following information: study title, method of calculation of the model or examined parameters (eg, statistical modelling or analysis, AI), scoring system versus analysis of single parameters, ‘summary of included parameters and AUROC for scoring systems’, ‘name and category of parameter (eg, biomarker)’ for single parameters and any additional salient findings.

RESULTS

After deduplication of original search results, title and abstracts of 589 studies were screened for relevance, and subsequently full-text articles were obtained and further assessed for eligibility. In all, 76 studies were identified that would inform our review.

Adapted current scoring systems

The sudden arrival of the pandemic has necessitated the application of existing prognostic systems to triage the influx of patients with COVID-19 to optimise distribution of limited resources and treatment. The accuracy of scoring systems adapted for COVID-19 mortality is detailed in online supplemental table 1 and then analysed to explore potential reasons for their differing predictive ability of mortality in patients with COVID-19.

Scoring systems are listed in order of descending AUROC values, as methodical differences between studies deem it inappropriate to merge AUROC results. For example, the Quick Sequential Organ Function Assessment (qSOFA) AUROC values ranged from 0.6200 to 0.8860 (online supplemental table 1), possibly due to different cut-off points. In addition, mortality was measured by 72 hours in some studies and up to 90 days in others, and sample sizes ranged from 105 to 864 across studies (online supplemental table 1).

The Acute Physiology and Chronic Health Evaluation II (APACHE II) score was found to have the highest AUROC values, followed by Modified Elixhauser Index (mEI) and Sequential Organ Function Assessment (SOFA) systems. APACHE II presides over other scores in terms of mortality prediction possibly due to its consideration of both age and comorbidities, whereas scores such as CURB-65 only assesses age and SOFA involves neither. Notably, however, the cut-off value for APACHE II is much lower when applied to patients with COVID-19 than under normal intensive care unit (ICU) conditions; while Glasgow Coma Scale (GCS) is an important component of APACHE II, the nervous system is typically less impacted than the respiratory system in COVID-19 infection.

COVID-19 scoring systems

Prediction scores play a vital role in guiding clinical decision-making for hospitalised patients with COVID-19. Online supplemental table 2 summarises recently developed scores and their AUROC values.

Different risk stratification tools use a variety of parameters to predict mortality. Online supplemental table 3 summarises the most common parameters used in novel COVID-19 mortality prediction scores. The two parameters associated with high predictive performance (higher AUROC) were lymphocyte count and D-dimer, with age being the most consistently used parameter. The most common parameter used in novel prediction models for mortality of patients with COVID-19 is age, followed by lymphocyte count, D-dimer, oxygen saturation, C reactive protein (CRP) and platelet count. Other less common parameters include respiratory rate (RR), lactate dehydrogenase, neutrophil-to-lymphocyte ratio (NLR), procalcitonin (PCT) and blood urea nitrogen.

The most common comorbidities for predicting mortality are hypertension (HTN), diabetes mellitus
(DM), obesity, cardiovascular disease, chronic kidney disease, smoking and malignancy.

**Single parameters**

COVID-19 has a different clinical picture to pneumonia and influenza, providing an avenue to explore what routinely available clinical information best predicts mortality. We explored blood parameters and imaging not currently extensively implemented into existing COVID-19 mortality prediction models, which are represented in online supplemental table 4.

Studies examining the associations of a range of laboratory biochemical tests and imaging at admission with mortality for patients with COVID-19 are extensive in the literature. Continued rapid identification of biomarkers that can accurately predict the likelihood of mortality is essential and has been proposed, including inflammatory, coagulation, renal, liver and cardiac biomarkers (online supplemental table 4).

Imaging, particularly chest CT scans, has been studied, with all three studies reporting independent associations with mortality, shown in online supplemental table 5. Alongside prognostic scores developed to assess risk of death, these must be updated to reflect the identification of imaging modalities that may need to be added or replace parameters in existing scores.

**AI in predicting mortality**

Machine learning (ML) is a subset of AI allowing systems to automatically improve based on new experiences. Online supplemental table 6 illustrates an overview of studies that used ML to predict mortality in patients with COVID-19.

Papers that used ML models have an AUROC greater than 0.8, conveying good discrimination of patients with high mortality risk. Models with a greater number of incorporated parameters did not find improvements in AUROC score. One model by Yuan et al. had a high AUROC of 0.9551 when looking at three parameters, while the model by Vaid et al. had a lower AUROC of 0.8400 when looking at 73 different parameters. This suggests that the total number of parameters was a less important factor than the interaction between the parameters in predicting mortality.

Deep learning (DL) is a subset of ML which uses algorithms to analyse multiple factors simultaneously; therefore, it would be more appropriate to handle multiple parameters. Online supplemental table 7 illustrates an overview of the studies that used ML to predict mortality in patients with COVID-19.

There are fewer studies assessing DL models, but similar to ML, these studies possess an AUROC >0.8.

**DISCUSSION**

**Adapted current scoring systems**

The variables used within existing scoring systems featured in online supplemental table 1 were analysed to explore potential reasons for their differing predictive ability of mortality in patients with COVID-19.

The APACHE II score was found to have the highest AUROC values, followed by mEI and SOFA systems. APACHE II presides over other scores in terms of mortality prediction possibly due to its consideration of both age and comorbidities, whereas scores such as CURB-65 only assesses age and SOFA involves neither. Notably, however, the cut-off value for APACHE II is much lower when applied to patients with COVID-19 than under normal ICU conditions; while GCS is an important component of APACHE II, the nervous system is typically less impacted than the respiratory system in COVID-19 infection.

