PROGNOSTIC MODELS

RECOMMENDATIONS

To guide the decision to admit patients with COVID-19 to the hospital:

We suggest the use of Age, BUN, number of Comorbidities, CRP, SpO2/FiO2 ratio, Platelet count, Heart rate (ABC2-SPH) risk score, Confusion Urea Respiration Blood Pressure (CURB-65) severity score, Risk Stratification in the Emergency Department in Acutely Ill Older Patients (RISE-UP) score, and Rapid Emergency Medicine Score (REMS). *(Low quality of evidence; Conditional recommendation)*

There is insufficient evidence to recommend the use of 4C Mortality Score, COVID Outcome Prediction in the Emergency Department (COPE) model, and Quick Sepsis-related Organ Failure Assessment (qSOFA) score. *(Very low quality of evidence)*

To guide in the expectant monitoring of hospitalized patients:

We suggest the use of the 4C Deterioration model. *(Low quality of evidence; Conditional recommendation)*

There is insufficient evidence to recommend the use of Modified Early Warning Score (MEWS) and National Early Warning Score 2 (NEWS2). *(Very low quality of evidence)*

Consensus Issues

There was a high certainty of evidence that the QCOVID model can predict mortality from COVID-19. However, there was an issue on applicability as some of the components of this model (i.e., geographic region and Townsend deprivation quintile) is specific for the general population of England. Hence, its use warrants reconsideration of the component prognostic factors and validation in the Philippine setting before any recommendations can be made.

It was noted that the qSOFA model was already being used by some hospitals and centers in the Philippines. Clinicians should be guided on its use as it was found to have a very low quality of evidence for prediction of mortality of inpatients. There are other prognostic models such as the CURB-65, RISE-UP and REMS which are pre-existing models designed for specific patient populations and the ABC2-SPH model which has a good discrimination performance. All of these were found to have better quality of evidence compared with qSOFA. The 4C Mortality score and COPE model were also found to have a very low quality of evidence to predict mortality. Further, it was observed that there was a decrease in the discriminatory ability of the COPE model when externally validated. In terms of clinical deterioration, the 4C deterioration score was found to have a better predictive ability and quality of evidence compared to MEWS and NEWS2 model. Like the QCOVID model, these prognostic models also need to be locally validated but the components of these models can be easily obtained especially in the hospital setting, making the validation process easier.

EVIDENCE SUMMARY

Among adult patients diagnosed with COVID-19, should prognostic models be used to predict the likelihood of severe disease and mortality?

Evidence Reviewers: Patricia Pauline Remalante-Rayco, MD, Evelyn Salido, MD, MSc and Maria Teresa S. Tolosa, MD, D Clin Epi, FPDS

Key Findings

Thirty-three (33) cohort studies on prognostic models for clinical deterioration and mortality of individuals with COVID-19 were found. Most of the studies (n = 28) were assessed to have high risk of bias due to issues in participant selection and analysis. There were four (4) studies with unclear, and three (3) with low, risk of bias.

There are a few models that have been validated in more than one population. The 4C mortality score, ABC_2-SPH, CURB-65, REMS, and RISE UP models have fair to good prediction of mortality for inpatients, while qSOFA has poor to fair prediction. The MEWS model has poor prediction of clinical deterioration while NEWS2 has inconsistent prediction (poor to good). The 4C deterioration score, which has been investigated in only one study but was found to have low risk of bias, has fair predictive ability for clinical deterioration. None of these models has been validated in the Philippine population. The QCOVID model for mortality, while with high certainty of evidence, is specific for use in England and needs to modified to and validated in the local setting before any recommendations can be made.

Introduction

As of March 23, 2021, around 2.72 million deaths out of 123.4 million cases have been reported globally [1], while in the Philippines, deaths have numbered to 12,972 out of 671,792 confirmed cases, with 1.8% of patients being classified as severe or critical. [2] In order to reduce the risk of severe disease and mortality, numerous studies have assessed the usefulness of prognostic models that aim to identify patients at high risk of adverse outcomes from COVID-19. These models include early warning scores that were originally developed to identify and monitor inpatients at risk of deterioration (in order to facilitate transfer to intensive care units), or new models that were developed for the purpose of predicting the likelihood of severity or mortality among COVID-19 patients (referred in this report as “pre-existing models” and “COVID-19-specific models”, respectively). Two systematic reviews of 232 prognostic models in January 2021 concluded that many of them were poorly reported with high concern for bias and applicability, with the exception of two studies. [3,4]

Review Methods

We did a search of the literature until March 8, 2021 using the MEDLINE, Cochrane Central, McMaster Evidence Based Alerts, Cornell Open Access Publication (COAP), and Living Overview of the Evidence (L•OVE) databases and hand search of their references, using the search terms “death” or “mortality”, “severe”, and “predict”. The search yielded both published articles and preprints. The criteria for inclusion of studies were as follows: (1) population: adults diagnosed with COVID-19 infection, (2) intervention: prognostic models with external validation, (3) outcome: worsening severity or clinical deterioration or poor outcomes and/or mortality; and (4) cross-sectional, case-control, or cohort studies. Studies were excluded if they met any of the following criteria: (1) development of prediction models through machine-learning algorithms or artificial

intelligence without an available online tool; (2) number of outcome events (severe disease or mortality) <100. There were 33 articles that met our inclusion criteria.

We used the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) to plan the review. [5] The Prediction model Risk Of Bias ASsessment Tool (PROBAST) was used as a guide to appraise the studies. [6] We extracted the following details of the studies: participants, setting, study design, the predictive performance of each model, and methods of calibration and discrimination. Discrepancies in appraisal and data extraction were resolved through discussion and eventual agreement between the two authors.

Results

Characteristics of included studies

We found 33 studies that met the inclusion criteria [9-41]. These are summarized in Appendix 1. Most of the studies were done in Europe (United Kingdom (11 studies), Spain (3), Netherlands (3), Italy (2), Switzerland (1), Denmark (1)). The rest were done in the US (6), Mexico (1), Brazil (1), China (3), and Turkey (1). All selected studies were cohorts – six prospective and 27 retrospective. Most studies were done during the early part of the pandemic (first half of 2020) but ten studies collected data extending to or exclusively during the second half of 2020 up to January 2021. These latter studies may reflect the possible effects of the changing incidence and prevalence of infection on model performance. [7,8]

Most studies obtained data from electronic medical databases of secondary or tertiary medical centers, wherein the prognostic model was applied during the emergency department visit or hospital admission. In seven articles, the data were obtained from cohorts of patients (both ambulatory and hospitalized) whose primary care physicians' electronic health records were linked with hospital data and enrolled in regional, national, or international databases, most of them established even prior to the pandemic. Most of the studies included confirmed PCR-positive adults with community-acquired COVID. Six studies included suspected or clinically-diagnosed COVID (typical symptoms, chest CT infiltrates, and the absence of an alternative diagnosis). Three studies specified the inclusion of nosocomial COVID in their population.

The prediction outcome of interest is death in 26 studies, admission to the intensive care unit in six, or a composite of death or clinical deterioration in ten studies, with some studies determining multiple outcomes. Death was determined either during the in-hospital stay or over a specified period of time (14, 28, or 30-day). Clinical deterioration or progression of COVID-19 severity was usually defined as admission to an intensive care setting, need for oxygen supplementation (non-invasive or mechanical ventilation), or death. One study looked at thromboembolism as an outcome. Among the studies done in the general population, the mortality rate was 0.020 to 0.077% in the England population and 8.2% in the Denmark population. The mortality rate among hospitalized patients was 5.5 to 36.1%.

Ten studies validated 14 different pre-existing prediction models (pre-COVID-19 pandemic). Twenty-two articles focused on the development and/or validation of 43 new models for prediction of outcomes in COVID-19 patients. One study included both pre-existing and COVID-specific models. The components of the models were of varying complexity – from a few patient characteristics to combinations of demographic features, comorbidity, clinical features, and laboratory tests.

Overall summary of methodological quality

We used the PROBAST tool to evaluate the risk of bias and concern for applicability for each article as low, high, or unclear. The assessments of the 33 articles are presented in Appendix 1.

The overall risk of bias was assessed to be high in 28 studies, low in three studies, and unclear in four studies. Contributory to this assessment is the high risk of bias for the participant domain in 23 studies and the analysis domain in 24 studies. Particularly for the participant domain, high risk of bias was due to inclusion of severely ill patients that was a result of the studies being done in tertiary centers and specialized COVID-designated hospitals. Some studies excluded patients with incomplete data.

For the analysis domain, causes for high risk of bias include the following: small number of participants with the outcome relative to the number of predictors; mishandling of continuous predictors; exclusion of patients lost to follow up and those with missing data from the analysis; lack of imputation for missing data; selection of predictors based on univariable analysis (leading to loss of information and consequent reduction of the model's predictive ability); lack of accounting for censoring and competing risks, as well as for model overfitting or optimism; and lack of information on model performance measures (usually on calibration).

Among the issues in the predictor and outcome domains are lack of blinding for outcome during data abstraction; lack of clear information on timing of determination of predictors; and insufficient time interval between assessment of predictor and outcome.

