FAST TRACK

Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal

Laure Wynants,[1,2] Ben Van Calster,[2,3] Gary S Collins,[4,5] Richard D Riley,[6] Georg Heinze,[7] Ewoud Schuit,[8,9] Marc M J Bonten,[8,10] Darren L Dahly,[11,12] Johanna A A Damen,[8,9] Thomas P A Debray,[8,9] Valentijn M T de Jong,[8,9] Maarten De Vos,[2,13] Paula Dhiman,[4,5] Maria C Haller,[7,14] Michael O Harhay,[15,16] Liesbet Henckaerts,[17,18] Pauline Heus,[8,9] Nina Kreuzberger,[19] Anna Lohmann,[20] Kim Luijken,[20] Jie Ma,[5] Glen P Martin,[21] Constanza L Andaur Navarro,[8,9] Johannes B Reitsma,[8,9] Jamie C Sergeant,[22,23] Chunhu Shi,[24] Nicole Skoetz,[19] Luc J M Smits,[1] Kym I E Snell,[6] Matthew Sperrin,[25] René Spijker,[8,9,26] Ewout W Steyerberg,[3] Toshihiko Takada,[8] Ioanna Tzoulaki,[27,28] Sander M J van Kuijk,[29] Florien S van Royen,[8] Jan Y Verbakel,[30,31] Christine Wallisch,[7,32,33] Jack Wilkinson,[22] Robert Wolff,[34] Lotty Hooft,[8,9] Karel G M Moons,[8,9] Maarten van Smeden[8]

For numbered affiliations see end of the article

Correspondence to: L Wynants laure.wynants@maastrichtuniversity.nl (ORCID 0000-0002-3037-122X)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2020;369:m1328
http://dx.doi.org/10.1136/bmj.m1328

Originally accepted: 31 March 2020

Final version accepted: 1 July 2020

ABSTRACT

OBJECTIVE
To review and appraise the validity and usefulness of published and preprint reports of prediction models for diagnosing coronavirus disease 2019 (covid-19) in patients with suspected infection, for prognosis of patients with covid-19, and for detecting people in the general population at increased risk of becoming infected with covid-19 or being admitted to hospital with the disease.

DESIGN
Living systematic review and critical appraisal by the COVID-PRECISE (Precise Risk Estimation to optimise covid-19 Care for Infected or Suspected patients in diverse sEttings) group.

DATA SOURCES
PubMed and Embase through Ovid, arXiv, medRxiv, and bioRxiv up to 5 May 2020.

STUDY SELECTION
Studies that developed or validated a multivariable covid-19 related prediction model.

DATA EXTRACTION
At least two authors independently extracted data using the CHARMS (critical appraisal and data extraction for systematic reviews of prediction modelling studies) checklist; risk of bias was assessed using PROBAST (prediction model risk of bias assessment tool).

RESULTS
14 217 titles were screened, and 107 studies describing 145 prediction models were included. The review identified four models for identifying people at risk in the general population; 91 diagnostic models for detecting covid-19 (60 were based on medical imaging, nine to diagnose disease severity); and 50 prognostic models for predicting mortality risk, progression to severe disease, intensive care unit admission, ventilation, intubation, or length of hospital stay. The most frequently reported predictors of diagnosis and prognosis of covid-19 are age, body temperature, lymphocyte count, and lung imaging features. Flu-like symptoms and neutrophil count are frequently predictive in diagnostic models, while comorbidities, sex, C reactive protein, and creatinine are frequent prognostic factors. C index estimates ranged from 0.73 to 0.81 in prediction models for the general population, from 0.65 to more than 0.99 in diagnostic models, and from 0.68 to 0.99 in prognostic models. All models were rated at high risk of bias, mostly because of non-representative selection of control patients, exclusion of patients who had not experienced the event of interest by the end of the study, high risk of model overfitting, and vague reporting. Most reports did not include any description of the study population or intended use of the models, and calibration of the model predictions was rarely assessed.

CONCLUSION
Prediction models for covid-19 are quickly entering the academic literature to support medical decision making at a time when they are urgently needed. This review indicates that proposed models are poorly reported, at high risk of bias, and their reported

WHAT IS ALREADY KNOWN ON THIS TOPIC

The sharp recent increase in coronavirus disease 2019 (covid-19) incidence has put a strain on healthcare systems worldwide; an urgent need exists for efficient early detection of covid-19 in the general population, for diagnosis of covid-19 in patients with suspected disease, and for prognosis of covid-19 in patients with confirmed disease

Viral nucleic acid testing and chest computed tomography imaging are standard methods for diagnosing covid-19, but are time consuming

Earlier reports suggest that elderly patients, patients with comorbidities (chronic obstructive pulmonary disease, cardiovascular disease, hypertension), and patients presenting with dyspnoea are vulnerable to more severe morbidity and mortality after infection

WHAT THIS STUDY ADDS

Four models identified patients at risk in the general population (using proxy outcomes for covid-19)

Ninety one diagnostic models were identified for detecting covid-19 (60 were based on medical images; nine were for severity classification); and 50 prognostic models for predicting, among others, mortality risk, progression to severe disease

Proposed models are poorly reported and at high risk of bias, raising concern that their predictions could be unreliable when applied in daily practice performance is probably optimistic. Hence, we do not recommend any of these reported prediction models for use in current practice. Immediate sharing of well documented individual participant data from covid-19 studies and collaboration are urgently needed to develop more rigorous prediction models, and validate promising ones. The predictors identified in included models should be considered as candidate predictors for new models. Methodological guidance should be followed because unreliable predictions could cause more harm than benefit in guiding clinical decisions. Finally, studies should adhere to the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) reporting guideline.

SYSTEMATIC REVIEW REGISTRATION

Protocol https://osf.io/ehc47/, registration https://osf.io/wy245.

READERS' NOTE

This article is a living systematic review that will be updated to reflect emerging evidence. Updates may occur for up to two years from the date of original publication. This version is update 2 of the original article published on 7 April 2020 (*BMJ* 2020;369:m1328), and previous updates can be found as data supplements (https://www.bmj.com/content/369/bmj.m1328/related#datasupp).

