Prediction Models for Severe Manifestations and Mortality due to COVID-19: A Rapid Systematic Review

Jamie L. Miller
MD Candidate
University of Iowa Carver College of Medicine

Masafumi Tada, MD
Department of Health Promotion and Human Behavior
Kyoto University Graduate School of Medicine / School of Public Health

Michihiko Goto MD MSCI
Division of Infectious Diseases
Department of Internal Medicine
University of Iowa Carver College of Medicine

Nicholas Mohr
Department of Emergency Medicine
Department of Anesthesia
Department of Epidemiology
University of Iowa Carver College of Medicine

Sangil Lee MD, MS
Clinical Associate Professor
Department of Emergency Medicine
The University of Iowa Carver College of Medicine

Acknowledgement

We thank the assistance to generate search strategy and identifying the literature from Jennifer Deberg, MLS, and Hans R House, MD, at the University of Iowa.

ABSTRACT

Background
Throughout 2020, the coronavirus disease 2019 (COVID-19) has become a threat to public health on national and global level. There has been an immediate need for research to understand
the clinical signs and symptoms of COVID-19 that can help predict deterioration including mechanical ventilation, organ support, and death. Studies thus far have addressed the epidemiology of the disease, common presentations, and susceptibility to acquisition and transmission of the virus; however, an accurate prognostic model for severe manifestations of COVID-19 is still needed because of the limited healthcare resources available.

Objective
This systematic review aims to evaluate published reports of prediction models for severe illnesses caused COVID-19.

Methods
Searches were developed by the primary author and a medical librarian using an iterative process of gathering and evaluating terms. Comprehensive strategies, including both index and keyword methods, were devised for PubMed and EMBASE. The data of confirmed COVID-19 patients from randomized control studies, cohort studies, and case-control studies published between January 2020 and July 2020 were retrieved. Studies were independently assessed for risk of bias and applicability using the Prediction Model Risk Of Bias Assessment Tool (PROBAST). We collected study type, setting, sample size, type of validation, and outcome including intubation, ventilation, any other type of organ support, or death. The combination of the prediction model, scoring system, performance of predictive models, and geographic locations were summarized.

Results
A primary review found 292 articles relevant based on title and abstract. After further review, 246 were excluded based on the defined inclusion and exclusion criteria. Forty-six articles were included in the qualitative analysis. Inter observer agreement on inclusion was 0.86 (95% confidence interval: 0.79 - 0.93). When the PROBAST tool was applied, 44 of the 46 articles were identified to have high or unclear risk of bias, or high or unclear concern for applicability. Two studied reported prediction models, 4C Mortality Score from hospital data and QCovid from general public data from UK, and were rated as low risk of bias and low concerns for applicability.

Conclusion
Several prognostic models are reported in the literature, but many of them had concerning risks of biases and applicability. For most of the studies, caution is needed before use, as many of
them will require external validation before dissemination. However, two articles were found to have low risk of bias and low applicability can be useful tools.

INTRODUCTION

Background
COVID-19 is the disease caused by SARS-CoV-2 that emerged in China in December 2019. Throughout 2020, COVID-19 has become an increasing threat to public health on national and global level. There has been an immediate need for research to help predict the clinical deterioration including death, mechanical ventilation, and organ support. A precise risk stratification of individuals with COVID-19 enables allocation of appropriate resources such as hospitalization, intensive care admission, ventilatory support, and antiviral therapy. As we start to understand the unique nature of this infection, an accurate risk stratification tool is urgently needed.

Studies thus far have addressed the epidemiology of the disease, common presentations, and susceptibility to disease acquisition and transmission of the virus. However, the short- and long-term implications of developing severe illnesses caused by COVID-19 is still widely unknown. Many studies with methodological shortcomings are published to facilitate the dissemination of rapidly evolving science of COVID-19. The quality of prediction models related to COVID-19 also suffered, as many studies did not have adequate validation steps after models were developed. The purpose of this review is to consolidate published data on validated predictive models of severe illnesses caused by COVID-19. We undertook a rapid systematic review following the format given by the King et al

Goals of this investigation
The goal of our review was to collectively evaluate published findings of prediction models for severe illness caused by COVID-19 to inform healthcare providers and the general population of the existing evidence in the midst of the pandemic.

METHODS
The review protocol was registered to PROSPERO (registration number CRD42020201484). Our study adheres to the Preferred Reporting Items for Systematic Reviews and Meta-analyses
guidelines for systematic reviews and was performed in accordance with best-practice guidelines (PRISMA).\textsuperscript{11}

Search strategies were developed with the assistance of a health sciences librarian with expertise in systematic reviews. Searches were developed by the primary author and librarian using an iterative process of gathering and evaluating terms. Searches were finalized in July 2020. Comprehensive strategies, including both index and keyword methods, were devised for PubMed and EMBASE. In order to maximize sensitivity, no pre-established database filters were used other than English language. We also screened references of articles identified by these systematic search strategies to look for any additional records potentially useful for the aim of this review. Lastly, we examined additional articles through a referral from investigators and online journal updates. The completed PubMed strategy is shown in Supplemental file 1. PubMed Strategy was then adapted for EMBASE and is available upon request. The last search was conducted on 7/27/2020.