Overall concern for applicability was unclear for all studies due to the inclusion of laboratory and/or imaging predictors, which may not be available in many local health facilities. Moreover, a delay in test results due to prolonged laboratory and imaging turnaround time may limit the application of prognostic models that are intended for immediate use to aid patient diagnosis or on admission at the emergency department. Models that include comorbidities as predictors with reliance on ICD-10 may be difficult to apply in our setting because most hospitals lack a readily-available database of comorbidities, leading to potential recall bias.

The GRADEpro Guidance Development Tool (GDT) was used to assess the certainty of evidence for six models that had multiple validation studies and five models with only one published validation study but had low to unclear risk of bias. (See Appendix 3 for GRADE Evidence Profiles)

Summary of results of included studies

We focus on models with more than one external validation study to obtain information both about accuracy of the model in predicting the outcome of interest and consistency of prediction. (See Table 1) We also briefly mention studies on prognostic models with only one external validation but were assessed to have unclear or low risk of bias. In all studies, the ability of a prediction model to discriminate among individuals who will develop an event or outcome (e.g., mortality or severity) from those who will not is measured by the area under the curve (AUC).

Models for prediction of mortality in the general population (ambulatory and hospitalized)

The QCOVID model developed by Clift et al., [9] and further validated by Nafilyan et al.[10] used data from a large network of primary care datasets with linkage to hospitalization data. It shows excellent discrimination of mortality risk but its use is quite specific to England by the nature of its components (geography, accommodation, Townsend deprivation index) and scoring system. It has the potential to help patients and doctors reach a shared understanding of mortality risk of COVID-19 diagnosed in the community even prior to the availability of laboratory tests. It was designed to be applied across the adult population for risk stratification for public health purposes during the pandemic, to support shared management of risk and occupational exposure, and in early targeting of vaccines to people most at risk [9]. The studies of Clift and Nafilyan have low

risk of bias and high certainty of evidence. However, for the QCOVID model to be used in the Philippines, it has to be modified and validated.

The COPE model was also developed from the general population (around 2 million Dutch recruited at point of RT-PCR testing). Its base model (age, sex, BMI) has excellent prediction of death at point of diagnosis (AUC 0.902) which falls to fair prediction at hospital admission (AUC 0.785). When validated in a UK cohort, prediction of death on diagnosis is much lower (AUC 0.742). This study has unclear risk of bias and the certainty of evidence is low.

Table 1. Models with more than one external validation study in this review

Model First Author Country	Population	Outcome	Predictors	Risk based on cut-off scores	Discrimination Performance (AUC, 95% CI if provided)*	Online risk calculator
QCOVID Clift [9] Nafilyan [10] (England)	General population	Mortality	Age, sex geographic region, ethnicity, Townsend deprivation quintile, accommodation,body mass index, chronic kidney disease (CKD), learning disability, chemotherapy, cancer/immunosuppression, other comorbidities	Online calculator gives absolute risk	Clift *Period 1* Jan-Apr 2020 *Men* 0.93 (0.92-0.93) *Women* 0.93 (0.92-0.94) *Period 2* May-Jun2020 *Men* 0.93 (0.92-0.95) *Women* 0.95 (0.94-0.96) Nafilyan *Period 1* *Men* 0.935 (0.933, 0.937) *Women* 0.945 (0.943, 0.947) *Period 2* *Men* 0.944 (0.942, 0.946) *Women* 0.956 (0.954, 0.958)	https://qcovid.org
4C Mortality Score Knight [12] (UK) Van Dam [39] (Netherlands)	Inpatients	Mortality (30-day)	Age, sex, number of comorbidities, respiratory rate, peripheral oxygen saturation, Glasgow coma scale (GCS), blood urea nitrogen (BUN), C-reactive protein (CRP)	Mortality risk (Score range: 0-21) Low: 0-3 Intermediate:4-8 High: 9-14 Very high:≥15	Knight 0.767 (0.76, 0.77) Van Dam 0.84 (0.79,0.88)	https://isaric4c.net/risk
Confusion Urea Respiration Blood Pressure (CURB-65) Artero [17]	Inpatients	Mortality (30-day)	Age, confusion, respiratory rate (RR), blood pressure, BUN		Artero 0.82 (0.82,0.84) Bradley 0.75 Liu FY 0.77	https://www.mdcalc.com/curb-65-score-pneumonia-severity

Model First Author Country	Population	Outcome	Predictors	Risk based on cut-off scores	Discrimination Performance (AUC, 95% CI if provided)*	Online risk calculator
(Spain) Bradley [21] (UK) Liu FY 30] (China) Nava [33] (US) van Dam [39] (Netherlands)					(0.72, 0.81) Nava 0.78 van Dam 0.75 (0.70, 0.80)	
Rapid Emergency Medicine Score (REMS) Gupta [27] (UK) Liu FY [30] (China) van Dam [39] (Netherlands	Patients at Emer- gency Department (ED)	Mortality (in-hospital)	Age, pulse rate (PR), mean arterial pressure (MAP), RR, GCS, oxygen saturation (SpO2)		Gupta 0.76 (0.71, 0.81) Liu FY 0.84 (0.8, 0.88) van Dam 0.73 (0.68, 0.78)	https://www.mdcalc.com/rapid-emergency-medicine-score-rems
Risk Stratification in the Emergency Department in Acutely Ill Older Patients (RISE UP) Van Dam [39] Van Dam [40] (Netherlands)	Patients at ED, Inpatients	30-day mortality	Age, HR, MAP, RR, SpO2, GCS, BUN, bilirubin, albumin, lactate dehydrogenase (LDH)	<10%- very low risk of mortality >30%- high risk of mortality	van Dam 0.77 (0.73, 0.81) van Dam 0.83 (0.79, 0.88)	P(30-day mortality)=1/(1+exp (-(−2.083+0.795 * (0.050*Age +1.115*≥2 Abnormal Vital Signs (yes=1, no=0)–0.112*Albumin (in g/L) +0.284* (BUN (in mmol/L)/5) +0.120* (LDH (in U/L)/100)+0.875* Bilirubin>20 µmol/L (yes=1, no=0))))
Quick Sepsis-related Organ Failure Assessment (qSOFA) Artero [17]	Inpatients	Mortality (in-hospital)	Mental status (GCS), RR, systolic blood pressure (SBP)		Artero 0.73 (0.71, 0.74) Bradley 0.62 Gupta 0.6 (0.54, 0.65)	https://www.mdcalc.com › qsofa-quick-sofa-score-sepsis

Model First Author Country	Population	Outcome	Predictors	Risk based on cut-off scores	Discrimination Performance (AUC, 95% CI if provided)*	Online risk calculator
(Spain) Bradley [21] (UK) Gupta [27] (UK) Liu FY [30] (China)					Liu FY 0.69 (0.64, 0.75)	
National Early Warning Score 2 (NEWS2) Baker [18] (UK) Bradley [21] (UK) Carr [22] (UK) Gupta [27] (UK) Gupta [14] (UK)	Inpatients on admission	Risk of clinical deterioration (CD)	RR, PR, hypercapneic respiratory failure, room air or with supplemental O2, Temp, SBP, consciousness		Baker 0.7 (0.65, 0.77) Carr CD in 3 days 0.72,0.77 CD in 14 days 0.70, 0.74 Gupta CD in 1-day 0.78 (0.73, 0.83) Gupta 0.69 (0.68, 0.70)	https://www.mdcalc.com/national-early-warning-score-news-2
Modified Early Warning Score (MEWS) Gupta [27] (UK) Gupta [14] (UK)	Inpatients	Risk of clinical deterioration	SBP, heart rate (HR), RR, temperature, Alert, Voice Response, Pain Response, Unresponsive (AVPU) score		Gupta 0.6 (0.56, 0.65) Gupta 0.63 (0.62, 0.64)	https://www.mdcalc.com/modified-early-warning-score-mews-clinical-deterioration

**Discrimination performance: ability of the model to discriminate between those who will and will not develop the outcome of interest. An area under the curve (AUC) of 0.90-1.0 indicates excellent discriminatory capacity of the model; 0.80-0.90 good; 0.70-0.80 fair; 0.60-0.70 poor; <0.60 failure to discriminate.*

Models for prediction of poor outcomes among hospitalized patients

Mortality

Pre-existing models such as CURB-65, REMS, and RISE-UP are established scoring systems that were designed for specific patient populations. CURB-65 is a tool that has been validated for use among patients with community-acquired pneumonia for the prediction of 30-day mortality. REMS was intended to predict in-hospital mortality for patients presenting at the emergency department regardless of disease, while RISE-UP was designed for mortality prediction specifically for elderly patients in the emergency room. Several studies have validated these models for use in COVID-19 as most of the affected patients, especially in the early part of the pandemic, were older than 65 years and presented with pneumonia. Their relative simplicity and popular use among clinicians have made them attractive models for risk stratification in COVID-19. These three models were found to have fair to good discriminative performance for mortality when validated in COVID-19 patients, with estimated AUCs ranging from 0.70 to 0.84. However, most validation studies were found to have high risk of bias, with low certainty of evidence, except the one by Gupta et al. (2021), which has unclear risk of bias.