Introduction

The novel coronavirus disease 2019 (covid-19) presents an important and urgent threat to global health. Since the outbreak in early December 2019 in the Hubei province of the People's Republic of China, the number of patients confirmed to have the disease has exceeded 8 963 350 in 188 countries, and the number of people infected is probably much higher. More than 468 330 people have died from covid-19 (up to 22 June 2020).[1] Despite public health responses aimed at containing the disease and delaying the spread, several countries have been confronted with a critical care crisis, and more countries could follow.[2-4] Outbreaks lead to important increases in the demand for hospital beds and shortage of medical equipment, while medical staff themselves could also get infected.

To mitigate the burden on the healthcare system, while also providing the best possible care for patients, efficient diagnosis and information on the prognosis of the disease is needed. Prediction models that combine several variables or features to estimate the risk of people being infected or experiencing a poor outcome from the infection could assist medical staff in triaging patients when allocating limited healthcare resources. Models ranging from rule based scoring systems to advanced machine learning models (deep learning) have been proposed and published in response to a call to share relevant covid-19 research findings rapidly and openly to inform the public health response and help save lives.[5] Many of these prediction models are published in open access repositories, ahead of peer review.

We aimed to systematically review and critically appraise all currently available prediction models for covid-19, in particular models to predict the risk of developing covid-19 or being admitted to hospital with covid-19, models to predict the presence of covid-19 in patients with suspected infection, and models to predict the prognosis or course of infection in patients with covid-19. We included model development and external validation studies. This living systematic review, with periodic updates, is being conducted by the COVID-PRECISE (Precise Risk Estimation to optimise covid-19 Care for Infected or Suspected patients in diverse sEttings) group in collaboration with the Cochrane Prognosis Methods Group.

Methods

We searched PubMed and Embase through Ovid, bioRxiv, medRxiv, and arXiv for research on covid-19 published after 3 January 2020. We used the publicly available publication list of the covid-19 living systematic review.[6] This list contains studies on covid-19 published on PubMed and Embase through Ovid, bioRxiv, and medRxiv, and is continuously updated. We validated whether the list is fit for purpose (online supplementary material) and further supplemented it with studies on covid-19 retrieved from arXiv. The online supplementary material presents the search strings. Additionally, we contacted authors for studies that were not publicly available at the time of the search,[7 8] and included studies that were publicly available but not on the living systematic review[6] list at the time of our search.[9-12]

We searched databases repeatedly up to 5 May 2020 (supplementary table 1). All studies were considered, regardless of language or publication status (preprint or peer reviewed articles; updates of preprints are only included and reassessed after publication in a peer reviewed journal). We included studies if they developed or validated a multivariable model or scoring system, based on individual participant level data, to predict any covid-19 related outcome. These models included three types of prediction models: diagnostic models for predicting the presence or severity of covid-19 in patients with suspected infection; prognostic models for predicting the course of infection in patients with covid-19; and prediction models to identify people at increased risk of covid-19 in the general population. No restrictions were made on the setting (eg, inpatients, outpatients, or general population), prediction horizon (how far ahead the model predicts), included predictors, or outcomes. Epidemiological studies that aimed to model disease transmission or fatality rates, diagnostic test accuracy, and predictor finding studies were excluded. Starting with the second update, retrieved records were initially screened by a text analysis tool developed by artificial intelligence to prioritise sensitivity (supplementary material). Titles, abstracts, and full texts were screened for eligibility in duplicate by independent reviewers (pairs from LW, BVC, MvS) using EPPI-Reviewer,[13] and discrepancies were resolved through discussion.

Data extraction of included articles was done by two independent reviewers (from LW, BVC, GSC, TPAD,

MCH, GH, KGMM, RDR, ES, LJMS, EWS, KIES, CW, AL, JM, TT, JAAD, KL, JBR, LH, CS, MS, MCH, NS, NK, SMJvK, JCS, PD, CLAN, RW, GPM, IT, JYV, DLD, JW, FSvR, PH, VMTdJ, and MvS). Reviewers used a standardised data extraction form based on the CHARMS (critical appraisal and data extraction for systematic reviews of prediction modelling studies) checklist[14] and PROBAST (prediction model risk of bias assessment tool) for assessing the reported prediction models.[15] We sought to extract each model's predictive performance by using whatever measures were presented. These measures included any summaries of discrimination (the extent to which predicted risks discriminate between participants with and without the outcome), and calibration (the extent to which predicted risks correspond to observed risks) as recommended in the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) statement.[16] Discrimination is often quantified by the C index (C index=1 if the model discriminates perfectly; C index=0.5 if discrimination is no better than chance). Calibration is often quantified by the calibration intercept (which is zero when the risks are not systematically overestimated or underestimated) and calibration slope (which is one if the predicted risks are not too extreme or too moderate).[17] We focused on performance statistics as estimated from the strongest available form of validation (in order of strength: external (evaluation in an independent database), internal (bootstrap validation, cross validation, random training test splits, temporal splits), apparent (evaluation by using exactly the same data used for development)). Any discrepancies in data extraction were discussed between reviewers, and remaining conflicts were resolved by LW and MvS. The online supplementary material provides details on data extraction. We considered aspects of PRISMA (preferred reporting items for systematic reviews and meta-analyses)[18] and TRIPOD[16] in reporting our article.

Patient and public involvement

It was not possible to involve patients or the public in the design, conduct, or reporting of our research. The study protocol and preliminary results are publicly available on https://osf.io/ehc47/ and medRxiv.

Results

We retrieved 14 209 titles through our systematic search (of which 9306 were included in the present update; supplementary table 1, fig 1). Two additional unpublished studies were made available on request (after a call on social media). We included a further six studies that were publicly available but were not detected by our search. Of 14 217 titles, 275 studies were retained for abstract and full text screening (of which 76 in the present update). One hundred seven studies describing 145 prediction models met the inclusion criteria (of which 56 papers and 79 models added in the present update, supplementary table 1).[7-12 19-119] These studies were selected for data extraction and critical appraisal (table 1, table 2, table 3, and table 4).