We included studies of randomized controlled trials, cohort studies, and case-control studies that discuss the possible short-term and long-term consequences of contracting COVID-19 in any clinical setting. All studies were considered regardless of publication status (preprint or peer review manuscript), as long as those are included to PubMed or EMBASE. The definition of positive COVID-19 included a positive nasopharyngeal or oropharyngeal swab for COVID-19 by polymerase-chain reaction (PCR) or antigen tests, serological test for COVID-19, or those who were clinically diagnosed as COVID-19 based on clinical presentation and epidemiologic information. Outcomes included intubation, mechanical ventilation, any other type of organ support (for example, hemodialysis), and/or death. We included studies that developed and validated a multivariable model or scoring system, based on individual participant level data, to predict any COVID-19 related outcome or externally validated a known scoring system. We excluded review articles, case reports, editorials, and comments. Studies reporting the predictive ability of a single test, for example, a study only focused on the prognostic value of lactate dehydrogenase (LDH), were excluded from the review, since we were primarily interested in the combination of test results with any internal or external validation. Epidemiological studies that
aimed to model disease transmission or diagnostic test accuracy, and predictor finding studies were also excluded.

**Data Collection and Processing**

Titles, abstracts, and full texts were screened in duplicate for eligibility by independent reviewers (JM and SL), and discrepancies were resolved through discussion moderated by the third reviewer (MG). The reasons for any exclusions were recorded for those that required full article review.

Two investigators (JM, MT) independently extracted data from the included studies. The investigators underwent initial training and extracted data into a predesigned data collection form. The following information was abstracted: author, year, setting, outcome, predictors in final model, sample size for development set, type of validation, sample size, performance, and geographical region.

When data were missing or ambiguous, we contacted the authors for clarification. Studies were independently assessed for risk of bias by two investigators (JM, MT) using the Prediction Model Risk Of Bias Assessment Tool (PROBAST) tool. Any discrepancies were resolved by consensus.

**Primary Data Analysis**

A narrative synthesis of the characteristics of the included studies, including the outcomes, sample size for development and validation, type of case ascertainment, relationship between the final model, and the development of severe illnesses, were summarized in a table. We listed risk of bias and applicability using the PROBAST tool. To maximize the synthesis of results, we reported the studies based on the low risk of bias, low concern for applicability, and geographical location for future use.

**RESULTS**

**Study selection**
We included a total of 292 of 3399 articles based on the title and abstract. Then, we included a total of 42 out of 292 during the secondary full text screening up to July 2020. We identified an additional four articles from an investigator, online journal update, and references. The details of the selection process are listed in Figure 1. Overall, kappa was 0.86 (95% 0.79 - 0.93) between two reviewers.

**Characteristics of studies**

A total of 41 (89%) articles were identified from peer review journals and 5 (11%) were from preprints (Figure 1). The majority of studies (n=25, 53%) were published in China, 9 studies (19%) in the U.S. and 7 (15%) were published in Europe. (Table 1).

There were 38 (82%) retrospective cohort studies, and four (9%) prospective cohort studies, and four (9%) retrospective and prospective studies. The majority (n=25, 54%) were conducted at a single institution, 21 studies (46%) were conducted at two or more sites. There were two studies that were conducted in multiple nations. The total sample was 8,621,479 patients, and the mean sample size used for model derivation in the 46 articles was 187,423 (range 66-8,256,158).

**Critical appraisal of quality**

Next, using the PROBAST, we evaluated the risk of bias for each study and concern for applicability using the low, high, or unclear categories. The articles where the PROBAST tool did not apply were categorized as not applicable.

The PROBAST composite scores of the 46 articles, at the domain and item level, are presented in Figure 2. The overall risk of bias was low in four (9%) studies, high in 34 (74%) studies, and unclear in eight (17%) studies. The overall concern for applicability was low in 22 (48%) studies, high in 17 (37%) studies, and unclear in seven (15%) studies. (Figure 2)

The lowest risk of bias was seen in the participants domain (29 studies with low risks of bias, 63%), and highest risk of bias was seen in the analysis domain (32 studies with high risks of bias, 70%). The lowest concern for applicability was seen in the participants domain (38 studies with low concern, 83%), and the highest risk of concern for applicability was seen in the outcome.
domain (high concern, n=16, 35%). When the PROBAST evaluation was stratified to multi-site studies, there were increase in the overall applicability (Figure 3).