Two studies on prognostic models for mortality among hospitalized patients at low risk of bias deserve some mention. The 4C Mortality Score was developed and validated from the

International Severe Acute Respiratory and Emerging Infections Consortium Coronavirus Clinical Characterisation Consortium (ISARIC4C) study, which involved 260 hospitals in the United Kingdom [12]. The model's components (age, sex, number of comorbidities, vital signs and BUN and CRP) are available in most hospitals. The model, which exhibited fair discriminative ability, was developed using a cohort of seriously-ill patients and may not be generalizable to patients in the community with lower mortality risk. The certainty of evidence from this study is low.

The ABC_2-SPH model was developed and validated in Brazil using variables commonly available in most emergency departments around the globe (age, number of comorbidity, heart rate, SpO2/FiO2 ratio, BUN, CRP and platelet count) [13]. It has an AUC above 0.8, indicating good discrimination. On external validation on a Spanish cohort with patients from the early part of the pandemic, the model was found to potentially underestimate mortality in patients who are at higher risk of death. The certainty of evidence from this study is low.

The rest of the models for prediction of mortality among hospitalized patients were at high risk of bias and with very low certainty of evidence.

Clinical Deterioration

A study of 66,136 adults with confirmed COVID-19 belonging to the ISARIC 4C cohort in the UK done by Gupta et al found that most in-hospital deterioration occurred around 4 days (1-9 days) from admission and declined with increasing time thereafter [14]. In-hospital clinical deterioration was defined as a need for non-invasive or invasive oxygen supplementation, admission to an intensive care unit, or death.

Early warning scores like NEWS2 and MEWS which use easy-to-obtain physical examination findings as predictors for in-hospital deterioration is desired during this early phase of hospitalization. They can easily be used in all clinical settings without need for additional training of medical staff and pose no socioeconomic issues like cost and availability of tests. However, validation studies of these two prognostic models for clinical deterioration showed poor to fair and poor discrimination indices of NEWS2 and MEWS, respectively.

The study by Gupta mentioned above was found to be of low risk of bias and will be mentioned here briefly. It developed and validated the 4C model for clinical deterioration, for use on admission for community-acquired COVID-19 cases, or at the initial assessment of suspected nosocomial COVID-19. It showed fair discrimination (AUC 0.77 [0.76, 0.78]). The certainty of evidence from this study is low. The rest of the studies on prognostic models for clinical deterioration of hospitalized patients with COVID-19 were at high risk of bias and with very low certainty of evidence.

Recommendations from Other Groups

The Australian guidelines for the clinical care of people with COVID-19 [42] recommend monitoring of markers of clinical progression, such as rapidly progressive respiratory failure and sepsis, especially on days 5 to 10 after onset of symptoms. This was not developed with an evidence-based framework, but formed through a consensus process.

The WHO COVID-19 Clinical management Living guidance (25 January 2021) [43] recommends close monitoring of moderate and severe COVID-19 patients for signs or symptoms of disease progression. The use of medical early warning scores (e.g. NEWS2, PEWS) that facilitate early recognition and escalation of treatment of the deteriorating patient is advised, where possible.

Moreover, in clinical decision-making in COVID-19, the WHO recommends the use of clinical judgment, including consideration of patients' values and preferences and local and national policy if available, to guide management decisions including admission to hospital and to the intensive care unit (ICU), rather than currently available prediction models for prognosis when caring for patients with COVID-19 of any severity assessed in a clinic or hospital (conditional recommendation, very low certainty).

Research Gaps

There are at least four ongoing studies registered in clinicaltrials.gov for prediction models for COVID-19 in various countries – one retrospective (case-only) study in China, one prospective case-control study from Germany (COVIP study), and two prospective cohort studies done in Switzerland (COVIVA study) and in Italy (CODED study). [44-47] While two of these studies focus on clinical deterioration and/or death among admitted patients similar to existing models, the COVIVA study looks at the short-term prognosis of suspected COVID-19 patients using personalized risk prediction models [46], and the CODED study aims to predict death or hospital admission (from and for any cause) among COVID-19 patients discharged from the emergency department [47]. These models look at combinations of demographic, clinical, biologic, and/or imaging parameters with or without use of machine-based learning.

Despite the large number of studies being produced, the majority of prediction models for adverse COVID-19 outcomes continue to suffer from selection bias, overfitting, and/or the lack of external validation. Article appraisals and the conduct of systematic reviews have paved the way for improved quality of studies. As the predictive performance of a model may differ depending on the setting and population to which it is applied, the importance of conducting external validation studies in settings where the model is intended to be used cannot be overemphasized. Validation of these prognostic models in the Filipino population is needed, and the use of these models must not replace clinical judgment with due consideration of patients' values and preferences. Pooling of multicenter data across heterogeneous settings and populations may help increase the robustness of model performance evaluation, especially when substantiated by meta-analyses.

References

[1] Jin Y, Yang H, Ji Wangquan, Wu Weidong, Chen S, Zhang W, Guangcai D. Virology, Epidemiology, Pathogenesis, and Control of COVID-19. Viruses. 2020;12(372):1–17.

[2] DOH COVID-19 Bulletin #374. https://doh.gov.ph/covid19casebulletin374. Accessed on March 23, 2021.

[3] Miller J, Tada M, Goto M, Mohr N, Lee S. Prediction Models for Severe Manifestations and Mortality due to COVID-19: a Rapid Systematic Review. medRxiv preprint doi: https://doi.org/10.1101/2021.01.28.21250718; this version posted February 1, 2021.

[4] Wynants L, Van Calster B, Collins G, Riley R, Heinze G, Schuit E et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020;369:m1328 http://dx.doi.org/10.1136/bmj.m1328

[5] Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. (2014) Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. PLoS Med 11(10): e1001744. https://doi.org/10.1371/journal.pmed.1001744

[6] Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med. 2019 Jan 1;170(1):W1-W33. doi: 10.7326/M18-1377. PMID: 30596876.

[7] Williamson 2021. Williamson E, Tazare J, Bhaskaran K, H McDonald, Walker A, Tomlinson L, OpenSAFELY Collaborative. Predicting COVID-19 related death using the OpenSAFELY platform. medRxiv preprint doi: https://doi.org/10.1101/2021.02.25.21252433; this version posted March 1, 2021.

[8] WHO weekly epidemiological update- 29 December 2020. https://www.who.int/publications/m/item/weekly-epidemiological-update---29-december-2020. Accessed on March 31, 2021.

[9] Clift AK, Coupland CAC, Keogh RH, Diaz-Ordaz K, Williamson E, Harrison EM et al. Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. BMJ 2020;371:m3731 http://dx.doi.org/10.1136/bmj.m3731

[10] Nafilyan V, Humberstone B, Mehta N, Diamond I, Coupland C, Lorenzi L et al. An external validation of the QCovid risk prediction algorithm for risk of mortality from COVID-19 in adults: national validation cohort study in England. medRxiv preprint https://doi.org/10.1101/2021.01.22.21249968 this version posted January 25, 2021

[11] Jimenez-Solem E et al. Developing And Validating Covid-19 Adverse Outcome Risk Prediction Models From A Bi-National European Cohort Of 5594 Patients. doi: https://doi.org/10.1101/2020.10.06.20207209

[12] Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ. 2020;370:m3339.

[13] Marcolino M, Pires M, Ramos L, Silva R, Oliveira L, Carvalho R, et al. ABC2-SPH risk score for in-hospital mortality in COVID-19 patients: development, external validation and comparison with other available scores. medRxiv preprint doi: https://doi.org/10.1101/2021.02.01.21250306; this version posted February 3, 2021

[14] Gupta R, Harrison E, Ho A, Docherty A, Knight S, Smeden M et al. Development and validation of the ISARIC 4C Deterioration model for adults hospitalised with COVID-19: a prospective cohort study. Lancet Respir Med 2021.https://doi.org/10.1016/S2213-2600(20)30559-2

[15] Adderley N, Taverner T, Price M, Sainsbury C, Greenwood D, Chandan J et al. Development and external validation of prognostic models for COVID-19 to support risk stratification in secondary care. medRxiv preprint doi: https://doi.org/10.1101/2021.01.25.21249942; this version posted January 30, 2021.

[16] Ageno W, Cogliati C, Perego M, Girelli D, Crisaffuli E, Pizzolo F et al. Clinical risk scores for the early prediction of severe outcomes in patients hospitalized for COVID-19. Internal and Emergency Medicine 2020. https://doi.org/10.1007/s11739-020-02617-4

[17] Artero A, Madrazo M, Fernandez-Garces M, Miguez A, Garcia A, Vieitez ASeverity Scores in COVID-19 Pneumonia: a Multicenter, Retrospective, Cohort Study. J Gen Intern Med DOI: 10.1007/s11606-021-06626-7.

[18] Baker KF, Hanrath A, van der Loeff I, Kay L, Back J, Duncan C. National Early Warning Score 2 (NEWS2) to identify inpatient COVID-19 deterioration: a retrospective analysis. Clinical Medicine 2021, Vol 21 No 2 March 2021.

[19] Bartoletti M, Giannella M, Scudeller L, Tedeschi S, Rinaldi M, Bussini L et al. Development and validation of a prediction model for severe respiratory failure in hospitalized patients with SARS-CoV-2 infection: a multicentre cohort study (PREDI-CO study). Clinical Microbiology and Infection. 2020 https://doi.org/10.1016/j.cmi.2020.08.003.