Primary datasets

Forty five studies used data on patients with covid-19 from China (supplementary table 2), six from Italy,[32 39 72 74 76 79] three from Brazil,[69 81 109] three from France,[71 77 110] three from the United States,[96 108 112] two from South Korea,[63 80] one from Belgium,[82] one from the Netherlands,[95] one from the United Kingdom,[75] one from Israel,[67] one from Mexico,[70] and one from Singapore.[40] Twenty two studies used international data (supplementary table 2) and two studies used simulated data.[35 41] Three studies used proxy data to estimate covid-19 related risks (eg, Medicare claims data from 2015 to 2016).[8 90 113] Twelve studies were not clear on the origin of covid-19 data (supplementary table 2).

Based on 59 studies that reported study dates, data were collected between 8 December 2019 and 21 April 2020. Four studies reported median follow-up time (4.5, 8.4, 15, and 18 days),[20 37 83 108] while another study reported a follow-up of at least five days.[42] Some centres provided data to multiple studies and several studies used open Github[120] or Kaggle[121] data repositories (version or date of access often unspecified), and so it was unclear how much these datasets overlapped across our identified studies (supplementary table 2). One study[25] developed prediction models for use in paediatric patients. The median age in studies on adults varied from 34 to 68 years, and the proportion of men varied from 35% to 75%, although this information was often not reported at all (supplementary table 2).

Among the studies that developed prognostic models to predict mortality risk in people with confirmed or suspected infection, the percentage of deaths varied between 1% and 59% (table 3). This wide variation is partly because of substantial sampling bias caused by studies excluding participants who still had the disease at the end of the study period (that is, they had neither recovered nor died).[7 21-23 44 96 98 100] Additionally, length of follow-up could have varied between studies (but was rarely reported), and there might be local and temporal variation in how people were diagnosed as having covid-19 or were admitted to the hospital (and therefore recruited for the studies). Among the diagnostic model studies, only nine reported on the prevalence of covid-19 and used a cross sectional or cohort design; the prevalence varied between 17% and 79% (table 2). Because 58 diagnostic studies used either case-control sampling or an unclear method of data collection, the prevalence in these diagnostic studies might not have been representative of their target population.

Table 1, table 2, and table 3 give an overview of the 145 prediction models reported in the 107 identified studies. Supplementary table 2 provides modelling details and box 1 discusses the availability of models in a format for use in clinical practice.

Models to predict risks of covid-19 in the general population

We identified four models that predicted risk of covid-19 in the general population. Three models from one study used hospital admission for non-

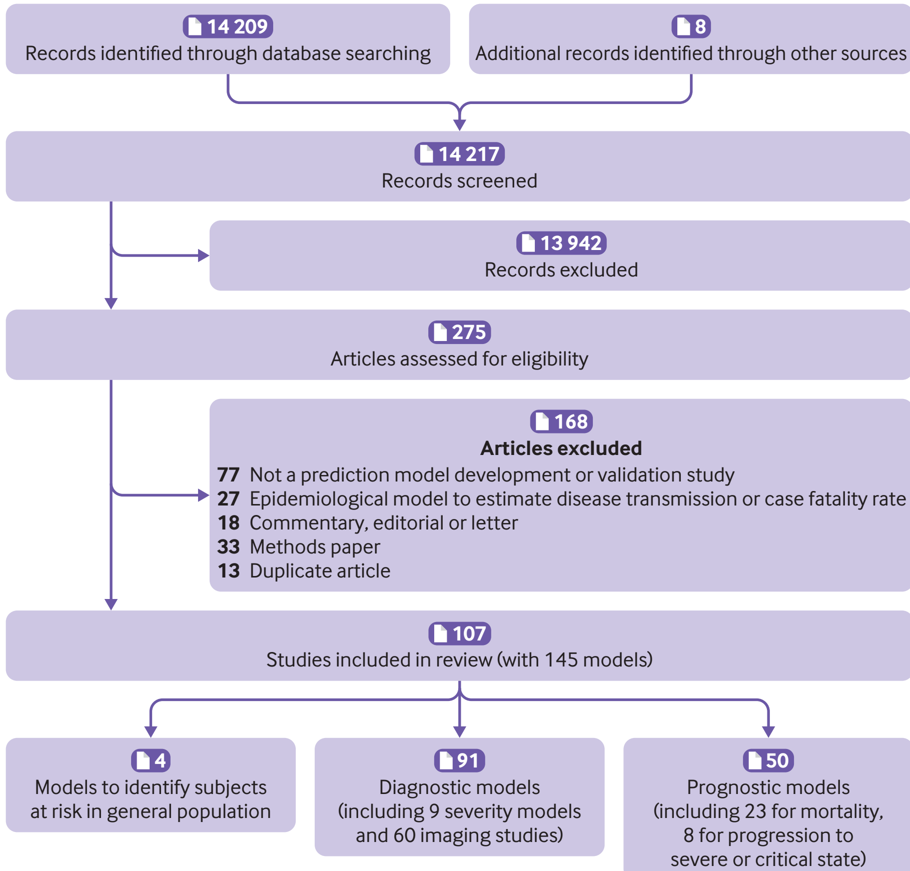

Fig 1 | PRISMA (preferred reporting items for systematic reviews and meta-analyses) flowchart of study inclusions and exclusions

tuberculosis pneumonia, influenza, acute bronchitis, or upper respiratory tract infections as proxy outcomes in a dataset without any patients with covid-19.[8] Among the predictors were age, sex, previous hospital admissions, comorbidity data, and social determinants of health. The study reported C indices of 0.73, 0.81, and 0.81. A fourth model used deep learning on thermal videos from the faces of people wearing facemasks to determine abnormal breathing (not covid related) with a reported sensitivity of 80%.[90]

Diagnostic models to detect covid-19 in patients with suspected infection

We identified 22 multivariable models to diagnose covid-19. Most models targeted patients with suspected covid-19. Reported C index values ranged between 0.65 and 0.99. A few models also evaluated calibration and reported good results.[69 78 117] The most frequently used diagnostic predictors (at least 10 times) were flu-like signs and symptoms (eg, shiver, fatigue), imaging features (eg, pneumonia signs on computed tomography scan), age, body temperature, lymphocyte count, and neutrophil count (table 2).