Considering the effects of COVID-19 on respiratory function are more marked than its cardiovascular impacts, it is unsurprising that most of the studies listed in online supplemental table 1 show respiratory parameters such as RR in CURB-65 to be independently more indicative of mortality than blood pressure and confusion, which are more related to haemodynamics. qSOFA’s focus on blood pressure and mental state may explain its lower AUROC and poorer predictive performance. Cetinkal et al., however, argue that as previous studies reveal worse clinical outcomes in patients with cardiac injury, non-respiratory variables in the CHA2D2-VASc system such as older age, DM, HTN and previous cardiovascular disease are valuable parameters for mortality risk stratification. However, AUROC values found for CHA2D2-VASc remain at the low end compared with other existing scoring systems, despite modifications catered to COVID-19 added to form the m-CHA2D2-VASc scale. Even this version, with an AUROC higher by 0.06, offers predictive ability similar to univariate NLR and inferior to troponin increase.

Ortiz et al. demonstrated A-DROP, a modified version of CURB-65, to provide more accurate mortality prediction than Pneumonia Severity Index (PSI), CURB-65, CRB-65, SMART-COP, qSOFA and National Early Warning Score 2 (NEWS2). Its superior discrimination may be due to its more accurate respiratory function evaluation (oxygen saturation [SpO2] >90% / arterial oxygen tension [PaO2] <60 mm Hg in A-DROP vs respiratory rate ≥30/min in CURB-65). The modified age cut-off (male >70 / female >75 in A-DROP vs age >65 in CURB-65) may also contribute to A-DROP’s advantage when applied to COVID-19, considering the median age of COVID-19 non-survivors is 69 years.

Ultimately, although APACHE II, SOFA, PSI and CURB-65 are well-founded in clinical practice, their requirement for sophisticated patient information makes rapid assessment impossible, an important benefit for triaging patients with COVID-19 in often overrun hospitals. Wang et al.’s study on MEWS suggests this system can overcome the issue of efficiency as a simple and rapid assessment able to be performed within minutes of patient admission while maintaining equal predictive ability.
Intriguingly, Gupta et al\textsuperscript{15} evaluated 22 prognostic models (including aforementioned systems), concluding that they should not be recommended for routine clinical implementation because none of them offered incremental value compared with univariable predictors to risk stratify COVID-19 mortality, of which patient’s age is a strong predictor of mortality. Similarly, Bradley et al\textsuperscript{16} concluded that CURB-65, NEWS2 and qSOFA all underestimate the mortality of patients with COVID-19.

**COVID-19 scoring systems**

To maximise the accuracy and effectiveness of mortality prediction models, novel scores should focus on identifying features that are COVID-19-specific. Examples of complications that are highly associated with COVID-19 include hypercoagulability and inflammation.\textsuperscript{17,18} However, only 27% of new prognostic scores included in this review incorporated CRP—an important inflammatory marker. Similarly, thrombopenia has been associated with higher rates of mortality,\textsuperscript{19} which reflects the importance of including platelet count in prognostic models, but only 16% of new scores took this into account.

Interestingly, the three prediction models with the highest AUROC values have all used D-dimer and lymphocyte count to predict mortality. This could reflect the importance of these two parameters in COVID-19 pathophysiology. However, these are all single-centre studies tested on significantly smaller sample sizes compared with other models with lower AUROC values. Models tested on a larger population, for instance, Mancilla-Galindo et al\textsuperscript{18} national cohort study with a sample size of 83779 (AUROC=0.8000), could be more representative and generalisable.

The most common parameter used in novel prediction models for mortality of patients with COVID-19 is age, followed by lymphocyte count, D-dimer, oxygen saturation, CRP and platelet count. Other less common parameters include RR, lactate dehydrogenase, NLR, PCT and blood urea nitrogen.

Fumagalli et al\textsuperscript{26} report age as the strongest predictor of severe outcomes and mortality. Similarly, Mei et al\textsuperscript{20,21} prognostic model included age as one of five indicators of mortality and reports a strong association between advanced age and death from COVID-19.

There seems to be no association between the number of parameters and the prognostic power and accuracy of a scoring system. Several mortality prediction models with a small number of parameters have had higher AUROC values, for example, Liu et al\textsuperscript{22} had an AUROC value of 0.9940 with only three variables compared with Mancilla-Galindo et al\textsuperscript{48} (COVID-GRAM) with an AUROC value of 0.7750 and 10 parameters.

The most common comorbidities for predicting mortality are HTN, DM, obesity, cardiovascular disease, chronic kidney disease, smoking and malignancy.

**Single parameters**

COVID-19 has a different clinical picture to pneumonia and influenza, providing an avenue to explore what routinely available clinical information best predicts mortality. We explored blood parameters not currently extensively implemented into existing COVID-19 mortality prediction models, which are represented in online supplemental table 4.

We discuss the feasibility of introducing the below blood tests and imaging modalities into routine practice for risk stratification of patients with COVID-19.

**Cardiac biomarkers**

Cardiac biomarkers were the most common parameters studied in our literature search. High-sensitivity cardiac troponins have been shown to be independently associated with all-cause mortality in patients with COVID-19 (p<0.05), after accounting for age, sex and comorbidities, shown in online supplemental table 4. High-sensitivity cardiac troponins (hs-cTnI and hs-cTnT) are markers of myocardial injury that are currently primarily used in the prognostication of acute coronary syndrome. Despite evidence that 50% with confirmed COVID-19 have elevated cardiac biomarkers at the time of hospital admission, the patient sample sizes are limited in current studies to less than 500 patients and single centres.\textsuperscript{22} Cao et al\textsuperscript{23} retrospectively observed 244 patients and incorporated hs-cTnI into a model of empirical prognostic factors. A proposed cut-off (>20 ng/L serum hs-cTnI levels) yielded an AUROC increase from 0.65 to 0.71 (p<0.01) and demonstrated feasibility of this parameter to increase predictive performance.\textsuperscript{24}