[20] Berenguer J, Borobia A, Ryan P, Rodriguez-Bano J, Bellon J, Harrin I et al. Development and validation of a prediction model for 30-day mortality in hospitalised patients with COVID-19: the COVID-19 SEIMC score. Thorax 2020. http://dx.doi.org/10.1136/ thoraxjnl-2020-216001).

[21] Bradley P, Frost F,Tharmaratnam K, et al. Utility of established prognostic scores in COVID-19 hospital admissions: multicenter prospective evaluation of CURB-65, NEWS2 and qSOFA. BMJ Open Resp Res 2020;7:e000729. doi:10.1136. bmjresp-2020-000729.

[22] Carr E, Bendayan R, Bean D, Stammers M. Wang W, Zhang H, Searle T et al. Evaluation and improvement of the NEWS2 for COVID-19: a multi-hospital study. BMC Medicine 19:23. https://doi.org/10.1186/s12916-020-01893-3.

[23] Castro V, McCoy T, Perlis R. Consistency of performance of adverse outcome prediction models for hospitalized COVID-19 patients. medRxiv preprint doi: https://doi.org/10.1101/2021.02.25.21252410; this version posted March 1, 2021.

[24] Chua F, Vancheeswaran R, Draper A*, Vaghela T, Knight M, Mogal R et al.* Early prognostication of COVID-19 to guide hospitalisation versus outpatient monitoring using a point-of-test risk prediction score*Thorax* Published Online First: 10 March 2021. doi: 10.1136/thoraxjnl-2020-216425

[25] Codon J, Lip G, Rey H, Iniesta A, Rosillo S, Castrejon S et al. Prediction of thromboembolic events and mortality by the $CHADS_2$ and the CHA_2DS_2-VASc in COVID-19. Europace (2021) 00, 1–11 doi:10.1093/europace/euab015.

[26] El-Solh AA, Lawson Y, Carter M, El-Solh DA, Mergenhagen KA (2020) Comparison of in-hospital mortality risk prediction models from COVID-19. PLoS ONE 15(12): e0244629. https://doi.org/10.1371/journal.pone.0244629

[27] Gupta RK, Marks M, Samuels THA, et al. Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: An observational cohort study. Eur Respir J 2020; in press (https://doi.org/10.1183/13993003.03498-2020).

[28] King JT, Jr., Yoon JS, Rentsch CT, Tate JP, Park LS, Kidwai-Khan F, et al. (2020) Development and validation of a 30-day mortality index based on pre-existing medical administrative data from 13,323 COVID-19 patients: The Veterans Health Administration COVID-19 (VACO) Index.PLoS ONE 15(11): e0241825. https://doi.org/10.1371/journal.pone.0241825

[29] Li, J., Chen, Y., Chen, S. *et al.* Derivation and validation of a prognostic model for predicting in-hospital mortality in patients admitted with COVID-19 in Wuhan, China: the PLANS (platelet lymphocyte age neutrophil sex) model. *BMC Infect Dis* 20, 959 (2020). https://doi.org/10.1186/s12879-020-05688-y

[30] Fang-Yan Liu, Xue-Lian Sun, Yong Zhang, Lin Ge, Jing Wang, Xiao Liang, Jun-Fen Li, Chang-Liang Wang, Zheng-Tao Xing, Jagadish K. Chhetri, Peng Sun, Piu Chan. Evaluation of the Risk Prediction Tools for Patients With Coronavirus Disease 2019 in Wuhan, China: A Single-Centered, Retrospective, Observational Study. Critical Care Medicine DOI: 10.1097/CCM.0000000000004549.

[31] Liu H, Chen J, Yang Q, Wang Y, Cai J, Li H. Development and validation of a risk score using complete blood count to predict in-hospital mortality in COVID-19 patients, Med (2020), https://doi.org/10.1016/j.medj.2020.12.013

[32] Mancilla-Galindo J, Vera- Zertuche JM, Navarro-Cruz AR, Segura-Badilla O, Reyes-Velázquez G, Tepepa-López FJ, Aguilar-Alonso P, Vidal-Mayo JdeJ, Kammar- García A (2020). Development and validation of the patient history COVID-19 (PH-Covid19) scoring system: a multivariable prediction model of death in Mexican patients with COVID-19. Epidemiology and Infection 148, e286, 1–8. https://doi.org/10.1017/ S0950268820002903

[33] Guillermo Rodriguez-Nava, Maria Adriana Yanez-Belloa, Daniela Patricia Trelles-Garciaa, Chul Won Chunga, Harvey J. Friedmanb,c, David W. Hinesd. Performance of the quick COVID-19 severity index and the Brescia-COVID respiratory severity scale in hospitalized patients with COVID-19 in a community hospital setting. International Journal of Infectious Diseases 102 (2021) 571–576

[34] Nicholson C, Wooster L, Sigurslide H, Li R, Jiang W, Tian W, Cardenas C, Malhotra R. Estimating risk of mechanical ventilation and in-hospital mortality among adult COVID-19 patients admitted to Mass General Brigham: The VICE and DICE scores. EClinicalMedicine 33 (2021) 100765

[35] Paranjape N, Staples L, Stradwick C, Ray H, Saldanha I. Development and validation of a predictive model for critical illness in adult patients requiring hospitalization for COVID-19 medRxiv preprint doi: https://doi.org/10.1101/2021.01.22.21250289; this version posted January 22, 2021.

[36] Richardson D, Faisal M, Fiori M, *et al*. Use of the first National Early Warning Score recorded within 24 hours of admission to estimate the risk of in-hospital mortality in unplanned COVID-19 patients: a retrospective cohort study. *BMJ Open* 2021;11:e043721. doi:10.1136/bmjopen-2020-043721

[37] Schöning V, Liakoni E, Baumgartner C, Exadaktylos A, Hautz W, Atkinson A, Hammann F Development and validation of a prognostic COVID-19 severity assessment (COSA) score and machine learning models for patient triage at a tertiary hospital*. J Transl Med (2021) 19:56* https://doi.org/10.1186/s12967-021-02720-w

[38] Tanboga IH, Canpolat U, Çetin E, Kundi H, Çelik O, Çağlayan M, Ata N, Özeke O, Çay S, Kaymaz C, Topaloğlu S. Development and validation of clinical prediction model to estimate the probability of death in hospitalized patients with COVID-19: Insights from a nationwide database. J Medical Virology 2021;1–8 DOI: 10.1002/jmv.26844

[39] van Dam P, Zelis N, van Kuijk S, Linkens A, Bruggemann R, Spaetgens B et al. Performance of prediction models for short term outcome in COVID-19 patients in the emergency department: a retrospective study. Ann Med 2021 Dec;53(1):402-409. doi: 10.1080/07853890.2021.1891453

[40] van Dam PMEL, Zelis N, Stassen P, *et al*. Validating the RISE UP score for predicting prognosis in patients with COVID-19 in the emergency department: a retrospective study. *BMJ Open* 2021;11:e045141. doi:10.1136/bmjopen-2020-045141

[41] van Klaveren D, Rekkas A, Alsma J, Verdonschot R, Koning D, Kamps M et al. COVID Outcome Prediction in the Emergency Department (COPE): Development and validation of a model for predicting death and need for intensive care in COVID-19 patients. medRxiv preprint doi: https://doi.org/10.1101/2020.12.30.20249023; this version posted February 8, 2021

[42] National COVID-19 Clinical Evidence Task Force. Living Guidelines. https://covid19evidence.net.au/

[43] COVID-19 CLinical Management: Living Guidance. https://www.who.int/publications/i/item/WHO-2019-nCoV-clinical-2021-1

[44] Sun JG. A Novel Nomogram to Predict Severity of COVID-19. 2020 Apr 28 [last updated 2020 Jul 14; cited 2021 Apr 8]. In: [Internet]. Bethesda (MD): U.S. National Library of Medicine. 2020 - Available from: https://clinicaltrials.gov/ct2/show/NCT04366024?term=prediction+model&cond=Covid19&draw=3&rank=15

[45] Twerenbold R, Pfister GK, et al. COVID-19 Survival - The COVIVA Study (COVIVA). 2020 Apr 29 [last updated 2020 May 4; cited 2021 Apr 8]. In: [Internet]. Bethesda (MD): U.S. National Library of Medicine. 2020 - Available from: https://clinicaltrials.gov/ct2/show/NCT04366765

[46] Jung C. Outcomes and Prognostic Factors in COVID-19 (COVIP). 2020 Mar 25 [last updated 2020 Dec 19; cited 2021 Apr 8]. In: [Internet]. Bethesda (MD): U.S. National Library of Medicine. 2020 - Available from: https://clinicaltrials.gov/ct2/show/NCT04321265

[47] Morello F. Risk Stratification of COVID-19 Patients Discharged From the Emergency Department (CODED). 2020 Nov 16 [last updated 2021 Feb 9; cited 2021 Apr 8]. In: [Internet]. Bethesda (MD): U.S. National Library of Medicine. 2020 - Available from: https://clinicaltrials.gov/ct2/show/NCT04629183?term=CODED&cond=Covid19&draw=2&rank=1