Nine studies aimed to diagnose severe disease in patients with covid-19: eight in adults with covid-19 with reported C indices between value of 0.80 and 0.99, and one in paediatric patients with reported perfect performance.[25] Predictors of severe covid-19 used more than once were comorbidities, liver enzymes, C reactive protein, imaging features, and neutrophil count.

Sixty prediction models were proposed to support the diagnosis of covid-19 or covid-19 pneumonia (and some also to monitor progression) based on images. Most studies used computed tomography images or chest radiographs. Others used spectrograms of cough sounds[53] and lung ultrasound.[73] The predictive performance varied widely, with estimated C index values ranging from 0.81 to more than 0.99.

Prognostic models for patients with diagnosis of covid-19

We identified 50 prognostic models (table 3) for patients with a diagnosis of covid-19. The intended use of these models (that is, when to use them, and for whom) was often not clearly described. Prediction horizons varied between one and 30 days, but were often unspecified.

Of these models, 23 estimated mortality risk and eight aimed to predict progression to a severe or critical state (table 3). The remaining studies used other outcomes (single or as part of a composite) including recovery, length of hospital stay, intensive care unit admission, intubation, (duration of) mechanical ventilation, and acute respiratory distress syndrome. One study used data from 2015 to 2019 to predict mortality and prolonged assisted mechanical ventilation (as a non-covid-19 proxy outcome).[113]

Table 1 | Overview of prediction models for use in the general population

Study; setting; and outcome	Predictors in final model	Sample size: total No of participants for model development set (No with outcome)	Predictive performance on validation: Type of validation*	Predictive performance on validation: Sample size: total No of participants for model validation (No with outcome)	Predictive performance on validation: Performance* (C index, sensitivity (%), specificity (%), PPV/NPV (%), calibration slope, other (95% CI, if reported))	Overall risk of bias using PROBAST
General population						
Original review						
Decaprio et al[8]; data from US general population; hospital admission for covid-19 pneumonia (proxy events)†	Age, sex, number of previous hospital admissions, 11 diagnostic features, interactions between age and diagnostic features	1.5 million (unknown)	Training test split	369 865 (unknown)	C index 0.73	High
Decaprio et al[8]; data from US general population; hospital admission for covid-19 pneumonia (proxy events)†	Age and ≥500 features related to diagnosis history	1.5 million (unknown)	Training test split	369 865 (unknown)	C index 0.81	High
Decaprio et al[8]; data from US general population; hospital admission for covid-19 pneumonia (proxy events)†	≥500 undisclosed features, including age, diagnostic history, social determinants of health, Charlson comorbidity index	1.5 million (unknown)	Training test split	369 865 (unknown)	C index 0.81	High
Update 2						
Jiang et al[90]; data from China, respiratory patients versus healthy volunteers; detection of respiratory diseases such as covid-19	Infrared/thermal video of face	Unknown	Training test split	Not applicable	Sensitivity 80, PPV 90	High

NPV=negative predictive value; PPV=positive predictive value; PROBAST=prediction model risk of bias assessment tool.
*Performance is given for the strongest form of validation reported. This is indicated in the column "type of validation." When a training test split was used, performance on the test set is reported. Apparent performance is the performance observed in the development data.
†Proxy events used: pneumonia (except from tuberculosis), influenza, acute bronchitis, or other specified upper respiratory tract infections (no patients with covid-19 pneumonia in data).

The most frequently used prognostic factors (for any outcome, included at least 10 times) included comorbidities, age, sex, lymphocyte count, C reactive protein, body temperature, creatinine, and imaging features (table 3).

Studies that predicted mortality reported C indices between 0.68 and 0.98. Some studies also evaluated calibration.[7 67 116] When applied to new patients, the model by Xie et al yielded probabilities of mortality that were too high for low risk patients and too low for high risk patients (calibration slope >1), despite excellent discrimination.[7] The mortality model by Zhang et al also showed miscalibrated (overfitted and underestimated) risks at external validation,[116] while the model by Barda et al showed underfitting.[67]

The studies that developed models to predict progression to a severe or critical state reported C indices between 0.73 and 0.99. Three of these studies also reported good calibration, but this was evaluated internally (eg, bootstrapped)[88] or in an unclear way.[83 119]

Reported C indices for other outcomes varied between 0.72 and 0.96. Singh et al and Zhang et al also evaluated calibration externally (in new patients). Singh showed that the Epic Deterioration Index overestimated the risk or a poor outcome, while the poor outcome model by Zhang et al underestimated the risk of a poort outcome.[108 116]

Risk of bias

All studies were at high risk of bias according to assessment with PROBAST (table 1, table 2, and table 3), which suggests that their predictive performance when used in practice is probably lower than that reported. Therefore, we have cause for concern that the predictions of the proposed models are unreliable when used in other people. Box 2 gives details on common causes for risk of bias for each type of model.

Fifty three of the 107 studies had a high risk of bias for the participants domain (table 4), which indicates that the participants enrolled in the studies might not be representative of the models' targeted populations. Unclear reporting on the inclusion of participants prohibited a risk of bias assessment in 26 studies. Fifteen of the 107 studies had a high risk of bias for the predictor domain, which indicates that predictors were not available at the models' intended time of use, not clearly defined, or influenced by the outcome measurement. One diagnostic imaging study used a simple scoring rule and was scored at low predictor risk of bias. The diagnostic model studies that used medical images as predictors in artificial intelligence were all scored as unclear on the predictor domain. The publications often lacked clear information on the preprocessing steps (eg, cropping of images). Moreover, complex machine learning algorithms transform images into predictors in a complex way, which makes it challenging to fully apply the PROBAST predictors section for such imaging studies. Most studies used outcomes that are easy to assess (eg, death, presence of covid-19 by laboratory confirmation). Nonetheless, there was cause for concern about bias induced by the outcome measurement in 19 studies, for example due to the use of subjective or proxy outcomes (eg, non covid-19 severe respiratory infections).