**Inflammatory biomarkers**

Liu et al\textsuperscript{25} confirmed the independent association of PCT with mortality in a cohort of 1525 patients through retrospective analysis. Due to the large cohort and continued follow-up of PCT levels throughout hospital stay, this study provides stronger evidence for the inclusion of PCT into scoring systems, which has begun to be implemented but is still in the minority of included parameters. Fois et al\textsuperscript{26} used the same study design and identified the systemic inflammation index (SII) as an independent predictor of mortality. However, the study quality was poor—with only 119 patients and the large number of different inflammation indexes being studied in different combinations. It is unclear whether any clinical utility is offered by implementation of SII, considering deranged lymphocyte count is already widely established as a useful predictor.\textsuperscript{26}

**Renal and hepatic function biomarkers**

Esposito et al\textsuperscript{27} identified estimated glomerular filtration rate (using a baseline of 60 mL/min/1.73 m\textsuperscript{2}), and Fu et al\textsuperscript{28} identified cholestasis and hypoprothrombinemia as independent predictors of mortality. Interestingly, as with cardiac biomarkers, these were predictors even after accounting for pre-existing comorbidities. The obvious benefit to clinical practice of renal and hepatic function markers is that they

---

\textsuperscript{1} Open access

---
are routinely done on hospital admission and straightforward to clinicians to score in a system. Replication of large-scale multicentre studies is needed before determining the diagnostic validity of such parameters in the stratification of patients with COVID-19 in a statistical or AI model. It must be acknowledged that additional parameters must be externally validated to determine AUROC values and appropriate cut-offs for parameters.6

Lung imaging

Trabulus et al.,28 Francone et al.80 and Xu et al.80 examined the relationship between chest CT findings and mortality, with all three studies reporting independent associations with mortality (p<0.05). Two studies31 32 used a methodology involving an overall severity score of each scan and proposed defined cut-offs above which there was yield of best predictive value. These cut-offs are of value for clinicians to allocate scans with a high/medium/low rating which can be used to triage patients with COVID-19. However, both these studies have limitations in their methodology and design, which need to be addressed before implementation of CT severity into scoring systems. In the study by Gao et al.31 follow-up was limited to 24 days; a minimum of at least 28-day mortality is recommended to better reflect the clinical course of COVID-19 in most cases.7 In addition, both severity score studies were retrospective in nature, which is susceptible to incomplete clinical records and bias in the interpretation of CT by different radiologists. Chest CT while highly sensitive is not a first-line test due to limited resources to CT scan in all COVID-19-positive hospital admissions. Routine implementation of admission CT scans would also carry a radiation burden to patients, which is arguably unnecessary if alternative parameters conferring equal predictive power without additional risk of iatrogenic effects could be used. Perhaps, chest CT is more appropriate in the discharge process of clinically stable, triaged patients with COVID-19 rather than as a first-line test as part of an admission scoring system.

AI in predicting mortality

Between ML and DL models, it is unclear which branch of AI modelling would be superior in predicting mortality due to the similar AUROC values. These similar values can be accounted for by limitations in the study methods. Within all AI modelling papers, Meng et al.83 and Vaid et al.10 were the only studies that conducted external validation. External validation is an important step to verify the effectiveness of the model in patient population. Internal validation would use the same cohort to test the model, which can lead to overfitting and an inaccurately high AUROC. The models created by Bertsimas et al.84 Gao et al.31 and Meng et al55 gathered training set data from multiple centres, whereas the other models used single-centre data. Therefore, these models would increase applicability to the general population.

As COVID-19 has only been prevalent for a year, not many models have had the chance to be prospectively tested. Vaid et al10 produced the only model that was prospectively tested. This is important as it demonstrates the model’s real-world performance. Many models with a large number of incorporated parameters included patients with missing values, leading to estimation. This may be useful in clinical practice as not all patients have every test carried out.

It is important to recognise that COVID-19 management and treatment guidelines are constantly being updated, which influences mortality rates. As AI models use dynamic data,10 reporting of model AUROC in earlier stages of the pandemic may not have been as accurate.

Limitations

There are inherent limitations to this review. Most studies included were single centre and retrospective, whereas multicentre, prospective research may provide more insight. Although AUROC scores are universally accepted outcome measures of the accuracy of prediction models,6 they are limited in their clinical interpretability as they lack a direct link to individual patient outcomes. Thus, future reviews could use additional performance metrics in addition to AUROC to assess the accuracy of different models.

CONCLUSION

The above systems and parameters have been evaluated for their ability to stratify patients with COVID-19 by mortality risk, with predictive ability depicted as AUROC scores. New scoring systems developed specifically for the pandemic demonstrated higher AUROC scores than currently existing scoring systems adapted for COVID-19. However, the predictive strength of AI systems was not notably higher than pandemic-specific scoring systems, potentially due to time restraints of development and incomplete refining of algorithms. Single parameters extracted from scoring systems, novel biomarkers and imaging modalities were also explored for the ability to predict mortality and potential incorporation into novel risk stratification systems.

As most studies in the current literature were retrospective, we propose further prospective, multicentre studies to validate these variables’ diagnostic accuracy and multivariate relationships, which may impact their compounded efficacy for COVID-19 mortality prediction. A meta-analysis would address the limitation of the current review of not being able to directly compare and statistically manipulate AUROC scores found in the literature due to differing cut-off points, study sample sizes and mortality periods used by different studies.

In all, refining strategies to triage patients with COVID-19 can bring immense value to healthcare professionals in their clinical decisions concerning optimal treatment for patients with varying mortality risks and allocating scarce resources effectively.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.