Appendix 1. Characteristics of Included Studies

Author and year Model	Study design and Setting	Specific Outcome	Time point of model application	Population and sample size	Candidate predictors	Determination of candidate predictors	Final predictors	Performance AUC (95% CI)
Adderley 2021 [15]	Retrospective cohort UK Jan 1-Sep 12, 2020 Last admission on Aug 16, 2020	28-day outcomes Mortality Training cohort (TC) n=288 (7.48%) Validation cohort (VC) n=1668 (27.35%) Admission to intensive care unit (ICU) TC, n= 183 (4.75%) VC, n= 722 (12.66%)	Admission or up to 72h from admission	TC Inpatients RT-PCR positive (RT-PCR+) n = 3849 VC RT-PCR+ or antibody test-positive n=6099	Demographic features Clinical features Laboratory features Imaging Frailty score Glasgow Coma Score (GCS) Comorbidity 63 candidate predictors for model development 27 candidate predictors in external validation population	Literature review Discussion with experts Available collected variables Time series analysis	Mortality model: Age, breathlessness, sputum, systolic blood pressure (SBP),Temperature (temp), Respiratory rate (RR), Oxygen saturation (O2sat) , FiO2, alkaline phosphatase (ALP), C-reactive protein (CRP), Calcium, Eosinophils, Glucose, pH, Urea, WBC count, platelets, and frailty score ICU admission model: age, gender, fever, new onset diarrhoea or vomiting, heart rate (HR),RR, FiO2, temp Albumin, CRP, eGFR, pH, monocytes, WBC, frailty score, and GCS Reduced model: Age, SBP, temp, RR, O2 sat, FiO2, frailty score, pH,urea, CRP	TC Mortality 0.778 (0.741, 0.815) ICU admission 0.892 (0.865, 0.920) Reduced model Mortality TC 0.791 (0.761, 0.822); VC 0.767 (0.754, 0.780) ICU admission TC 0.906 (0.883, 0.929) VC 0.811 (0.795, 0.828)

Author and year Model	Study design and Setting	Specific Outcome	Time point of model application	Population and sample size	Candidate predictors	Determination of candidate predictors	Final predictors	Performance AUC (95% CI)
Ageno 2020 [16] Italy	Retrospective cohort Five centers Feb 17 - May 8, 2020	Severe outcome defined as non-invasive ventilation (NIV), intubation, or death n= 275 (45.08%)	Admission	Inpatients RT-PCR+ n = 610	Demographic features Comorbidity Laboratory tests	No explanation for selection Multivariate logistic regression with backward selection, LASSO, Random Forest	Age Coronary heart disease PCR, AST, D-dimer, NLR 6 variables 13 points	0.80 Cut-off 7 points in VCt Sensitivity0.93 Specificity 0.34 PPV 0.59 NPV 0.82.
Artero et al., 2020 [17] Spain	Retrospective cohort Multi-center March-May 2020	Mortality, all-cause n= 2135 (20.9%) ICU/mechanical ventilation n=907 (8.9%)	Admission	Inpatients RT-PCR+/ antibody+ n =10,238	N/A	N/A	Pneumonia Severity Index CURB-65 qSOFA (altered mental status, RR, SBP) MuLBSTA: Age, smoking, bacteria infection, HPN Lymphocytes Imaging-multilobar infiltrates	*Mortality* PSI 0.835 (0.826, 0.845) CURB-65 0.825 (0.815, 0.835) MuLBSTA 0.715 (0.703, 0.727) qSOFA: 0.728 (0.715, 0.741) *ICU admission* PSI 0.539 (0.521, 0.557) CURB-65 0.562 (0.544, 0.580) MuLBSTA 0.658 (0.640, 0.677) qSOFA 0.616 (0.598, 0.635) *Mechanical ventilation* PSI 0.560 (0.540, 0.579) CURB-65 0.572 (0.553, 0.592) MuLBSTA 0.678 (0.657, 0.698) qSOFA 0.624 (0.603, 0.644)

Author and year Model	Study design and Setting	Specific Outcome	Time point of model application	Population and sample size	Candidate predictors	Determination of candidate predictors	Final predictors	Performance AUC (95% CI)
Baker 2021 [18] UK	Retrospective cohort Single-center Jan- Apr, 2020	Clinical deterioration defined as initiation of NIV or mechanical ventilation (MV), ICU admission, end of care, or in-hospital death *Data censored at 28 days for patients still admitted n= 133 (44.9%)	Admission	Inpatients RT-PCR+ n = 296 131 with severe covid on admission	N/A	N/A	NEWS2:HR, BP, temp, RR, O2 sat, level of consciousness	0.70 (0.65–0.77)
Bartoletti 2020 [19] Italy	Retrospective cohort study Multi-centre Feb-April, 2020	Severe respiratory failure- SpO2 <93% with 100% FiO2, RR>30/ min or respiratory distress n=367 (32.97%)	Admission	Inpatients RT-PCR+ n = 1113	Demographic, Comorbidities Symptoms on admission Laboratory tests	No explanation	PREDI-CO score: Age Obesity, RR Fever at hospitalization Lymphocytes Creatinine, CRP, LDH	0.85 (0.81-0.88)
Berenguer 2021 [20] Spain	Retrospective cohort Feb-April, 2020	30-day mortality n= 341 (5.5%)	Admission	Inpatients RT-PCR+ n = 6161	Demographic Comorbidities Signs and symptoms Laboratory tests	17 baseline variables found to be independently associated with death in the COVID-19 Spain cohort	COVID-19 SEIMC score Age, sex Dyspnea, age-adjusted SaO2 NLR, eGFR by CKD-EPI	0.831 (0.806–0.856)
Bradley 2020 [21] UK	Prospective Seven respiratory hospitals in NW England April 1-	30-day mortality n=300 (36.14%) ICU admission n=142	Admission	Consecutive adults admitted meeting the Public Health England inpatient case definition for COVID-			CURB-65 NEWS2 qSOFA	30-day mortality CURB-65- 0.75 NEWS2- 0.67 qSOFA- 0.62 72-h mortality CURB-65- 0.76 NEWS2- 0.78 qSOFA 0.65 ICU admission

Author and year Model	Study design and Setting	Specific Outcome	Time point of model application	Population and sample size	Candidate predictors	Determination of candidate predictors	Final predictors	Performance AUC (95% CI)
	14, 2020	(17.10%)		19 and PCR+ n=800				CURB-65 0.63 NEWS2 0.65 qSOFA 0.55
Carr 2021 [22] UK	Retrospective cohort study Multi-center Feb-Aug, 2020	Severe covid-19 outcomes (transfer to ICU or death at 3 and 14 days from admission or symptom onset for nosocomial COVID TC 3-day n = 389 (30.48%) 14-day n= 163 VC 3-day n= 27-289 14-day n=39-391	Admission (up to 48h after admission)	Inpatients RT-PCR+ n = 7513 TC, n = 1276 (1 hospital of NHS Hospitals Trust) VC,n = 6237 (5 centers of NHS Hospitals Trust, 1 hospital in Norway, 2 hospitals in Wuhan, China)	Age, sex, ethnicity, select comorbidity, physiologic measures (NEWS2), biomarkers (alb, CRP, GFR, lymphocyte ct, neutrophil ct, platelet ct, NLR, lym-CRP ratio, urea)	routinely obtained parameters available in a wide range of settings Regularised logistic regression with least absolute shrinkage & selection operator (LASSO) estimator	Model 1: NEWS2 only Model 2: NEWS2 + age Model 3: Supplemented NEWS2 score Age Supplemental O2 flow rate, O2 sat, eGFR Urea, CRP, Neutrophil count NLR	NEWS2 3-day: 0.717-0.772 14-day: 0.697-0.743 NEWS+age 3-day: 0.717-0.772 14-day severe outcome: 0.686-0.815 All features 3-day: 0.716-0.831 14-day: 0.762-0.864 *All models showed evidence of increasing miscalibration as new predictors are added to the model
Castro 2021 [23] USA	Retrospective cohort study Multi-center- 2 academic centers and 4 community hospitals June 7, 2020-Jan 22, 2021	1) Severe illness (composite of ICU admission, mechanical ventilation, or mortality) n=241 (8.3%) 2) mortality n=167 (5.8%)	Admission	Inpatients RT-PCR+ within 5 days of admission (n = 2,892) 2 academic medical centers, 4 community hospitals Mass General Brigham Data Registry Enterprise Data Warehouse (Temporal validation)	Age SpO2 Comorbidity (CCI)	N/A Logistic regression Survival analysis Right-censoring	Severe illness model: Age, SpO2, BUN, CRP, crea, low eGFR, eosinophils, ,glucose, LDH, lymphocytes, low ALC,monocytes, neutrophil, high ANC,plt, Trop T Charlson comorbidity index (CCI), prior respiratory infections Mortality model: Same as severe illness model plus low MCH ,high ANC, high absolute nucleated RBC , low plt PCT, RDW,Trop T, high WBC CCI, COPD or bronchiectasis,	severe illness: 0.79 (95% CI:0.75-0.81) mortality: 0.83 (95% CI:0.80-0.87)