All but one of these studies[50] were at high risk of bias for the analysis domain (table 4). Many

Table 2 | Overview of prediction models for diagnosis of covid-19

Study; setting; and outcome	Predictors in final model	Sample size: total No of participants for model development set (No with outcome)	Predictive performance on validation: Type of validation*	Predictive performance on validation: Sample size: total No of participants for model validation (No with outcome)	Predictive performance on validation: Performance* (C index, sensitivity (%), specificity (%), PPV/NPV (%), calibration slope, other (95% CI, if reported))	Overall risk of bias using PROBAST
Diagnosis						
Original review						
Feng et al[10]; data from China, patients presenting at fever clinic; suspected covid-19 pneumonia	Age, temperature, heart rate, diastolic blood pressure, systolic blood pressure, basophil count, platelet count, mean corpuscular haemoglobin content, eosinophil count, monocyte count, fever, shiver, shortness of breath, headache, fatigue, sore throat, fever classification, interleukin 6	132 (26)	Temporal validation	32 (unclear)	C index 0.94	High
Lopez-Rincon et al[35]; data from international genome sequencing data repository, target population unclear; covid-19 diagnosis	Specific sequences of base pairs	553 (66)	10-fold cross validation	Not applicable	C index 0.98, sensitivity100, specificity 99	High
Meng et al[12]; data from China, asymptomatic patients with suspected covid-19; covid-19 diagnosis	Age, activated partial thromboplastin time, red blood cell distribution width SD, uric acid, triglyceride, serum potassium, albumin/globulin, 3-hydroxybutyrate, serum calcium	620 (302)	External validation	145 (80)	C index 0.87‡	High
Song et al[31]; data from China, inpatients with suspected covid-19; covid-19 diagnosis	Fever, history of close contact, signs of pneumonia on CT, neutrophil to lymphocyte ratio, highest body temperature, sex, age, meaningful respiratory syndromes	304 (73)	Training test split	95 (18)	C index 0.97 (0.93 to 1.00)	High
Update 1						
Martin et al[41]; simulated patients with suspected covid-19; covid-19 diagnosis	Unknown	Not applicable	External validation only (simulation)	Not applicable	Sensitivity 97, specificity 96	High
Sun et al[40]; data from Singapore, patients with suspected infection presenting at infectious disease clinic; covid-19 diagnosis	Age, sex, temperature, heart rate, systolic blood pressure, diastolic blood pressure, sore throat	292 (49)	Leave-one-out cross validation	Not applicable	C index 0.65 (0.57 to 0.73)	High
Sun et al[40]; data from Singapore, patients with suspected infection presenting at infectious disease clinic; covid-19 diagnosis	Sex, temperature, heart rate, respiration rate, diastolic blood pressure, sore throat, sputum production, shortness of breath, gastrointestinal symptoms, lymphocytes, neutrophils, eosinophils, creatinine	292 (49)	Leave-one-out cross validation	Not applicable	C index 0.88 (0.83 to 0.93)	High
Sun et al[40]; data from Singapore, patients with suspected infection presenting at infectious disease clinic; covid-19 diagnosis	Sex, temperature, heart rate, respiration rate, diastolic blood pressure, sputum production, gastrointestinal symptoms, chest radiograph or CT scan suggestive of pneumonia, neutrophils, eosinophils, creatinine	292 (49)	Leave-one-out cross validation	Not applicable	C index 0.88 (0.83 to 0.93)	High
Sun et al[40]; data from Singapore, patients with suspected infection presenting at infectious disease clinic; covid-19 diagnosis	Sex, covid-19 case contact, travel to Wuhan, travel to China, temperature, heart rate, respiration rate, diastolic blood pressure, sore throat, sputum production, gastrointestinal symptoms, chest radiograph or CT scan suggestive of pneumonia, neutrophils, eosinophils, creatinine, sodium	292 (49)	Leave-one-out cross validation	Not applicable	C index 0.91 (0.86 to 0.96)	High
Wang et al[43]; data from China, patients with suspected covid-19; covid-19 pneumonia	Epidemiological history, wedge shaped or fan shaped lesion parallel to or near the pleura, bilateral lower lobes, ground glass opacities, crazy paving pattern, white blood cell count	178 (69)	External validation	116 (68)	C index 0.85, calibration slope 0.56	High
Wu et al[45]; data from China, inpatients with suspected covid-19; covid-19 diagnosis	Lactate dehydrogenase, calcium, creatinine, total protein, total bilirubin, basophil, platelet distribution width, kalium, magnesium, creatinine kinase isoenzyme, glucose	108 (12)	Training test split	107 (61)	C index 0.99, sensitivity 100, specificity 94	High
Update 2						
Batista et al[69]; data from Brazil, inpatients with suspected covid-19 admitted to the emergency care department; covid-19 diagnosis	Age, sex, haemoglobin, platelets, red blood cells, mean corpuscular haemoglobin concentration, mean corpuscular haemoglobin, red cell distribution width , mean corpuscular volume, leukocytes, lymphocytes, monocytes, basophils, eosinophils and C reactive protein	234 (102)	Training test split	31 (unknown)	C index 0.85, sensitivity 68, specificity 85	High

(*Continued*)