Two studies (4.3%) reported the prediction from images and seven (15%) studies reported the combinations of clinical data and images. These studies included chest X-ray and Computed Tomography (CT findings) either by itself or in conjunction with clinical data. The predictive ability of these models (AUC) reached 0.8 to 0.9, but most of these studies had a high or unclear risk of bias, which was similar to the recent review reported by Wynants et al.11

Overall, we identified two studies with a low risk of bias and low concern for applicability based on the PROBAST tool, which were Clift et al. and Knight et al. (Table 3)13,15

**Individual Study Characteristics**

Detailed characteristics of each of the 46 studies are presented in Table 1. Importantly, 5 (11%) of studies used external validation of the existing clinical prediction scores, and 41 (89%) of studies developed a new model and internally or externally validated it. Two studies (4.3%) reported the prediction from images and 7 (15%) reported the combinations of clinical data and images.

**Synthesis of Results**

Based on the PROBAST tools, we listed the studies with low risk of bias, low concerns for applicability, and predictive ability, metrics, setting, geographic location, and the link to an online calculator in Table 2.

Two studies were found to have low risk of bias and low concern for applicability (Table 2), and therefore warrant further description. Clift et al.15 developed a prediction model using various risk factors including age, sex, ethnicity, BMI, postal (zip) code, housing, and comorbidity. It was conducted in the general population in the UK during the first wave, and the absolute risk depends on the incidence in the UK at the time. An example shows a 55-year-old black African man with type 2 diabetes, a body mass index of 27.7, and no other risk factors. His absolute risk of catching and dying from covid-19 over the 90 day period is 0.1095% (or 1 in 913). His relative risk compared with a white man aged 55 years and a body mass index of 25 is 10.84.
The model shows that he is in the top 10% of the population at the highest risk. Knight et al.\textsuperscript{13} developed and validated a risk score based on common parameters that are available at hospital admission. This included factors such as age, sex, comorbidity, respiratory rate, SpO\textsubscript{2}, GSC, BUN, and CRP. The study was on the inpatient population, and study findings, particularly the probability of mortality, could easily be adapted for use outside the UK. The study reported the low, intermediate, high, and very high risk for mortality based on predictors. For example, patients in the intermediate-risk group (score 4-8, 21.9%) had a mortality rate of 9.9%. Patients in the high-risk group (score 9-14, n=11 664, 52.2%) had a mortality rate of 31.4%.\textsuperscript{13} However, one should account for differences in the criteria for admission and the quality of care (Table 2).

**DISCUSSION**

This systematic review identified a total of 46 articles that predicted severe COVID-19 in the literature. The majority of articles are from China, Europe, and North America where COVID-19 hit the hardest. These prediction models are focused on prognosticating patients with COVID-19. Overall, models reported the validation AUC from 0.54 to 0.98, but risk of bias was unclear or high for many studies, likely due to limited transparency on the predictor and outcome assessment, limited number of samples and outcomes. Many of these studies are retrospective studies and from early in the pandemic; most of them are limited due to the event rate and unclear case and outcomes definitions. Sample size was variable throughout studies, and lastly, analysis domain, such as overfitting were often seen as the limitation. These are commonly identified problems with building prediction models, and likely causes overfitting.\textsuperscript{16}

Although many of the studies are reported out of urgency, several of them did not seem to follow TRIPOD guidelines,\textsuperscript{17} which likely contributed to the overall risk of bias for the included studies. After review of the current literature, our impression is that many of the reported prediction models have high/unclear risk of bias, thus it is not recommended that any particular tool should be used until further external validation is completed, unless the risk of bias is low.

This review identified several implications for potential application of these prediction models in the appropriate clinical settings. As stated earlier, there are 41 studies that reported original models during the pandemic, and four studies\textsuperscript{24,32,41,45} that reported the external validation
models from the severity prediction models which existed before the COVID-19. Clift et al., Knight et al., reported the models that were developed from nationwide data which enabled robust re-sampling methods to minimize overfitting.\textsuperscript{13,15} We conclude that two of them\textsuperscript{13,15} can be useful tools (Table 3).

Clift et al\textsuperscript{15} reported population-based risk algorithm in the UK, showing high levels of discrimination for deaths and hospital admissions due to covid-19. The absolute risks presented, however, will change over time in line with the prevailing SARS-CoV-2 infection rate and the extent of social distancing measures in place, so they should be interpreted with caution. Because the incidence of COVID19 varies widely according to region and timing and therefore generalizability of the absolute risk is limited. The authors reported relative risk, which is less sensitive to incidence, so that it can be used in other countries. Knight et al.\textsuperscript{13} reported mortality estimates from hospitalized adults in the UK. The authors categorized the risk group into very low, low, intermediate, high, and very high for in-hospital mortality. This study took place when the first wave hit the UK, and to account for concerns regarding the level of care at the time of the pandemic, the authors employed temporal validation. The prediction model could be used to determine the level of care, such as discharge or hospitalization and escalation of care after hospitalization.