Author and year Model	Study design and Setting	Specific Outcome	Time point of model application	Population and sample size	Candidate predictors	Determination of candidate predictors	Final predictors	Performance AUC (95% CI)
							dementia or delirium external causes of injury lung CA respiratory failure or insufficiency	
Chua 2020 [24] UK	Prospective Multi-center March 1-May 16, 2020	In-hospital mortality TC, n=294 (29.9%) (university hospital NHS Hospitals Trust) VC1 ISARIC n=4319 (30.35%) VC2 n= 94 (32.41%) (university hospital NHS Hospitals)	Presentation to ED	Adults ≥18 years old PCR+ at Emergency Department (ED) TC n=983 VC 1 n=14231 VC 2 n=290 n=5 confined at cut-off day	NEWS Demographic Routine laboratory tests	Usual data collected at the ED	SOARS11 SpO2, obesity, age, RR, stroke, smoking, dementia, CKD with stage, Wbc count, lymphocytes CXR (≥4 zones affected) SOARS5: SpO2, Obesity, Age, RR Stroke history	SOARS11 TC 0.82 VC1 0.80 VC2 0.74
Clift 2020 [9] UK	Retrospective	Time to death from COVID-19 TC n=4384 deaths (0.07%)[#] VC1: Jan 24-Apr 30, 2020 n=1722 (0.07%)[#] VC 2: May1-Jun 30,2020 n =621 (0.02%)[#] [#]% deaths in whole population (includes those with and without COVID-19	Not specified	Adults 19-100 years old QResearch database (1205 general practices in England linked to death and hospital registries TC n=6.08 million VC n= 2.17 million	Demographic data Comorbidity	Data available in the database	QCOVID Age Ethnicity Deprivation index BMI Comorbidity	Period 1 Women 0.93 (0.92, 0.94) Men0.93 (0.92, 0.93) Period 2 Women 0.95 (0.94, 0.96) Men0.93 (0.92, 0.95)

Author and year Model	Study design and Setting	Specific Outcome	Time point of model application	Population and sample size	Candidate predictors	Determination of candidate predictors	Final predictors	Performance AUC (95% CI)
Codon 2021 [25] CHADS CHA2DS 2-VASc Spain	Retrospe ctive Mar1-Apr 20, 2020	Thromboemb olism n= 115 (3.78%) Mortality n=626 (20.58%)	Not stated i Presumed to be on admission	Inpatients confirmed COVID completed 1-month follow-up or died, Mar 1-Apr 20,2020	CHADS and CHA2DS2-VASc	–	CHADS CHF or LV ejection fraction ≤40, HPN, Age Stroke, transient ischaemic attack (TIA), systemic embolism CHA2DS2-VASc score: age, sex, history of stroke, TIA, CHF, HPN, thromboembolis m, diabetes mellitus	Thromboembolis m CHADS 0.497 (0.452,0.542) CHA2DS2-VASc 0.490 (0.440,0.541) Mortality CHADS 0.788 (0.770–0.807) CHA2DS2-VASc 0.794 (0.775,0.812)
El-Solh 2020 [26] US	Retrospe ctive cohort Jan-May, 2020	In-hospital mortality n=475 (29.07%)	Admission	Inpatients RT-PCR+ n = 1634	–	–	Chen: Age, CHD, CVD, dyspnea, PCT, AST Shang: Age, CHD, % lymphocytes, PCT, D-dimer Wang Clinical model: age, HPN, CHD Laboratory model: Age, hsCRP, peripheral capillary O2 sat, neutrophil and	Chen 14-day mortality: 0.67 (0.64–0.70) 21-day mortality: 0.68 (0.65–0.71) 28-day mortality: 0.69 (0.66–0.72) Shang.: 0.72 (0.69–0.74) Yu: 0.63 (0.60, 0.66) Wang: 0.69 (0.66, 0.72)

Author and year Model	Study design and Setting	Specific Outcome	Time point of model application	Population and sample size	Candidate predictors	Determination of candidate predictors	Final predictors	Performance AUC (95% CI)
							lymphocyte count, D-dimer, AST, GFR Yu: age, male sex, history of diabetes, lymphopenia, increased PCT	
Gupta 2020 [27] Multiple models 2(UK)	Retrospective Feb-Apr 2020	Mortality N=115 (28%) Clinical deterioration (ventilatory support or death) N=180 (43.8%	On hospital admission	Consecutive adults admitted to University College Hospital London, PCR+ or clinically diagnosed COVID-19 (diagnosis of ID Specialist - clinical features, laboratory, radiologic results without alternative diagnosis. n=411	n/a	n/a	22 models	Mortality Lu: 12-day 0.72 (0.67, 0.76) 30-day CURB-65: 0.74 (0.69, 0.79) BelloChavolla 0.66 (0.6, 0.72) In-hospital REMS: 0.76 (0.71, 0.81) qSOFA: 0.6 (0.54, 0.65) Xie: 0.76 (0.69, 0.82) Hu 0.74 (0.68, 0.79) Caramelo: 0.71 (0.66, 0.76) Zhang: 0.7 (0.65, 0.76) Yan: 0.58 (0.49, 0.67) Deterioration NEWS1 day 0.78 (0.73, 0.83) Ji: 10 days 0.56 (0.5, 0.62) Carr: 14 days 0.78 (0.74, 0.82) Guo: 0.67 (0.61, 0.73) Zhang: 0.74 (0.69, 0.79) Galloway: 0.72 (0.68, 0.77) TACTIC: 0.7 (0.65, 0.75) Colombi: 0.69 (0.63, 0.74) Huang: 0.67 (0.61, 0.73) Shi: 0.61 (0.56, 0.66) MEWS: 0.6 (0.56, 0.65)

Author and year Model	Study design and Setting	Specific Outcome	Time point of model application	Population and sample size	Candidate predictors	Determination of candidate predictors	Final predictors	Performance AUC (95% CI)
Gupta , 2020 [14] UK	Prospective cohort Feb-Aug 2020	in-hospital clinical deterioration- initiation of ventilatory support (NIV, MV, ECMO); admission to ICU or death n = 31 924 (43.17%)	Admission or first clinical suspicion of covid	Inpatients Suspected/ Confirmed RT-PCR+ n = 73 948	Demographic, clinical, laboratory features, comorbidities	Review of iterature, availability in >60% of the study population	4C Deterioration Score: Age, sex + comorbidity + nosocomial infection + radiographic infiltrates + periph O2 sat + room air or o2, GCS, Urea, CRP, lymphocyte	0.77 (95% CI 0.76, 0.78)
King , 2021 [28] VACO USA	Retrospective Feb-July 2020	30-day mortality TC n=480 (13%) VC1 n=253 (12%) VC2 n=403 (5%)	Admission	Inpatients RT-PCR+ inpatients testing + within 14 days before or in the hospital (D1/ later)	Demographic Comorbidity	Multivariable logistic regression	Age Sex Comorbidity MI or PVD	Hospital cohort: 0.80 (0.77, 0.83) Medicare cohort: 0.67 (0.67, 0.68) 0.68 (0.68 – 0.68)
Knight 2020 [12] 4C Mortality Score	Prospective cohort Feb-Jun 2020	Mortality TC n=11426 (32.22%) Feb 6-May 20, 2020 VC n=6729 (30.09%) May 21-June 29, 2020 No recorded outcome considered alive.	Hospital admission	Inpatients ≥ 18 years RT-PCR+ TC n= 35463 VC n= 22,361 (Temporal validation)	Patient and clinical variables Clinical biomarkers for COVID-19	41 candidate predictors selected a priori based on influence on outcome of pneumonia & flu-like illness, COVID-19, available for 2/3 of patients in TC 3-step model development	Age, sex, number of comorbidities RR, O2 sat, GCS, Urea,CRP	TC 0.786 (0.781,0.79) VC 0.767 (0.76, 0.773)
Li 2020 [29] PLANS China	Retrospective Jan-Mar 2020	In-hospital mortality TC n=211 (20.93%) VC n=162 (15.71%)	On admission	Inpatients Adults RT-PCR+ TC (2 hospitals) Jan 1-Feb 10, 2020	Patient characteristic Laboratory tests	Clinical knowledge, literature, data availability Multivariable Fine-Gray model	Platelet count, lymphocyte count, Age Neutrophil count Sex	TC 0.85 (0.83, 0.87) VC 0.87 (0.85, 0.89)