Table 2 | Continued

Study; setting; and outcome	Predictors in final model	Sample size: total No of participants for model development set (No with outcome)	Predictive performance on validation: Type of validation*	Sample size: total No of participants for model validation (No with outcome)	Performance* (C index, sensitivity (%), specificity (%), PPV/NPV (%), calibration slope, other (95% CI, if reported))	Overall risk of bias using PROBAST
Brinati et al[74]; data from Italy, inpatients with suspected covid-19; covid-19 diagnosis	Age, aspartate aminotransferase, lymphocytes, lactodehydrogenase, PCR, WBC count, eosinophils, alanine transaminase, neutrophils, gamma-glutamyltransferase, monocytes, basophils, alkaline phosphatase, platelets	279 (102)	Training test split	56 (20)	C index 0.84, sensitivity 92, specificity 65	High
Brinati et al[74]; data from Italy, inpatients with suspected covid-19; covid-19 diagnosis	Age, aspartate aminotransferase, lymphocytes, lactodehydrogenase, PCR, WBC count, eosinophils, alanine transaminase, neutrophils, gamma-glutamyltransferase, monocytes, basophils, alkaline phosphatase, platelets	279 (102)	Training test split	56 (20)	Sensitivity 95, specificity 75, PPV 86	High
Chen et al[78]; data from China, inpatients with suspected covid-19; covid-19 diagnosis	Total number of mixed GGO in peripheral area, Tree-in-bud, offending vessel augmentation in lesions, respiration, heart ratio, temperature, WBC count, cough, fatigue, lymphocyte count	98 (51)	Training test split	38 (19)	C index 0.94 (0.87 to 1.00), sensitivity 74, specificity 79	High
Diaz-Quijano et al[81]; data from Brazil, inpatients with suspected covid-19; covid-19 diagnosis	Age, days after reporting first confirmed case in federal unit, fever, cough, sore throat, diarrhoea, coryza, chills, pulmonary manifestation, other signs, HIV, kidney diseaes, trip outside Brazil up to 14 days before onset	1243 (541)	External validation (new centres, Brazil)	4192 (785)	C index 0.73 (0.71 to 0.75), sensitivity 46, specificity 80	High
Kurstjens et al[95]; data from The Netherlands, inpatients with suspected covid-19; covid-19 diagnosis	Age, sex, CRP, LD, ferritin, absolute neutrophil count, absolute lymphocyte count, chest radiograph	375 (276)	External (Unclear)	592 (393)	C index 0.91 (0,89 to 0,94)	High
Mei et al[101]; data from China; inpatients with suspected covid-19; covid-19 diagnosis	Age, sex, CT imaging, exposure history, symptoms (present or absent of fever, cough and/or sputum), WBC counts, neutrophil count, percentage neutrophils, lymphocyte counts, percentage lymphocytes	534 (242)	Training test split	279 (134)	C index 0.92 (0.89 to 0.95), sensitivity 84 (77 to 90), specificity 83 (76 to 89), PPV 81.9 (76 to 87), NPV 85 (79 to 90)	High
Menni et al[102]; data from UK and USA, suspected covid-19; covid-19 diagnosis	Age, sex, loss of smell and taste, severe or significant persistent cough, severe fatigue, skipped meals	12 510 (5162)	External validation (new centres, USA)	2763 (726)	C index 0.76 (0.74 to 0.78), sensitivity 66 (62 to 69), specificity 83 (82 to 85), PPV 58 (55 to 62), NPV 87 (86 to 89)	High
Soares et al[109]; data from Brazil; patients with suspected infection presenting at triage centre; covid-19 diagnosis	Age, red blood cells, mean corpuscular volume, mean corpuscular haemoglobin concentration, mean corpuscular haemoglobin, red blood cell distribution width, leukocytes, basophils, monocytes, lymphocytes, platelets, mean platelet volume, creatinine, potassium, sodium, CRP	599 (81)	Repeated 10-fold cross validation	Not applicable	C index 0.87 (0.86 to 0.88), sensitivity 70 (67 to 73), specificity 86 (85 to 87), NPV 95 (94 to 95), PPV 45 (43 to 47)	High
Tordjman et al[110]; data from France; suspected patients; covid-19 diagnosis	Eosinophils, lymphocytes, neutrophils, basophils	100 (50)	External validation (new centres, France)	300 (208)	C index 0.89 (0.85 to 0.93), sensitivity 80, specificity 85, PPV 92	High
Zhao et al[117]; data from China; inpatients with suspected covid-19; covid-19 diagnosis	Fever, chest CT, CRP, PCT, WBC	547 (unknown)	Training test split	275 (unknown)	C index 0.97 (0.96 to 0.97)	High
Diagnostic severity classification						
Original review						
Yu et al[25]; data from China, paediatric inpatients with confirmed covid-19; severe disease (yes/no) defined based on clinical symptoms	Direct bilirubin, alanine transaminase	105 (8)	Apparent performance only	Not applicable	F1 score 1.00	High
Update 1						
Zhou et al[46]; data from China, inpatients with confirmed covid-19; severe pneumonia	Age, sex, onset-admission time, high blood pressure, diabetes, CHD, COPD, white blood cell counts, lymphocyte, neutrophils, alanine transaminase, aspartate aminotransferase, serum albumin, serum creatinine, blood urea nitrogen, CRP	250 (79)	Training test split	127 (38)	C index 0.88 (0.94 to 0.92), sensitivity 89, specificity 74	High

(*Continued*)