**Strength and Limitations**

Our review used a clear inclusion and exclusion criteria, which likely led to almost perfect agreement between two reviewers. Furthermore, the use of PROBAST strengthened our review, as this tool enabled accurate evaluation of risk of bias and applicability for each of the included studies.

There are several limitations. First, we used only two databases to identify the literature, which is less than we need for a traditional systematic review. We undertook this review to disseminate a rapidly progressing science of COVID-19 for healthcare providers. Second, our search was completed in July 2020, but this review required additional time to train reviewers to use the PROBAST tool to evaluate each article. Due to the rapid development of COVID-19 pandemic, it is important to note the gap between our search timeline and the current status at the time of
article publication. Third, variables and types of prediction models were highly variable among studies, and it was infeasible to conduct meta-analysis or meta-regression, as we specified in the protocol. Our aim was to systematically curate the current state of scientific knowledge to predict severe COVID-19, and we argue that our findings are still informative to healthcare providers without a pooling of results. Fourth, we were primarily interested in multivariable models, which may not have captured innovative studies focused on one predictive marker. Lastly, the volume of literature related to this topic area was a challenge, particularly for preprint articles. Due to concerns with feasibility, database search for preprints were limited to the records included in the mentioned databases.

CONCLUSION
Our review identified several prognostic models for COVID-19, and they showed various discriminative performances. The risk of bias was overall high or unclear for most of the studies, and the risk of bias was highest for the analysis domain. Concern for applicability was low for the majority of included studies. Thus, it is possible that most studies are prone to bias and require external validation before dissemination. The use of prediction models for severe COVID-19 cases requires caution since risk of bias is not negligible. We reported two studies that had a low risk of bias and low concern for applicability, one from a general public population and hospital setting in the UK.

Figure 1. PRISMA diagram
Figure 2. Risk of Bias and Concern for Applicability
Figure 3. Risk of Bias and Concern for Applicability Restricted to Multi-sites
Table 1. Data Summary
Table 2. A List of Studies with Low Risk of Bias and Low Concern for Applicability
REFERENCES


11. Alessandro Liberati, M., DrPH; Douglas G. Altman, DSc; Jennifer Tetzlaff, BSc; Cynthia Mulrow, MD, MSc; Peter C. Götzsche, MD, DrMedSci, MSc; John P.A. Ioannidis, MD; Mike Clarke, BA, DPhil; P.J. Devereaux, MD, BSc, PhD; Jos Kleijnen, MD, PhD; and David Moher, PhD. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. Annals of Internal Medicine, 151(4), 65-94.


Supplemental file 1. Search terms for PubMed.

# 1.

# 2.

# 3.
"Animals"[Mesh] NOT ("Animals"[Mesh] AND "Humans"[Mesh])

# 4.
"Letter" [Publication Type] OR "Editorial" [Publication Type] OR "Comment" [Publication Type] OR "Review" [Publication Type]