BMJ Health Care Inform: first published as 10.1136/bmjhci-2021-100389 on 14 September 2021. Downloaded from http://informatics.bmj.com/ on December 15, 2021 by guest. Protected by copyright.
Competition interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD

Kelly Chu https://orcid.org/0000-0003-0527-373X

REFERENCES


Table 1 – Predictive ability of Existing Scoring Systems for Covid-19 in-hospital mortality, sorted by AUROC (n=40)

<table>
<thead>
<tr>
<th>Authors of paper evaluating system</th>
<th>Prognostic System</th>
<th>Intended application</th>
<th>Variables used</th>
<th>Number of variables included</th>
<th>AUROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang, L et al.</td>
<td>APACHE II (acute physiology and chronic health evaluation II)</td>
<td>Scoring system for use in intensive care units to provide gross estimate of mortality risks</td>
<td>APS (Acute physiology score), Temp, MABP, HR, RR, PaO2 or aPO2, arterial pH or HCO3, serum Na, serum K, serum creatinine, Hct, WBCC, GCS, age, chronic health evaluation</td>
<td>12</td>
<td>0.9370</td>
</tr>
<tr>
<td>Zou, X et al.</td>
<td>mEEI (modified Elixhauser Index)</td>
<td>Specific development of a modified Elixhauser Index</td>
<td>Age, sex, presence of renal diseases, neurological dis-orders, lymphoma, solid tumour with metastasis, IHD, CHD, coagulopathy, fluid and electrolyte disorders, liver disease, weight loss, metastatic cancer</td>
<td>13</td>
<td>0.9660</td>
</tr>
<tr>
<td>De Giorgi, A et al.</td>
<td>SOFA (sequential organ function assessment)</td>
<td>Score for calculation of number and severity of organ dysfunction in six organ systems (respiratory, coagulatory, liver, cardiovascular, renal, and neurologic)</td>
<td>Oxygenation index, MABP, GCS, creatinine or urine volume, bilirubin, platelets</td>
<td>6</td>
<td>0.918</td>
</tr>
<tr>
<td>S. Liu et al.</td>
<td>RAS (respiratory assessment scoring)</td>
<td>Assessment for progression and mortality in respiratory disease</td>
<td>RR, resting SpO2, Alveolar-arterial O2 gradient, Minimal exercise desaturation test</td>
<td>4</td>
<td>0.9150</td>
</tr>
<tr>
<td>Wang, L et al.</td>
<td>PSI (pneumonia severity index)</td>
<td>Index to identify CAP patients at a low risk of mortality who could safely be treated as outpatients</td>
<td>Age, sex, residence, comorbidity and acute pneumonia-associated morbidity</td>
<td>20</td>
<td>0.9270</td>
</tr>
<tr>
<td>Zou, X et al.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8500</td>
</tr>
<tr>
<td>D. Ji et al.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9100</td>
</tr>
<tr>
<td>C. Satci et al.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: SBP = Systolic blood pressure, RR = respiratory rate, DM = diabetes mellitus, LoC = level of consciousness, GCS = Glasgow Coma Scale, HR = heart rate, HTN = hypertension, TIA = transient ischaemic attack, MABP = mean arterial blood pressure, IHD = ischaemic heart disease, COPD = chronic obstructive pulmonary disease, HCT = haematocrit, SpO2 = oxygen saturation, AF = atrial fibrillation, CVD = cardiovascular disease, CHF = congestive heart failure, CAP = community acquired pneumonia
**Table 1 cont. part 2**