Table 2 | Continued

Study; setting; and outcome	Predictors in final model	Sample size: total No of participants for model development set (No with outcome)	Predictive performance on validation: Type of validation*	Sample size: total No of participants for model validation (No with outcome)	Performance* (C index, sensitivity (%), specificity (%), PPV/NPV (%), calibration slope, other (95% CI, if reported))	Overall risk of bias using PROBAST
Update 2						
Benchoufi et al[71]; data from France, inpatients with suspected or confirmed covid-19; Lung injury severity (pathologic vs normal)	Lung ultrasound scores for 8 quadrants in a global score	90 (unknown)	Internal validation by resampling (bootstrap)	Not applicable	C index 0.93, sensitivity 95, specificity 83	High
Chassagnon et al[77]; data from France, inpatients with confirmed covid-19; severe covid-19	Unclear	50 (unknown)	External validation (new centres, France)	130 (unknown)	C index 0.80, sensitivity 69, specificity 79	High
Li et al[97]; data from China, target population unclear; severe covid-19	Portion of infection, average infection Hounsfield unit, a measure of radio density	196 (32)	Apparent performance only	Not applicable	C index 0.97 (0.94 to 0.98), sensitivity 94 (87 to 98), specificity 88 (85 to 91)	High
Lyu et al[99]; data from China, target population unclear; severe/critical covid-19 pneumonia	Unclear	51 (39)	Apparent performance only	Not applicable	C index 0.99 (0.88 to 1.00), sensitivity 90, specificity 100	High
Lyu et al[99]; data from China , target population unclear; critical covid-19 pneumonia	Unclear	39 (24)	Apparent performance only	Not applicable	C index 0.92 (0.73 to 0.99), sensitivity 92, specificity 87	High
Wang et al[114]; data from China, inpatients with confirmed covid-19; severe covid-19	Neutrophil-to-lymphocyte ratio, red cell volume distribution width	45 (10)	Apparent performance only	Not applicable	C index 0.94 (0.90 to 0.97), sensitivity 90, specificity 85, PPV 52, NPV 96	High
Zhu et al[118]; data from China, inpatients with confirmed covid-19; severe covid-19	Peripheral blood cytokine IL-6, CRP, hypertension	127 (16)	Apparent performance only	Not applicable	C index 0.90 (0.83 to 0.97), sensitivity 100 (79 to 100), specificity 66 (56 to 75)	High
Diagnostic imaging						
Original review						
Barstugan et al[32]; data from Italy, patients with suspected covid-19; covid-19 diagnosis	Not applicable	53 (not applicable)	Cross validation	Not applicable	Sensitivity 93, specificity 100	High
Chen et al[27]; data from China, people with suspected covid-19 pneumonia; covid-19 pneumonia	Not applicable	106 (51)	Training test split	27 (11)	Sensitivity 100, specificity 82	High
Gozes et al[26]; data from China and US,§ patients with suspected covid-19; covid-19 diagnosis	Not applicable	50 (unknown)	External validation with Chinese cases and US controls	Unclear	C index 0.996 (0.989 to 1.000)	High
Jin et al[11]; data from China, US, and Switzerland,¶ patients with suspected covid-19; covid-19 diagnosis	Not applicable	416 (196)	Training test split	1255 (183)	C index 0.98, sensitivity 94, specificity 95	High
Jin et al[33]; data from China, patients with suspected covid-19; covid-19 pneumonia	Not applicable	1136 (723)	Training test split	282 (154)	C index: 0.99, sensitivity 97, specificity 92	High
Li et al[34]; data from China, patients with suspected covid-19; covid-19 diagnosis	Not applicable	2969 (400)	Training test split	353 (68)	C index 0.96 (0.94 to 0.99), sensitivity 90 (83 to 94), specificity 96 (93 to 98)	High
Shan et al[29]; data from China, people with confirmed covid-19; segmentation and quantification of infection regions in lung from chest CT scans	Not applicable	249 (not applicable)	Training test split	300 (not applicable)	Dice similarity coefficient 91.6%**	High
Shi et al[36]; data from China, target population unclear; covid-19 pneumonia	Five categories of location features from imaging: volume, number, histogram, surface, radiomics	2685 (1658)	Fivefold cross validation	Not applicable	C index 0.94	High

(Continued)

Table 2 | Continued

Study; setting; and outcome	Predictors in final model	Sample size: total No of participants for model development set (No with outcome)	Predictive performance on validation: Type of validation*	Predictive performance on validation: Sample size: total No of participants for model validation (No with outcome)	Predictive performance on validation: Performance* (C index, sensitivity (%), specificity (%), PPV/NPV (%), calibration slope, other (95% CI, if reported))	Overall risk of bias using PROBAST
Wang et al[30]; data from China, target population unclear; covid-19 diagnosis	Not applicable	259 (79)	Internal, other images from same people	Not applicable	C index 0.81 (0.71 to 0.84), sensitivity 83, specificity 67	High
Xu et al[28]; data from China, target population unclear; covid-19 diagnosis	Not applicable	509 (110)	Training test split	90 (30)	Sensitivity 87, PPV 81	High
Song et al[24]; data from China, target population unclear; diagnosis of covid-19 *v* healthy controls	Not applicable	123 (61)	Training test split	51 (27)	C index 0.99	High
Song et al[24]; data from China, target population unclear; diagnosis of covid-19 *v* bacterial pneumonia	Not applicable	131 (61)	Training test split	57 (27)	C index 0.96	High
Zheng et al[38]; data from China, target population unclear; covid-19 diagnosis	Not applicable	Unknown	Temporal validation	Unknown	C index 0.96	High
Update 1						
Abbas et al[47]; data from repositories (origin unspecified), target population unclear; covid-19 diagnosis	Not applicable	137 (unknown)	Training test split	59 (unknown)	C index 0.94, sensitivity 98, specificity 92	High
Apostolopoulos et al[48]; data from repositories (US, Italy); patients with suspected covid-19; covid-19 diagnosis	Not applicable	1427 (224)	Tenfold cross validation	Not applicable	Sensitivity 99, specificity 97	High
Bukhari et al[49]; data from Canada and US; patients with suspected covid-19; covid-19 diagnosis	Not applicable	223 (unknown)	Training test split	61 (17)	Sensitivity 98, PPV 91	High
Chaganti et al[50]; data from Canada, US, and European countries; patients with suspected covid-19; percentage lung opacity	Not applicable	631 (not applicable)	Training test split	100 (not applicable)	Correlation§§ 0.98	High
Chaganti et al[50]; data from Canada, US, and European countries; patients with suspected covid-19; percentage high lung opacity	Not applicable	631 (not applicable)	Training test split	100 (not applicable)	Correlation§§ 0.98	High
Chaganti et al[50]; data from Canada, US, and European countries; patients with suspected covid-19; severity score	Not applicable	631 (not applicable)	Training test split	100 (not applicable)	Correlation§§ 0.97	High
Chaganti et al[50]; data from Canada, US, and European countries; patients with suspected covid-19; lung opacity score	Not applicable	631 (not applicable)	Training test split	100 (not applicable)	Correlation§§ 0.97	High
Chowdhury et al[39]; data from repositories (Italy and other unspecified countries), target population unclear; covid-19 *v* "normal"	Not applicable	Unknown	Fivefold cross validation	Not applicable	C index 0.99	High
Chowdhury et al[39]; data from repositories (Italy and other unspecified countries), target population unclear; covid-19 *v* "normal" and viral pneumonia	Not applicable	Unknown	Fivefold cross validation	Not applicable	C index 0.98	High
Chowdhury et al[39]; data from repositories (Italy and other unspecified countries), target population unclear; covid-19 *v* "normal"	Not applicable	Unknown	Fivefold cross validation	Not applicable	C index 0.998	High