Author and year Model	Study design and Setting	Specific Outcome	Time point of model application	Population and sample size	Candidate predictors	Determination of candidate predictors	Final predictors	Performance AUC (95% CI)
				n=1008 VC (1 hospital) Jan 14-Mar 8, 2020), n=1031				
Liu FY 2020 [30] NEWS NEWS2 REMS CURB-65 qSOFA China	Retrospective Single-COVID center Jan 30-Mar 14, 2020	In-hospital death N=121 (17.98%)	On admission	Inpatients Adults RT-PCR+	n/a	n/a	NEWS NEWS2 REMS CURB-65 qSOFA	0.882 (0.847–0.916) 0.880 (0.845–0.914) 0.839 (0.800–0.879) 0.766 (0.718–0.814) 0.694 (0.641–0.746)
Liu H 2021 [31] PAWNN China	Retrospective Jan-Apr 2020	In-hospital death TC n=773 (7.88%) VC: China 211 (7.7%) Italian 77 (33.92%)	Admission and throughout hospitalization	Inpatients RT-PCR+ or clinically diagnosed Excluded leukemia inpatients at study end TC n=9810 VC1 n=2739 VC2 n=227	38 candidate predictors Demographic Clinical findings Laboratory tests- CBC	Generalized linear mixed modelling, Cox regression model	Platelet count Age WBC count Neutrophil count Neutrophil:lymphocyte ratio	TC 0.92-93 (0.91, 0.94) VC Chinese 0.97 (0.96–0.98) Italian 0.80 (0.74, 0.86)
Mancilla-Galindo, 2021 [32] Mexico	Retrospective	Mortality TC n=9228 (11.01%) VC n=5278 (5.28%)	Not mentioned	Inpatients Ambulatory RT-PCR+ TC n=83779 VC n=100,000	Demographic, Clinical Comorbidity	Demographic Patient history predictors for low-resource settings Univariate and multivariate regression	Age Sex Diabetes COPD Immunosuppression Hypertension Obesity CKD	0.8 (0.796, 0.804)
Marcolino 2021 [13] ABC_2-SPH Brazil	Retrospective Mar-Jul 2020. Aug-Sep 2020	In-hospital deaths TC n=806 (20.26%) VC1 Brazil n=208 (19.73%) VC2 Spain 82 (17.29%)	Admission	Inpatients PCR+ TC n=3978 VC1 n=1054 VC2 n=474	20 predictors chosen a priori Demographic Comorbidity Vital signs Laboratory tests	Least absolute shrinkage and selection operator (LASSO) logistic regression	Age, blood urea nitrogen, number of comorbidities, HR, CRP, SpO2/FiO2 ratio, platelet count	TC 0.844 (0.829,0.859) VC1 0.859 (0.833, 0.885) VC2 0.899 (0.864,0.934)

Author and year Model	Study design and Setting	Specific Outcome	Time point of model application	Population and sample size	Candidate predictors	Determination of candidate predictors	Final predictors	Performance AUC (95% CI)
Nafilyan 2021 [10] QCOVID England	Retrospective Jan-July 2020	Suspected/ confirmed COVID-related death (ICD code) n=26,985 (0.077%)[#] deaths in Period 1 (Jan 24-Apr 30, 2020) n=13,177 (0.037%)[#] in Period 2 (May1-Jul 28, 2020) [#]% deaths in whole population (includes those with and without COVID-19	No particular time	Adults n=34.897M (ONS PH Health Linked Data Asset linked to primary care and hospital databases) Patients entered the COVID cohort on Jan 24 & follow-up till Jul 28,2020	Demographic data Comorbidity	n/a	QCOVID Age, Sex Region Ethnicity Townsend deprivation scale Accommodation BMI Comorbidity (CKD, Cancer, Chemotherapy, Immunosuppression, Learning Disability Others)	C statistic Period 1 Men 0.935 (0.933, 0.937) Women 0.945 (0.943, 0.947) Period 2 Men 0.944 (0.942, 0.946) Women 0.956 (0.954, 0.958)
Nava 2020 [33] US	Retrospective Teaching community hospital Mar-May 2020	In-hospital mortality n=101 (32.26%) ICU admission n=98 (31.3%)		Inpatients adults COVID-19 pneumonia n=313			CURB-65 Quick COVID-19 Severity Index (qCSI) Brescia- COVID Respiratory Severity Scale (BCRSS)	CURB-65 0.781 qCSI score 0.711 BCRSS prediction rule 0.663
Nicholson 2021 [34] VICE DICE US	Retrospective Metropolitan hospital network until May 19, 2020	Death TC n=111 (19.2%) VC n=99 (21.33%) Mechanical Ventilation TC n=243 (42.04%) VC n=161 (34.69%	On admission Laboratory tests within 24 hours of admission (Research Patient Data Repository)	Inpatients Adults RT-PCR+ Observed until discharge TC n=1042 VC1= 578 (1 hospital) VC2 n= 464 (4 hospitals)	Demographic, clinical, and admission laboratory data	Multivariate logistic regression analysis with backwards selection stepwise method	DICE: Age, male sex, CAD, DM, chronic statin use, SpO2:FiO2 ratio,BMI, NLR, platelet count, procalcitonin VICE: DM SpO2:FiO2 ratio, CRP, LDH	DICE TC 0.91 (0.87,0.94) VC 0.79 (0.74,0.84) VICE TC 0.84 (0.80,0.87) VC 0.86 (0.82,0.90)
Paranjape 2021 [35] Calculator for ICU transfer US	Retrospective Large metropolitan health system	Transfer to ICU service TC n=804 (39.92%) VC n= 192 (28,61%)	On admission	Inpatients Adults RT-PCR+, TC Mar-Jul n=2014 VC: July	Demographic (age, sex, race, BMI), temp, SpO2 on room air CRP, LDH, ferritin, D dimer,	Multivariate Logistic Regression analysis with backwards selection stepwise method	DM, CAD, CKD, CRP, LDH	TC 0.752 VC 0.769

Author and year Model	Study design and Setting	Specific Outcome	Time point of model application	Population and sample size	Candidate predictors	Determination of candidate predictors	Final predictors	Performance AUC (95% CI)
	Mar-Jul 2020			n=671	absolute lymphocyte count Comorbidities - HTN, DM CKD,Asthma, COPD, CAD			
Richardson 2021 [36] NEWS2/ NEWS UK	Retrospective	Death 24 hrs-9 48 hrs-15 72 hrs-33 In-hospital-199 (32.09%)	Within 24 hours of admission	Adults non-elective admission (COVID or not) discharged Mar11-Jun 13, 2020 COVID as ICD-coded in EMR	n/a	n/a	NEWS: RR, HR, temp, O2Sat, O2 supplement, AVPU, SBP NEWS2 NEWS and alertness includes confusion	NEWS 0.64 NEWS2 0.64 72hours NEWS: 0.75 NEWS2: 0.71 48 hours NEWS: 0.78 NEWS2: 0.76 24hours NEWS: 0.84 NEWS2: 0.86
Schoning 2021 [37] COSA Switzerland	Retrospective Prospective validation Feb-Nov 2020	Severe TC n=63 (31.82%) VC n=105 (22.87%)	On admission Lab values 3 days before or up to 1 day after PCR+	In- and outpatients 198 PCR+ TC n=198 Feb-Aug VC n=459 Sep-Nov	Medical history, demographics Top 20 laboratory tests routinely assessed on admission	Logistic regression Repeated cross-validation	COSA score Sex CRP Sodium Hemoglobin eGFR Glucose Leukocyte count	TC 0.94 (0.87, 0.95) VC 0.85
Solem 2021 [11] COPE Denmark	Prospective cohort 2 regions Denmark Mar 1 to June 16, 2020	TC Hospital admission n=1359 (34.5%) ICU n=181 (4.6%) Death n=324 (8.2% of COVID-19 patients or 0.01% of whole population tested) VC Hospitalized n=753 (45.63%) ICU n=131 (17.4%) Deaths n=305 (18.48%)	Different time points: Diagnosis First 12 h of hospital admission 12 hours prior to ICU admission -12 hours after ICU admission	Adults at PCR test (Regional EMR with in-hospital data) 3944 RT-PCR+ out of 2.6M tested TC Denmark VC: UK Biobank n=1650	Demographic data Comorbidity Temporal features In-hospital laboratory tests	Available information in the EMR Random forests Cross-validation	Base model-Age, sex, BMI Model 2: Base +comorbidity Model 3: Model 2+temporal features Model 4: Model 3+in-hospital laboratory tests	TC: Risk of death 0.906 at diagnosis, 0.818, at admission 0.721 at ICU admission ICU admission On diagnosis, Model 2,3 &4 with improved prediction On admission Model 4 significantly improves prediction VC On Diagnosis Mortality 0.742 ICU admission 0.529 Hospital admission 0.661