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Supplemental Material</th>
<th>Description</th>
<th>Parameters</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>F. Liu et al.⁹</td>
<td>NEWS (National Early Warning Score)</td>
<td>Tool for early detection of in-hospital patient deterioration</td>
<td>RR, SpO₂, Supplemental oxygen, SBP, temperature, HR, AVPU score</td>
<td>7</td>
</tr>
<tr>
<td>X. Tang et al.⁹</td>
<td>A-DROP</td>
<td>Modified CURB-65 system for prediction of mortality in hospitalized patients with CAP</td>
<td>Age, Dehydration, SpO₂ or MABP, Confusion, SBP</td>
<td>5</td>
</tr>
<tr>
<td>C. Satici et al.⁹</td>
<td>CCI (Charlson Comorbidity Index)</td>
<td>Predicts survival in patients with multiple comorbidities, and is widely used as a measure of total comorbidity burden</td>
<td>Age and Comorbidities (MI, CHF, peripheral vascular disease, cerebrovascular disease, dementia, COPD, peptic ulcer disease, liver disease, DM, hemiplegia, moderate to severe CKD, solid tumor, leukaemia, lymphoma, AIDS)</td>
<td>2</td>
</tr>
<tr>
<td>H. Hu et al.⁹</td>
<td>REMS (Rapid emergency medicine score)</td>
<td>Predicts in-hospital mortality in non-surgical patients admitted to the ED</td>
<td>HR, BP, RR, GCS, SpO₂, age</td>
<td>6</td>
</tr>
<tr>
<td>H. Hu et al.⁹</td>
<td>MEWS</td>
<td>Tool for assessment and early identification of pneumonia deterioration</td>
<td>HR, SBP, RR, body temperature, LoC</td>
<td>5</td>
</tr>
<tr>
<td>P. Bradley et al.¹⁰</td>
<td>CURB-65</td>
<td>Scoring system specific for CAP to predict all-cause mortality within 30 days</td>
<td>Confusion, Urea, RR, BP, Age ≥65</td>
<td>5</td>
</tr>
<tr>
<td>D. Ortiz et al.¹¹</td>
<td></td>
<td></td>
<td></td>
<td>0.7200</td>
</tr>
<tr>
<td>X. Tang et al.⁹</td>
<td></td>
<td></td>
<td></td>
<td>0.8500</td>
</tr>
<tr>
<td>Zou, X et al.²</td>
<td></td>
<td></td>
<td></td>
<td>0.8440</td>
</tr>
<tr>
<td>R. Gupta et al.¹²</td>
<td></td>
<td></td>
<td></td>
<td>0.7500</td>
</tr>
<tr>
<td>F. Liu et al.¹⁶</td>
<td></td>
<td></td>
<td></td>
<td>0.7665</td>
</tr>
<tr>
<td>C. Satici et al.¹³</td>
<td></td>
<td></td>
<td></td>
<td>0.8800</td>
</tr>
<tr>
<td>P. Bradley et al.¹⁰</td>
<td>NEWS2 (National Early Warning Score 2)</td>
<td>Disease agnostic early warning tool used to trigger escalation of care in the deteriorating patient, with high scores being associated with death or unanticipated intensive care unit (ICU) admission within 24 hours</td>
<td>RR, SpO₂, air or oxygen, systolic BP, HR, LoC, temperature</td>
<td>9</td>
</tr>
<tr>
<td>X. Tang et al.⁹</td>
<td></td>
<td></td>
<td></td>
<td>0.8100</td>
</tr>
<tr>
<td>F. Liu et al.⁹</td>
<td></td>
<td></td>
<td></td>
<td>0.8797</td>
</tr>
<tr>
<td>Table 1 cont. part 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P. Bradley et al</strong>&lt;sup&gt;10&lt;/sup&gt;</td>
<td><strong>qSOFA</strong> (quick sequential organ function assessment)</td>
<td>Tool for predicting mortality and ICU admission among patients with suspected infection in prehospital, emergency department and ward settings</td>
<td>Mental status, RR &lt;22, SBP &lt;100</td>
<td>3</td>
</tr>
<tr>
<td>S. Liu et al&lt;sup&gt;11&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang, L et al&lt;sup&gt;12&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zou, X et al&lt;sup&gt;13&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. Liu et al&lt;sup&gt;14&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A. Halalau et al</strong>&lt;sup&gt;15&lt;/sup&gt;</td>
<td><strong>m-CHA&lt;sub&gt;2&lt;/sub&gt;D&lt;sub&gt;3&lt;/sub&gt;VASC</strong> (Modified CHA&lt;sub&gt;2&lt;/sub&gt;D&lt;sub&gt;3&lt;/sub&gt;VASC)</td>
<td>Risk score created from CHA&lt;sub&gt;2&lt;/sub&gt;D&lt;sub&gt;3&lt;/sub&gt;VASC to improve predictive ability for COVID-19 mortality</td>
<td>Same as CHA&lt;sub&gt;2&lt;/sub&gt;D&lt;sub&gt;3&lt;/sub&gt;VASC but with gender criteria switched from female to male (male sex is reported by recent studies to be an important predictor of mortality in COVID-19 patients)</td>
<td>8</td>
</tr>
<tr>
<td><strong>G. Cetinkal et al</strong>&lt;sup&gt;16&lt;/sup&gt;</td>
<td><strong>CHA&lt;sub&gt;2&lt;/sub&gt;D&lt;sub&gt;3&lt;/sub&gt;VASC</strong></td>
<td>Risk score principally used for estimating the risk of ischemic stroke in patients with AF and also predicts mortality in various CVD</td>
<td>CHF, HTN, Age (65 to 74), DM, Vascular disease, Female gender, 2 points for age ≥75 and history of TIA and/or stroke</td>
<td>7</td>
</tr>
<tr>
<td><strong>D. Ortiz et al</strong>&lt;sup&gt;17&lt;/sup&gt;</td>
<td><strong>SMART-COP</strong></td>
<td>Assessing severity of CAP (community acquired pneumonia) confirmed by CXR</td>
<td>SBP, multilobar CXR involvement, Albumin, RR, HR, Confusion, SpO2, pH</td>
<td>8</td>
</tr>
</tbody>
</table>
### Table 2: New COVID-19 in-hospital mortality prediction models, displaying the number of patients and number of incorporated parameters, sorted by AUROC (n=37)