#1 AND #2 NOT (#3 OR #4)= 2896, limited to english[Filter] AND 2020:2020[pdat]
<table>
<thead>
<tr>
<th>First Author</th>
<th>Type of Study</th>
<th>Outcome</th>
<th>Prediction Model</th>
<th>Sample Size</th>
<th>Type of validation</th>
<th>Performance (AUC)</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-Najjar H et al.</td>
<td>Retrospective</td>
<td>Death</td>
<td>Country, infection reason, sex, group, confirmation date, birth year, and region</td>
<td>659</td>
<td>Split</td>
<td>.908 for recovery, .985 for death</td>
<td>Korea</td>
</tr>
<tr>
<td>Bae J et al.</td>
<td>Retrospective</td>
<td>Mechanical ventilation, death</td>
<td>Chest X-Ray (deep learning)</td>
<td>514</td>
<td>Split</td>
<td>.904 for ventilation .936 for mortality</td>
<td>US</td>
</tr>
<tr>
<td>Bahl et al.</td>
<td>Retrospective</td>
<td>Death</td>
<td>Age, RR, SpO2, Cr, ALT, procalcitonin, lactic acid</td>
<td>1629</td>
<td>LOOCV</td>
<td>V 0.80</td>
<td>US</td>
</tr>
<tr>
<td>Bello-Chavolla, O. Y. et al.</td>
<td>Retrospective</td>
<td>Death</td>
<td>Age, diabetes, diabetes, obesity, CKD, hypertension, and immunosuppression</td>
<td>177 133</td>
<td>Split</td>
<td>V .823</td>
<td>Mexico</td>
</tr>
<tr>
<td>Bi X et al.</td>
<td>Retrospective</td>
<td>Severe case</td>
<td>Fibrinogen to albumin ratio, platelet</td>
<td>141</td>
<td>External validation</td>
<td>V .754</td>
<td>China</td>
</tr>
<tr>
<td>Borghesi A et al.</td>
<td>Retrospective</td>
<td>Death</td>
<td>Brixia score, age, immunosuppression</td>
<td>302</td>
<td>External validation</td>
<td>V .853</td>
<td>Italy</td>
</tr>
<tr>
<td>Chen R et al.</td>
<td>Retrospective</td>
<td>Death</td>
<td>Age, CHD, CVD, dyspnea, procalcitonin, AST</td>
<td>1590</td>
<td>Bootstrap</td>
<td>V .91 (95% CI, .85-.97)</td>
<td>China</td>
</tr>
<tr>
<td>Chen A et al.</td>
<td>Retrospective</td>
<td>Death</td>
<td>BUN, D-dimer</td>
<td>305</td>
<td>Bootstrap</td>
<td>V 0.929</td>
<td>China</td>
</tr>
<tr>
<td>Cheng F et al.</td>
<td>Retrospective</td>
<td>ICU</td>
<td>RR, WBC, Lymphocyte, DBP, CRP, SpO2, age, temp, pulse, QRS duration, BUN, sodium, T wave axis, anion gap, SBP, PR interval, R wave axis, RBC, Calcium, Albumin</td>
<td>1987</td>
<td>Split</td>
<td>V .799 (95% CI: .752-.846)</td>
<td>US</td>
</tr>
<tr>
<td>Clift A et al.</td>
<td>Retrospective</td>
<td>Death</td>
<td>age, ethnicity, deprivation, BMI, comorbidity</td>
<td>8256158</td>
<td>Two temporary validations</td>
<td>V 0.928 (0.919 to 0.938)</td>
<td>UK</td>
</tr>
<tr>
<td>Study</td>
<td>Study Type</td>
<td>Setting</td>
<td>Outcomes</td>
<td>Discussion Points</td>
<td>N</td>
<td>Method</td>
<td>Country</td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
<td>---------</td>
<td>----------</td>
<td>-------------------</td>
<td>----</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>Dong Y et al.</td>
<td>Retrospective</td>
<td>Death</td>
<td>HTN, neutrophil to lymphocyte, and NT-Pro BNP</td>
<td>628</td>
<td>Split</td>
<td>V 0.922 (14day), V 0.881 (21 day)</td>
<td>China</td>
</tr>
<tr>
<td>Galloway J et al.</td>
<td>Retrospective</td>
<td>ICU, death</td>
<td>age, male, non-white ethnicity, SpO2, radiological severity score&gt;3, neutrophil count, CRP, albumin, Cr, diabetes, hypertension and chronic lung disease</td>
<td>1157</td>
<td>Split</td>
<td>V 0.714 (95%CI .665, .762)</td>
<td>UK</td>
</tr>
<tr>
<td>Gong J et al.</td>
<td>Retrospective</td>
<td>Severe case</td>
<td>Age; LDH, CRP, RDW, BUN, and direct bilirubin; and albumin</td>
<td>372</td>
<td>Split</td>
<td>V .853 (95% CI, .790-.916)</td>
<td>China</td>
</tr>
<tr>
<td>Haimovich et al.</td>
<td>Retrospective</td>
<td>Respiratory failure</td>
<td>Quick COVID severity index (RR, SpO2, and O2 flow rate), COVID severity score (O2 flow, SpO2, AST, Chloride, Procalcitonin, RR, SBP, ALT, WBC, BUN, Glucose, CRP, Ferritin, age, Cr)</td>
<td>1792</td>
<td>10-fold cross validation</td>
<td>V.81 (95% CI.73, .89) for the quick COVID severity index, V 0.76 [95% CI 0.65 to 0.86] for COVID severity score</td>
<td>US</td>
</tr>
<tr>
<td>Hu H et al.</td>
<td>Retrospective</td>
<td>Death</td>
<td>Rapid Emergency Medicine Score (REMS)</td>
<td>138</td>
<td>External</td>
<td>V 0.833 (95% confidence interval [CI] = 0.737 to 0.928)</td>
<td>China</td>
</tr>
<tr>
<td>Jang JG et al.</td>
<td>Retrospective</td>
<td>Death and critical outcomes</td>
<td>SIRS, qSOFA, NEWS</td>
<td>110</td>
<td>External</td>
<td>V SIRS = .639 (95% CI, .423-.856) qSOFA = .779 (95% CI, .600-.957) NEWS = 0.867 (95% CI .709–1.000) for mortality, SIRS = .744 (95% CI, .602–.886), qSOFA = .760 (95% CI, .620–.899), and NEWS = .918 (95%)</td>