Author and year Model	Study design and Setting	Specific Outcome	Time point of model application	Population and sample size	Candidate predictors	Determination of candidate predictors	Final predictors	Performance AUC (95% CI)
Tanboga 2021 [38] CORONATION TURKEY	Retrospective National database Mar-Jun 2020	Total deaths n=2682 (4.4%) ICU n=7688 (13%) Mechanical ventilation n=4867 (8%) 30-day death TC n=2343 (6%) VC n=339 (2%) ICU TC n=6160 (15%) VC n=1528 (8%) Mechanical ventilation TC n=4158 (10%) VC n=709 (4%)	Admission	Inpatient RT-PCR+ n=60,980 TC n=41,300 VC n=19,6809	Demographic data Comorbidities Lung CT Laboratory tests	Data from public health management system for COVID-19 specific data during index hospitalization (symptoms, biomarkers, medication, comorbidity, clinical outcomes)	Age, findings from lung CT, LDH, CRP, comorbidities, NLR, and D-dimer	Temporal validation TC 0.933 (0.929–0.937) VC 0.956 (0.948–0.964) Geographic validation Istanbul TC Istanbul 0.958 (0.939–0.972) VC: Anatolia region 0.896 (0.890–0.902)
Van Dam 2020 [39] Netherlands	Retrospective study ED of a single secondary/ tertiary hospital Mar-May 2020	30-day mortality N=95/403 (23.57%)	On admission	Adults on ED consult RT-PCR+ or clinical diagnosis (symptom, CT findings with consent) n=403 307were admitted	n/a	n/a	RISE-UP CURB-65 MEWS REMS abbMEDS SOFA APACHE II	30-day mortality 0.83 (0.79-0.88) 0.75 (0.70-0.80) 0.64 (0.58-0.70) 0.73(0.68-0.78) 0.75(0.70-0.81) 0.72(0.67-0.78) 0.71(0.65-0.78)
Van Dam 2021[40] RISE UP Netherlands	Retrospective 2 EDs Mar-May 2020	30-day mortality n=167 (26%) Composite of 30-day mortality,ICU ICU n=102	During Emergency Department Visit	Adults on ED visits RT-PCR+ or clinically diagnosed (symptom, CT finding)			RISE-UP score Age, abnormal vital signs (any of HR, MAP, RR, O2Sat, temp, GCS) serum albumin BUN,	Mortality 0.77 (0.73, 0.81) Composite 0.72 (0.68,0.76)

Author and year Model	Study design and Setting	Specific Outcome	Time point of model application	Population and sample size	Candidate predictors	Determination of candidate predictors	Final predictors	Performance AUC (95% CI)
		(15.9%)					LDH, bilirubin with or without O2 supplement	
Van Klaveren 2021 [41] COPE Netherlands	Retrospective 4 hospitals Mar-Aug 2020	326 deaths (10.02%)	Admission to hospital	Admitted from ED suspected COVID 19 n=3252 Temporal validation	Patient characteristics (sex, age, BMI) Vital statistics Laboratory tests	Literature review Available at ED setting Logistic regression with post-hoc uniform shrinkage	Age, RR, CRP, LDH, alb, urea	AUC in 4 hospitals: 0.82 [0.78; 0.86] 0.82[0.74; 0.90] 0.79 [0.70;0.88] 0.83 [0.79; 0.86]

AUC = Area Under the Curve; VC=Validation cohort; TC=Training cohort

Appendix 2. Clinical appraisal of included studies using the PROBAST tool

Study	ROB				Applicability			Overall judgment	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
Adderley et al.	L	L	L	H	U	U	L	**H**	**U**
Ageno et al.	L	L	L	H	U	U	L	**H**	**U**
Artero et al.	H	U	L	H	U	U	L	**H**	**U**
Baker et al.	H	L	H	H	U	L	L	**H**	**U**
Bartoletti et al.	H	L	H	H	U	U	L	**H**	**U**
Berenguer et al.	L	L	L	H	U	U	L	**H**	**U**
Bradley et al.	H	L	H	U	U	U	L	**H**	**U**
Carr et al.	L	L	L	H	L	U	U	**H**	**U**
Castro et al.	L	L	L	H	U	U	L	**H**	**U**
Chua	H	H	U	U	U	U	L	**H**	**U**
Clift et al.	L	L	L	L	U	U	L	**L**	**U**
Codon	H	U	U	H	U	U	L	**H**	**U**
El Sohl et a.	H	L	L	H	U	U	L	**H**	**U**
Gupta 2020	H	L	L	U	U	U	L	**H**	**U**
Gupta 2021 4C	L	U	L	L	U	U	L	**U**	**U**
King et al.	L	L	L	H	L	U	L	**H**	**U**
Knight	L	L	L	U	U	U	L	**U**	**U**
Li J	H	U	H	L	U	U	L	**H**	**U**
Liu FY	H	L	L	H	U	U	L	**H**	**U**
Liu H	L	L	U	H	U	U	L	**H**	**U**
Mancilla-Galindo	H	L	L	H	L	U	L	**H**	**U**
Marcolino	L	L	L	U	U	U	L	**U**	**U**
Nafilyan	L	L	L	L	U	U	L	**L**	**U**
Nava	U	U	H	H	U	U	L	**H**	**U**
Nicholson	L	U	L	H	U	U	L	**H**	**U**
Paranjape	L	U	U	H	U	U	L	**H**	**U**
Richadson	L	L	L	H	U	U	L	**H**	**U**
Schoning	U	U	L	H	U	U	L	**H**	**U**
Solem	L	L	L	U	U	U	L	**U**	**U**
Tanboga	U	U	H	H	U	U	L	**H**	**U**
van Dam RISE UP	L	U	L	H	U	U	L	**H**	**U**
van Dam	L	L	L	H	U	U	L	**H**	**U**
van Klaveren	U	H	U	H	U	U	L	**H**	**U**

Appendix 3. GRADE Evidence Profile

Author(s): Patricia Pauline Remalante-Rayco, Evelyn Salido
Question: Prognostic models compared to no prognostic models for prediction of worsening severity/clinical deterioration/poor outcomes and mortality in adults with COVID-19
Setting:
Bibliography:

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
QCOVID									
2	observational studies	not serious	not serious	not serious	not serious	very strong association		⨁⨁⨁⨁ HIGH	
4C Deterioration Score									
1	observational studies	not serious	not serious	not serious	not serious	none		⨁⨁◯◯ LOW	
COPE									
1	observational studies	serious [a]	not serious	not serious	not serious	none		⨁◯◯◯ VERY LOW	
ABC2-SPH									
1	observational studies	serious [b]	not serious	not serious	not serious	strong association		⨁⨁◯◯ LOW	
4C Mortality Score									
2	observational studies	serious [b]	not serious	not serious	not serious	none		⨁◯◯◯ VERY LOW	
CURB-65									
5	observational studies	serious [c,d,e]	not serious	not serious	not serious	strong association		⨁⨁◯◯ LOW	
REMS									
3	observational studies	serious [c,e]	not serious	not serious	not serious	strong association		⨁⨁◯◯ LOW	
RISE-UP									
2	observational studies	serious [c,d,e]	not serious	not serious	not serious	strong association		⨁⨁◯◯ LOW	
qSOFA									
4	observational studies	serious [c,e]	not serious	not serious	not serious	none		⨁◯◯◯ VERY LOW	
MEWS									
2	observational studies	serious [e]	not serious	not serious	not serious	none		⨁◯◯◯ VERY LOW	
NEWS2									
5	observational studies	serious [c,d,e,f]	serious [g]	not serious	not serious	none		⨁◯◯◯ VERY LOW	

CI: Confidence interval

Explanations

a. Did not account for censoring
b. Continuous predictors were dichotomized
c. Exclusion of participants with missing data
d. Calibration not done
e. Complexities in data not addressed
f. Small number of events relative to number of predictors
g. Outlier study (1 study with higher AUC)

Appendix 4. Characteristics of Ongoing Studies

Study name	Methods	Participants	Interventions	Outcomes	Starting date	Notes
NCT04366024 A Novel Nomogram to Predict Severity of COVID-19	Retrospective observational study (case-only)	COVID-19 disease patients confirmed by virus nucleic acid RT-PCR and CT	clinical diagnosis	Consistency of predicted severe rate and observed severe rate of COVID-19 patients (Time frame: up to 3 months) Duration of severe illness (Time frame: up to 3 months)	January 17, 2020	recruiting

Study name	Methods	Participants	Interventions	Outcomes	Starting date	Notes
NCT04366765 COVID-19 Survival - The COVIVA Study (COVIVA)	Prospective case-control study	Patients ÷18 years old with clinically suspected or confirmed SARS-CoV-2 infection triaged to the emergency department for which swab test was performed	–	Primary: Incidence of death during index hospital stay (Time Frame: up to 30 days) Secondary: ICU admission at 30 days; invasive ventilation; need for extracorporeal membrane oxygenation; hemodynamic support; Length of ICU stay; Acute respiratory distress Syndrome; Myocardial injury; ST-segment elevation myocardial infarction; In-hospital resource use; quality of life using the EQ-5D questionnaire (Time Frame for all secondary outcomes: up to 30 days)	March 19, 2020	recruiting
NCT04321265 Outcomes and Prognostic Factors in Coronavirus Disease (COVID-19) in Very Old Intensive Care Patients (COVIP)	Prospective cohort study	Patients infected with SARS-COV-2 aged >70 years, admitted to the ICU	–	Primary: Survival (Time Frame: up to 30 days) Secondary: Fragilty (Time Frame: pre-admission) Fragilty will be measured by using the Clinical frailty scale (CFS)	March 19, 2020	Recruiting

Study name	Methods	Participants	Interventions	Outcomes	Starting date	Notes
NCT04629183 Risk Stratification of COVID-19 Patients Discharged From the Emergency Department (CODED)	Prospective cohort study	Adult patients (>18 years) subjected to a first ED visit for physician-confirmed COVID-19, discharged from the ED based on attending physician's or patient's decision (independent from study participation)	integrated clinical evaluation	Primary: Composite outcome of death (any cause), hospital admission (any cause) (Time Frame: up to 30 days) Secondary: Death from COVID-19; death from other disease; hospital admission for COVID-19; hospital admission for other disease (Time Frame for all secondary outcomes: 30 days)	November 1, 2020	Recruiting