<table>
<thead>
<tr>
<th>Author with scoring system</th>
<th>AUROC</th>
<th>Number of patients</th>
<th>Number of incorporated parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liu et al.</td>
<td>0.9940</td>
<td>336</td>
<td>3</td>
</tr>
<tr>
<td>Qin et al.</td>
<td>0.9920</td>
<td>118</td>
<td>4</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>0.9905</td>
<td>126</td>
<td>7</td>
</tr>
<tr>
<td>Weng et al.</td>
<td>0.9750</td>
<td>301</td>
<td>5</td>
</tr>
<tr>
<td>Soto-Mota et al.</td>
<td>0.9600</td>
<td>400</td>
<td>7</td>
</tr>
<tr>
<td>Mei et al.</td>
<td>0.9600</td>
<td>1088</td>
<td>4</td>
</tr>
<tr>
<td>Luo et al.</td>
<td>0.9560</td>
<td>739</td>
<td>3</td>
</tr>
<tr>
<td>Luo et al.</td>
<td>0.9550</td>
<td>1115</td>
<td>9</td>
</tr>
<tr>
<td>Zhou et al.</td>
<td>0.9550</td>
<td>118</td>
<td>11</td>
</tr>
<tr>
<td>Cheng et al.</td>
<td>0.9400</td>
<td>305</td>
<td>8</td>
</tr>
<tr>
<td>Laguna-Goya et al.</td>
<td>0.9400</td>
<td>501</td>
<td>9</td>
</tr>
<tr>
<td>El-Sohl et al.</td>
<td>0.9200</td>
<td>1634</td>
<td>6</td>
</tr>
<tr>
<td>Shang et al.</td>
<td>0.9190</td>
<td>1830</td>
<td>5</td>
</tr>
<tr>
<td>Mei et al.</td>
<td>0.9120</td>
<td>492</td>
<td>6</td>
</tr>
<tr>
<td>El-Sohl et al.</td>
<td>0.9100</td>
<td>1634</td>
<td>6</td>
</tr>
<tr>
<td>Fumagalli et al.</td>
<td>0.9000</td>
<td>516</td>
<td>5</td>
</tr>
<tr>
<td>Gude-Sampedro et al.</td>
<td>0.8900</td>
<td>10545</td>
<td>6</td>
</tr>
<tr>
<td>Torres-Macho et al.</td>
<td>0.8830</td>
<td>1968</td>
<td>5</td>
</tr>
<tr>
<td>El-Sohl et al.</td>
<td>0.8800</td>
<td>1634</td>
<td>6</td>
</tr>
<tr>
<td>Li et al.</td>
<td>0.8700</td>
<td>1008</td>
<td>3</td>
</tr>
<tr>
<td>Hajifathalian et al.</td>
<td>0.8600</td>
<td>265</td>
<td>8</td>
</tr>
<tr>
<td>Manocha et al.</td>
<td>0.8340</td>
<td>1053</td>
<td>5</td>
</tr>
<tr>
<td>Hajifathalian et al.</td>
<td>0.8300</td>
<td>265</td>
<td>7</td>
</tr>
<tr>
<td>Tusha et al.</td>
<td>0.8130</td>
<td>163</td>
<td>8</td>
</tr>
<tr>
<td>Fernandez et al.</td>
<td>0.8129</td>
<td>487</td>
<td>10</td>
</tr>
<tr>
<td>Cai et al.</td>
<td>0.8070</td>
<td>126</td>
<td>5</td>
</tr>
<tr>
<td>Nunez-Gil et al.</td>
<td>0.8070</td>
<td>1021</td>
<td>10</td>
</tr>
<tr>
<td>Varol et al.</td>
<td>0.8020</td>
<td>383</td>
<td>6</td>
</tr>
<tr>
<td>Mancilla-Galindo et al.</td>
<td>0.8000</td>
<td>83779</td>
<td>3</td>
</tr>
<tr>
<td>Altschul et al.</td>
<td>0.7980</td>
<td>4711</td>
<td>3</td>
</tr>
<tr>
<td>Gue et al.</td>
<td>0.7933</td>
<td>316</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>AUC</td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>----------------</td>
<td>---------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Knight et al.</td>
<td>0.7900</td>
<td>35463</td>
<td>3</td>
</tr>
<tr>
<td>Lorente-Ros et al.</td>
<td>0.7900</td>
<td>707</td>
<td>7</td>
</tr>
<tr>
<td>Shi et al. (COVID-GRAM score)</td>
<td>0.7750</td>
<td>257</td>
<td>5</td>
</tr>
<tr>
<td>El-Solh et al. (Yu-score)</td>
<td>0.7700</td>
<td>1634</td>
<td>3</td>
</tr>
<tr>
<td>Rodriguez-Nava et al.</td>
<td>0.7110</td>
<td>313</td>
<td>10</td>
</tr>
<tr>
<td>Shi et al. (CALL-score)</td>
<td>0.6400</td>
<td>257</td>
<td>4</td>
</tr>
</tbody>
</table>
Table 3: Most common parameters incorporated into new prediction models of in-hospital mortality in patients with COVID-19 (n=37)

Abbreviations: CRP = C-reactive protein, SpO2 = Peripheral capillary oxygen saturation

<table>
<thead>
<tr>
<th>Author with scoring system</th>
<th>AUROC</th>
<th>Lymphocyte count</th>
<th>D-dimer</th>
<th>CRP</th>
<th>SpO2</th>
<th>Platelet count</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liu et al.</td>
<td>0.9940</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qin et al.</td>
<td>0.9920</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang et al.</td>
<td>0.9905</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weng et al.</td>
<td>0.9750</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soto-Mota et al.</td>
<td>0.9600</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Mei et al.</td>
<td>0.9600</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Luo et al.</td>
<td>0.9560</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Luo et al.</td>
<td>0.9550</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Zhou et al.</td>
<td>0.9550</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheng et al.</td>
<td>0.9400</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laguna-Goya et al.</td>
<td>0.9400</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>El-Solh et al.</td>
<td>0.9200</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Shang et al.</td>
<td>0.9190</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Mei et al.</td>
<td>0.9120</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>El-Solh et al.</td>
<td>0.9100</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Fumagalli et al.</td>
<td>0.9000</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Gude-Sampedro et al.</td>
<td>0.8900</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Torres-Macho et al.</td>
<td>0.8830</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>El-Solh et al.</td>
<td>0.8800</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Li et al.</td>
<td>0.8700</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Hajifathalian et al.</td>
<td>0.8600</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Manoche et al.</td>
<td>0.8340</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Hajifathalian et al.</td>
<td>0.8300</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Tusha et al.</td>
<td>0.8130</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Fernandez et al.</td>
<td>0.8129</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Cai et al.</td>
<td>0.8070</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nunez-Gil et al.</td>
<td>0.8070</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Varol et al.</td>
<td>0.8020</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mancilla-Galindo et al.</td>
<td>0.8000</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altschul et al.</td>
<td>0.7980</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Study</td>
<td>Score</td>
<td>AIC</td>
<td>BIC</td>
<td>Cross-Validation</td>
<td>ROC Curves</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------</td>
<td>-----</td>
<td>-----</td>
<td>-----------------</td>
<td>------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gue et al. 42</td>
<td>0.7933</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knight et al. 43</td>
<td>0.7900</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lorente-Ros et al. 44</td>
<td>0.7900</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shi et al. 45 (COVID-GRAM score)</td>
<td>0.7775</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>El-Soh et al. 26 (Yu-score)</td>
<td>0.7700</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rodriguez-Nava et al. 46</td>
<td>0.7110</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shi et al. 45 (CALL-score)</td>
<td>0.6400</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4: studies examining the association of novel blood parameters with mortality in patients admitted to hospital with COVID-19 (n=12)