<td>Korea</td>
</tr>
<tr>
<td>Study Authors</td>
<td>Study Design</td>
<td>Study Type</td>
<td>Predictors</td>
<td>n</td>
<td>Validation Type</td>
<td>Area Under Curve (95% CI)</td>
<td>Location</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------</td>
<td>--------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>----</td>
<td>--------------------</td>
<td>-----------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Ji D et al.31</td>
<td>Retrospective</td>
<td>Severe case</td>
<td>Age, comorbidity, lymphocyte, LDH</td>
<td>208</td>
<td>bootstrap</td>
<td>( V^{.91} (95% CI, \ .86-.94) )</td>
<td>China</td>
</tr>
<tr>
<td>Knight et al.32</td>
<td>Retrospective</td>
<td>Death</td>
<td>age, sex, comorbidities, RR, SpO2, level of consciousness, BUN, and CRP</td>
<td>57824</td>
<td>Split</td>
<td>( V\text{ range } .767 (95%CI .760-.773) )</td>
<td>UK, Wales, Ireland</td>
</tr>
<tr>
<td>Li M et al.34</td>
<td>Retrospective</td>
<td>Mechanical ventilation, death</td>
<td>Pulmonary X ray severity score</td>
<td>263</td>
<td>External</td>
<td>( V^{.80} (95%CI .75-.85) )</td>
<td>US</td>
</tr>
<tr>
<td>Li Q et al.35</td>
<td>Retrospective</td>
<td>Severe disease</td>
<td>Age, LDH, CD4</td>
<td>639</td>
<td>Prospective</td>
<td>( V^{.92} )</td>
<td>China</td>
</tr>
<tr>
<td>Liang W et al.36</td>
<td>Retrospective</td>
<td>Mechanical ventilation, death, ICU</td>
<td>chest radiographic abnormality, age, hemoptysis, dyspnea, unconsciousness, number of comorbidities, cancer history, neutrophil-to-lymphocyte ratio, LDH and direct bilirubin.</td>
<td>2300</td>
<td>Split</td>
<td>( D^{.88} (95% CI, .85-.91) ), ( V^{.88} (95% CI, .84-.93) )</td>
<td>China</td>
</tr>
<tr>
<td>Liu Y et al.37</td>
<td>Retrospective</td>
<td>Severe/Critical illness</td>
<td>Neutrophil to lymphocyte ratio, CRP</td>
<td>84</td>
<td>Split</td>
<td>( D^{.804} (95% CI: .702-.883), V^{.881} (95% CI: .782-.946) )</td>
<td>China</td>
</tr>
<tr>
<td>Matos J et al.38</td>
<td>Prospective</td>
<td>Mechanical ventilation, death</td>
<td>Gender, chronic lung disease, duration of symptom, predominant type, WBC count, aortic and coronary calcification, comorbidity, lymphocyte, age, volume of disease, CRP</td>
<td>106</td>
<td>Cross validation</td>
<td>( V^{0.92} )</td>
<td>Italy</td>
</tr>
<tr>
<td>McRae et al.39</td>
<td>Prospective</td>
<td>Death</td>
<td>biomarkers (cTnI, PCT, MYO, and CRP), age, and sex</td>
<td>172</td>
<td>External validation</td>
<td>( V^{0.94 (0.89–0.99)} )</td>
<td>China</td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Outcomes</td>
<td>Predictors</td>
<td>Sample size</td>
<td>Validation Method</td>
<td>Clinical/Lab</td>
<td>Country</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>----------</td>
<td>------------</td>
<td>-------------</td>
<td>------------------</td>
<td>-------------</td>
<td>---------</td>
</tr>
<tr>
<td>Myrstad et al.40</td>
<td>Prospective</td>
<td>Severe disease</td>
<td>NEWS2 score&gt;=6</td>
<td>66</td>
<td>External validation</td>
<td>V .822, (95% CI .690-.953)</td>
<td>Norway</td>
</tr>
<tr>
<td>Laguna-Goya et al.41</td>
<td>Prospective</td>
<td>Death</td>
<td>SpO2/FiO2 ratio, neutrophil-to-lymphocyte ratio, LDH level, IL-6 level, and age</td>
<td>611</td>
<td>bootstrap</td>
<td>V.93</td>
<td>Spain</td>
</tr>
<tr>
<td>Satci C et al.42</td>
<td>Retrospective</td>
<td>Death</td>
<td>PSI, CRP</td>
<td>681</td>
<td>External validation</td>
<td>V .92 (95% CI .89-.94)</td>
<td>Turkey</td>
</tr>
<tr>
<td>Shang Y et al.43</td>
<td>Retrospective</td>
<td>Death</td>
<td>Age, CVD, lymphocytes, Procalcitonin and D-dimer</td>
<td>452</td>
<td>External validation</td>
<td>V .938(95%CI .902-.973)</td>
<td>China</td>
</tr>
<tr>
<td>Shashikumar SP et al.44</td>
<td>Retrospective and prospective</td>
<td>Intubation</td>
<td>Deep learning⁶</td>
<td>777</td>
<td>External validation</td>
<td>V.918</td>
<td>US</td>
</tr>
<tr>
<td>Singh K et al.45</td>
<td>Retrospective</td>
<td>Death, ICU, intubation</td>
<td>Epic deterioration index</td>
<td>174</td>
<td>External validation</td>
<td>V0.67 (95% CI 0.59-0.75)</td>
<td>US</td>
</tr>
<tr>
<td>Sun H et al.46</td>
<td>Retrospective and prospective</td>
<td>Hospitalization, critical illness, death</td>
<td>CoVAS model⁵</td>
<td>11,586</td>
<td>Cross validation and External</td>
<td>Hospitalization V.76, critical illness V.79, death V.93</td>
<td>US</td>
</tr>
<tr>
<td>Sun L et al.47</td>
<td>Retrospective</td>
<td>Severe and critical symptoms</td>
<td>age, GSH, CD3 percentage and total protein</td>