<table>
<thead>
<tr>
<th>Authors</th>
<th>Sample size</th>
<th>Category of parameters</th>
<th>Blood parameter(s) shown to be associated with mortality (p&lt;0.05)</th>
<th>Proposed cut-offs for independently associated parameters</th>
<th>Sensitivity %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fu et al.</td>
<td>355</td>
<td>Hepatic function markers</td>
<td>Cholestasis markers (ALP, γ-GGT and TBA) Hypoproteinaemia markers (albumin and globulin)</td>
<td>Outside of normal range</td>
<td>Not reported</td>
</tr>
<tr>
<td>García et al.</td>
<td>639</td>
<td>arterial blood gas analyses, and laboratory values such as inflammatory, coagulation, renal, liver, cardiac</td>
<td>creatinine, D-dimer, lactate, potassium, P/F-ratio, alveolar-arterial gradient</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Bellmann-Weiler et al.</td>
<td>259</td>
<td>Presence of anaemia subgroups (mild/moderate/severe)</td>
<td>Presence of moderate-severe anaemia</td>
<td>moderate-severe anaemia; defined as haemoglobin &lt;109 g/L</td>
<td>Not reported</td>
</tr>
<tr>
<td>Liu et al.</td>
<td>1525</td>
<td>Inflammatory biomarker</td>
<td>Procalcitonin</td>
<td>PCT≥0.05 ng/ml</td>
<td>Not reported</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>605</td>
<td>Admission fasting blood glucose</td>
<td>FBG</td>
<td>FBG ≥7.0 mmol/l</td>
<td>Not reported</td>
</tr>
<tr>
<td>Singh et al.</td>
<td>276</td>
<td>Cardiac biomarker</td>
<td>Elevated initial high sensitivity cardiac troponin-T (hs-TnT)</td>
<td>- initial hs-TnT above the median (≥17 ng/L)</td>
<td>Not reported</td>
</tr>
<tr>
<td>Fois et al.</td>
<td>119</td>
<td>complete blood cell count (CBC)-derived inflammation indexes</td>
<td>Systemic inflammation index (SSI)</td>
<td>&gt;1835 ×10^9 cells/L</td>
<td>SSI-55</td>
</tr>
<tr>
<td>Aloisio et al.</td>
<td>427</td>
<td>Range of serum biomarkers</td>
<td>Lactate dehydrogenase Albumin</td>
<td>lactate dehydrogenase: &gt;731 U/L , albumin: 18 g/L or lower</td>
<td>Not reported</td>
</tr>
<tr>
<td>Foy et al.</td>
<td>1641</td>
<td>Complete blood count (CBC) derived parameter</td>
<td>red blood cell distribution width (RDW)</td>
<td>elevated RDW was defined as greater than 14.5%</td>
<td>Not reported</td>
</tr>
<tr>
<td>Stefanini et al.</td>
<td>397</td>
<td>Cardiac biomarkers</td>
<td>high-sensitivity cardiac troponin I (hs-TnI), B-type natriuretic peptide (BNP)</td>
<td>≥19.6 ng/L, BNP ≥100 pg/mL hs-TnI serum levels</td>
<td>Not reported</td>
</tr>
<tr>
<td>Trabulus et al.</td>
<td>336</td>
<td>Kidney function biomarker</td>
<td>eGFR</td>
<td>eGFR under 60 mL/min/1.73m²</td>
<td>Not reported</td>
</tr>
<tr>
<td>Cao et al.</td>
<td>244</td>
<td>Cardiac biomarkers</td>
<td>serum high-sensitivity cardiac Troponin I (hs-cTnI)</td>
<td>&gt;20ng/L serum hs-cTnI levels</td>
<td>hs-cTnI -85.7</td>
</tr>
</tbody>
</table>
**Table 5: studies examining the association of imaging modalities with mortality in patients admitted to hospital with COVID-19 (n=4)**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Sample size</th>
<th>Imaging modality and site</th>
<th>Imaging features(s) shown to be associated with mortality (p&lt;0.05)</th>
<th>Proposed cutoffs for independently associated parameters</th>
<th>Sensitivity %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Esposito et al. 59</td>
<td>1394</td>
<td>Chest CT</td>
<td>Enlarged main pulmonary artery diameter (MPAD)</td>
<td>Enlargement (≥ 31 mm)</td>
<td>Not reported</td>
</tr>
<tr>
<td>Francone et al. 50</td>
<td>130</td>
<td>Chest CT</td>
<td>CT-based semi-quantitative score of pulmonary lobar involvement (range 0-25)</td>
<td>CT score ≥ 18</td>
<td>Not reported</td>
</tr>
<tr>
<td>Lichter et al. 51</td>
<td>120</td>
<td>Lung ultrasound (LUS)</td>
<td>LUS severity score (range 0-36)</td>
<td>Baseline LUS score &gt; 18</td>
<td>LUS- 62</td>
</tr>
<tr>
<td>Xu et al. 62</td>
<td>703</td>
<td>Chest CT</td>
<td>CT severity score</td>
<td>CT severity score &gt; 14</td>
<td>Not reported</td>
</tr>
</tbody>
</table>
Table 6: Machine learning models that are used to predict mortality in patients admitted to hospital from COVID-19, with the training set being the number of people used to create the model and the test set being the number of people used when validating the model, sorted by highest AUROC. (n=12)

<table>
<thead>
<tr>
<th>Author</th>
<th>Number of incorporated parameters</th>
<th>AUROC</th>
<th>Training set</th>
<th>Test set</th>
<th>Parameters looked at</th>
</tr>
</thead>
<tbody>
<tr>
<td>An et al.</td>
<td>5</td>
<td>0.9630</td>
<td>7166</td>
<td>3071</td>
<td>age &gt; 80, taking of acarbose, age &gt; 70, taking of metformin, and underlying cancer</td>
</tr>
<tr>
<td>Yuan et al.</td>
<td>3</td>
<td>0.9551</td>
<td>1479</td>
<td>573</td>
<td>qSOFA, CURB 65, CRB65</td>
</tr>
<tr>
<td>Booth et al.</td>
<td>5</td>
<td>0.9300</td>
<td>318</td>
<td>80</td>
<td>CRP, blood urea nitrogen (BUN), serum calcium, serum albumin, and lactic acid.</td>
</tr>
<tr>
<td>Yu et al.</td>
<td>32</td>
<td>0.9220</td>
<td>172</td>
<td>74</td>
<td>lactate dehydrogenase, a-hydroxybutyrate dehydrogenase, bnp, urea nitrogen, hr cp I, myoglobin, age, d dimer, lymphocyte, cystatin c, igG, neutrophils, albumin, creatinine kinase isozyme, creatinine, % eosinophil RR, total platelet, x blood glucose, eosinophil count, platelet distribution width, average platelet volume, hsc reactive protien, alkaline phosphate, basophil count, thrombin time, x platelet hematocrit, temperature, lipoprotien a, hbeab, phosherous, aptt</td>
</tr>
<tr>
<td>Gao et al.</td>
<td>14</td>
<td>0.9186</td>
<td>2160</td>
<td>116</td>
<td>Consciousness, male sex, sputum, blood urea nitrogen [BUN], respiratory rate [RR], D-dimer, number of comorbidities, age, platelet count [PLT], fever, albumin [ALB], SpO2, lymphocyte, and chronic kidney disease</td>
</tr>
<tr>
<td>Bertsimas et al.</td>
<td>7</td>
<td>0.9019</td>
<td>2755</td>
<td>307</td>
<td>IL-2R, IL-6 , IL-8, TNF-α, B cells, CD4+ T cells, CD8+ T cells, NK cells</td>
</tr>
<tr>
<td>Abdulaal et al.</td>
<td>22</td>
<td>0.9012</td>
<td>398</td>
<td>40</td>
<td>Consciousness, male sex, sputum, blood urea nitrogen [BUN], respiratory rate [RR], D-dimer, number of comorbidities, age, platelet count [PLT], fever, albumin [ALB], SpO2, lymphocyte, and chronic kidney disease</td>
</tr>
<tr>
<td>Hu et al.</td>
<td>4</td>
<td>0.8810</td>
<td>183</td>
<td>64</td>
<td>Consciousness, male sex, sputum, blood urea nitrogen [BUN], respiratory rate [RR], D-dimer, number of comorbidities, age, platelet count [PLT], fever, albumin [ALB], SpO2, lymphocyte, and chronic kidney disease</td>
</tr>
<tr>
<td>Pan et al.</td>
<td>8</td>
<td>0.8600</td>
<td>98</td>
<td>25</td>
<td>LYM%, PT, lactate dehydrogenase (LDH), total bilirubin (T-Bil) , eosinophil percentage (EOS%), creatinine (Cr), NEUT%, and ALB.</td>
</tr>
</tbody>
</table>
### Table 6 cont:

<table>
<thead>
<tr>
<th>Study (Ref)</th>
<th>n</th>
<th>p-value</th>
<th>y</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parchure et al. [72]</td>
<td>55</td>
<td>0.8550</td>
<td>396</td>
<td>170</td>
</tr>
<tr>
<td>Vaid et al. [73]</td>
<td>73</td>
<td>0.8400</td>
<td>1514</td>
<td>2201</td>
</tr>
<tr>
<td>Yu et al. [66]</td>
<td>9</td>
<td>0.8060</td>
<td>172</td>
<td>74</td>
</tr>
</tbody>
</table>

- Age, Gender, Length of stay, Admission type, Admission Source, Respiratory rate, Pulse, Diastolic blood pressure, Percutaneous oxygen saturation, Systolic blood pressure, Temperature, Blood urea nitrogen, Serum Creatinine, Platelet count, Serum chloride, Anion gap, Serum sodium, Corrected WBC count, C-reactive protein, Red blood cell count, Partial pressure of carbon dioxide in arterial blood, (PACO2), Partial pressure of oxygen in arterial blood (PAO2), Partial pressure of carbon dioxide in venous blood (PVCO2), Partial pressure of oxygen in venous blood (PVO2), Serum potassium, Activated partial thromboplastin time, Serum lactate, pH of arterial blood, Serum total protein, Hemoglobin, Complement C3, Complement C4, Interleukin 1 beta, Interleukin 6, Interleukin 17, D-dimer, Aspartate aminotransferase, Alanine aminotransferase, Serum calcium, Serum ferritin, Lymphocyte count, Lactate dehydrogenase, Serum albumin, NT-pro hormone B-type natriuretic peptide, pH of venous blood, Bicarbonates by arterial blood gas analysis, Serum direct bilirubin, Serum total bilirubin, T wave axis, P wave axis, R wave axis, Atrial rate, Ventricular rate, PR interval, QRS duration.

- age, BNP, urea nitrogen, total platelet count, average platelet volume, D-dimer, high-sensitivity troponin I, LDH and creatinine kinase isoenzyme.
Table 7: Deep Learning models that are used to predict patient mortality in patients admitted to hospital from COVID-19, with the training set being the number of people used to create the model and the test set being the number of people used when validating the model, sorted by highest AUROC (n=3)

<table>
<thead>
<tr>
<th>Author</th>
<th>Number of incorporated parameters</th>
<th>AUROC</th>
<th>Training set</th>
<th>Test Set</th>
<th>Parameters looked at</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhu et al.</td>
<td>5</td>
<td>0.9540</td>
<td>187</td>
<td>33</td>
<td>D-dimer, oxygen index, neutrophil to lymphocyte ratio (NE:LY), C-reactive protein (CRP), and lactate dehydrogenase (LDH).</td>
</tr>
<tr>
<td>Meng et al.</td>
<td>5</td>
<td>0.9430</td>
<td>246</td>
<td>120</td>
<td>sex, age, severity grade, and with/without chronic disease and image features</td>
</tr>
<tr>
<td>Li et al.</td>
<td>6</td>
<td>0.8480</td>
<td>997</td>
<td>111</td>
<td>age, LDH, procalcitonin, troponin, CRP and SpO2</td>
</tr>
</tbody>
</table>

References