<td>336</td>
<td>Not specified</td>
<td>D 0.9997 and V 0.9757</td>
<td>China</td>
</tr>
<tr>
<td>Wang B et al.48</td>
<td>Retrospective</td>
<td>Death</td>
<td>Age, Chest tightness, AST, BUN</td>
<td>104</td>
<td>Bootstrap</td>
<td>V.893 (95% CI .807-.980)</td>
<td>US</td>
</tr>
<tr>
<td>Wang K et al.49</td>
<td>Retrospective</td>
<td>Death</td>
<td>Age, HTN, CAD (clinical) Age, CRP, SpO2, Neutrophil, Lymphocyte, d-dimer, AST, GFR (lab based)</td>
<td>340</td>
<td>Split</td>
<td>Clinical V .83 (95%CI .68-.93), Lab V .88 (95%CI .75-.96)</td>
<td>China</td>
</tr>
<tr>
<td>Wei W et al.50</td>
<td>Retrospective</td>
<td>Severe COVID-19</td>
<td>CT score, CD8, CRP, tight chest</td>
<td>81</td>
<td>LGOCV</td>
<td>Clinical D.97 V.90 Radiomics D.95 V.89</td>
<td>China</td>
</tr>
<tr>
<td>Study Title</td>
<td>Study Design</td>
<td>Study Population</td>
<td>Outcome Measures</td>
<td>Sample Size</td>
<td>Validation Type</td>
<td>Validation Results</td>
<td>Country</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>-------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Wollenstein-Betech S* et al.51</td>
<td>Retrospective</td>
<td>1) Hospitalization, 2) death, 3) ICU, 4) mechanical ventilation</td>
<td>1) age, gender, CKD, diabetes, immunosuppression, 2) age, SARS-CoV-2 test, immunosuppression and pregnancy, 3) pneumonia, CVD, asthma, and SARS-CoV-2 test, 4) ICU and pneumonia, age, gender, CVD, obesity, pregnancy, and SARS-CoV-2 test</td>
<td>91,179</td>
<td>Cross validation</td>
<td>1) 0.612-0.622, 2) 0.674-0.687, 3) 0.538-0.554, 4) 0.541-0.560</td>
<td>Mexico</td>
</tr>
<tr>
<td>Wu G et al.13</td>
<td>Retrospective</td>
<td>Severe or critical illness</td>
<td>Age, Lymphocytes, CRP, LDH, CK, BUN, Ca, Total protein</td>
<td>725</td>
<td>Prospective validation</td>
<td>V range of 0.84 to 0.93</td>
<td>China, Italy, Belgium</td>
</tr>
<tr>
<td>Wu Q et al.51</td>
<td>Retrospective</td>
<td>death, mechanical ventilation, ICU</td>
<td>Clinic-radiomics signature (CrrScore) early/late phase</td>
<td>492</td>
<td>Split</td>
<td>Early phase D 0.826 (0.714-0.937), V 0.850 (0.763-0.935), Late phase D 0.911 (0.796-0.999), V 0.886 (0.675-0.999)</td>
<td>China</td>
</tr>
<tr>
<td>Wu S et al.52</td>
<td>Retrospective</td>
<td>Severe COVID-19</td>
<td>Age, neutrophil, lymphocyte, procalcitonin, and C-reactive protein</td>
<td>270</td>
<td>Split</td>
<td>D 0.955 V 0.945</td>
<td>China</td>
</tr>
<tr>
<td>Xiao, S. et al.53</td>
<td>Retrospective</td>
<td>Severe case</td>
<td>Hypertension, Neutrophil, CRP, Lymphocyte, LDH</td>
<td>690</td>
<td>Internal and external validation</td>
<td>V 0.871, 95% CI: 0.769-0.972 (Internal), V 0.826, 95% CI: 0.746-0.907 (external)</td>
<td>China</td>
</tr>
<tr>
<td>Yadaw A et al.54</td>
<td>Retrospective and prospective</td>
<td>Death</td>
<td>Age, SpO2, type of encounter, hydroxychloroquine, and max temperature</td>
<td>5051</td>
<td>External</td>
<td>V 0.91</td>
<td>US</td>
</tr>
<tr>
<td>Zhang C et al.3</td>
<td>Retrospective</td>
<td>Severe COVID-19</td>
<td>Age, WBC, Neutrophil, GFR, Myoglobin</td>
<td>80</td>
<td>Scoring validated internally</td>
<td>V 0.906</td>
<td>China</td>
</tr>
<tr>
<td>Zhang S et al.55</td>
<td>Retrospective</td>
<td>Death</td>
<td>Age, LDH, Neutrophil to lymphocyte, and bilirubin</td>
<td>828</td>
<td>External</td>
<td>V #1 0.878, #2 0.839</td>
<td>China</td>
</tr>
<tr>
<td>Authors</td>
<td>Study Design</td>
<td>Study Population</td>
<td>Predictors</td>
<td>Sample Size</td>
<td>Method</td>
<td>AUC (95% CI)</td>
<td>Country</td>
</tr>
<tr>
<td>---------</td>
<td>--------------</td>
<td>------------------</td>
<td>------------</td>
<td>-------------</td>
<td>--------</td>
<td>--------------</td>
<td>---------</td>
</tr>
<tr>
<td>Zheng Y et al.(^56)</td>
<td>Retrospective</td>
<td>Composite outcome ICU, mechanical ventilation, death Secondary Death</td>
<td>Age, sex, comorbidity, lymphocyte, extent, crazy paving sign, change in liver density</td>
<td>238</td>
<td>Split</td>
<td>V .89 (95% CI, .82--.96)</td>
<td>China</td>
</tr>
<tr>
<td>Zheng Y et al.(^57)</td>
<td>Retrospective</td>
<td>Severe COVID-19</td>
<td>Neutrophil, Lymphocytes, and Platelets (NLP score)</td>
<td>141</td>
<td>Validation with survival nomogram</td>
<td>V .821 (95% CI, .746--.896)</td>
<td>China</td>
</tr>
<tr>
<td>Zhou Y et al.(^58)</td>
<td>Retrospective</td>
<td>Severe COVID-19 per ATS guidelines</td>
<td>temperature, cough, dyspnea, HTN, CVD, CLD, and CKD</td>
<td>366</td>
<td>bootstrapping</td>
<td>Derivation .86 Validation .84 Nomogram .86</td>
<td>China</td>
</tr>
</tbody>
</table>

*Preprint only, # Deep learning model does not give an importance of variables for prediction. $ Model variables include age, diastolic BP, SpO2, Positive COVID, respiratory rate, stroke, chest x ray finding, heart rate, temperature, systolic BP, Charlson score, intracranial hemorrhage, subarachnoid hemorrhage, male, hematologic malignancy, renal cancer, pancreatitis, cystic fibrosis, cardiac arrest, seizure, amyotrophic lateral sclerosis, metabolic acidosis, myasthenia gravis, pneumothorax, spinal muscular dystrophy, pericarditis, high or low BMI, ARDS. D: Derivation AUC, V: Validation AUC.
Table 2. A list of studies with low risk of bias and low concern for applicability

<table>
<thead>
<tr>
<th>First Author</th>
<th>Outcome</th>
<th>Predictors</th>
<th>Performance (AUC)</th>
<th>Metrics</th>
<th>Setting</th>
<th>Geographic location</th>
<th>Online Risk Calculator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clift A et al.</td>
<td>Death, Hospitalization</td>
<td>Age, Sex, Ethnicity, BMI, postal (zip) code, housing, comorbidity</td>
<td>0.928, second validation 0.776</td>
<td>Absolute risk and relative risk</td>
<td>General</td>
<td>UK</td>
<td><a href="https://qcovid.org/BMJ/">https://qcovid.org/BMJ/</a></td>
</tr>
<tr>
<td>Knight S et al.</td>
<td>Death</td>
<td>Age, sex, comorbidity, respiratory rate, SpO2, GSC, BUN, CRP</td>
<td>0.767 (0.760 to 0.773)</td>
<td>Probability</td>
<td>Hospital</td>
<td>UK, Wales, Ireland</td>
<td><a href="https://isaric4c.net/risk">https://isaric4c.net/risk</a></td>
</tr>
</tbody>
</table>

For more information, visit www.prisma-statement.org.