(*Continued*)

Table 2 | Continued

Study; setting; and outcome	Predictors in final model	Sample size: total No of participants for model development set (No with outcome)	Predictive performance on validation: Type of validation*	Sample size: total No of participants for model validation (No with outcome)	Performance* (C index, sensitivity (%), specificity (%), PPV/NPV (%), calibration slope, other (95% CI, if reported))	Overall risk of bias using PROBAST
Chowdhury et al[39]; data from repositories (Italy and other unspecified countries), target population unclear; covid-19 *v* "normal" and viral pneumonia	Not applicable	Unknown	Fivefold cross validation	Not applicable	C index 0.99	High
Fu et al[51]; data from China, target population unclear; covid-19 diagnosis	Not applicable	610 (100)	External validation	309 (50)	C index 0.99, sensitivity 97, specificity 99	High
Gozes et al[52]; data from China, people with suspected covid-19; covid-19 diagnosis	Not applicable	50 (unknown)	External validation	199 (109)	C index 0.95 (0.91 to 0.99)	High
Imran et al[53]; data from unspecified source, target population unclear; covid-19 diagnosis	Not applicable	357 (48)	Twofold cross validation	Not applicable	Sensitivity 90, specificity 81	High
Li et al[54]; data from China, inpatients with confirmed covid-19; severe and critical covid-19	Severity score based on CT scans	Not applicable	External validation of existing score	78 (not applicable)	C index 0.92 (0.84 to 0.99)	High
Li et al[55]; data from unknown origin, patients with suspected covid-19; covid-19	Not applicable	360 (120)	Training test split	135 (45)	C index 0.97	High
Hassanien et al[56]; data from repositories (origin unspecified), people with suspected covid-19; covid-19 diagnosis	Not applicable	Unknown	Training test split	Unknown	Sensitivity 95, specificity 100	High
Tang et al[57]; data from China, patients with confirmed covid-19; covid-19 severe *v* non-severe	Not applicable	176 (55)	Threefold cross validation	Not applicable	C index 0.91, sensitivity 93, specificity 75	High
Wang et al[42]; data from China, inpatients with suspected covid-19; covid-19	Not applicable	709 (560)	External validation in other centres	508 (223)	C index (average) 0.87	High
Zhang et al[58]; data from repositories (origin unspecified), people with suspected covid-19; covid-19	Not applicable	1078 (70)	Twofold cross validation	Not applicable	C index 0.95, sensitivity 96, specificity 71	High
Zhou et al[59]; data from China, patients with suspected covid-19; covid-19 diagnosis	Not applicable	191 (35)	External validation in other centres	107 (57)	C index 0.92, sensitivity 83, specificity 86	High
Update 2						
Angelov et al[64]; data from unknown origin; covid-19 diagnosis	Not applicable	Unknown	Apparent performance only	Not applicable	C index 0.89, sensitivity 89, PPV 90	High
Arpan et al[65]; data from repositories (multiple countries); covid-19 diagnosis	Not applicable	3516 (80)	Training test split	424 (19)	C index >0.99, sensitivity 100, PPV 94	High
Bai et al[66]; data from China and US, target population unclear; covid-19 diagnosis	Not applicable	830 (377)	Training test split	119 (42)	C index 0.95, sensitivity 95 (83 to 100) , specificity 96 (90 to 99)	High
Bassi et al[68]; data from Italy, target population unclear, covid-19 diagnosis	Not applicable	Unknown	Training test split	Unknown	Sensitivity 98, PPV 98	High
Borghesi et al[72]; data from Italy, target population unclear, ; severity of COVID-19 pneumonia	Sum score for lung abnormalities based on chest radiograph	Not applicable	External validation only	100 (unknown)	Agreement, kappa 0.82 (0.79 to 0.86)	High
Born et al[73]; data from repositories (origin unspecified), target population unclear, covid-19 diagnosis	Not applicable	64 (37)	Fivefold cross validation	Not Applicable	C index 0.94 (0.82 to 1.00), sensitivity 96, specificity 79	High

(*Continued*)

Table 2 | Continued

Study; setting; and outcome	Predictors in final model	Sample size: total No of participants for model development set (No with outcome)	Predictive performance on validation: Type of validation*	Predictive performance on validation: Sample size: total No of participants for model validation (No with outcome)	Predictive performance on validation: Performance* (C index, sensitivity (%), specificity (%), PPV/NPV (%), calibration slope, other (95% CI, if reported))	Overall risk of bias using PROBAST
Castiglioni et al[76]; data from Italy, inpatients suspected of covid-19; covid-19 diagnosis	Not applicable	500 (250)	Temporal validation	110 (36)	C index 0.81 (0.73 to 0.87), sensitivity 80 (72 to 86), specificity 81 (73 to 87), PPV 89 (82 to 94), NPV 66 (57 to 75)	High
Guiot et al[82]; data from Belgium, inpatients suspected of covid-19; covid-19 diagnosis	30 radiomics features	727 (unknown)	Training test split	165 (unknown)	C index 0.94 (0.88 to 1.00) , sensitivity 79, specificity 91, PPV 54, NPV 97	High
Hu et al[86]; data from unknown origin, target population unclear; covid-19 diagnosis	Not applicable	629 (313)	Training test split	201 (104)	C index 92 (84 to 100), sensitivity 86, specificity 85	High
Islam et al[87]; data from unknown origin, inpatients suspected of covid-19; covid-19 diagnosis	Not applicable	16130 (98)	Unknown origin	210 (10)	Sensitivity 80	High
Kana et al[91]; data from unknown origin, target population unclear; covid-19 diagnosis	Not applicable	5092 (161)	External validation, different repository (unknown origin)	600 (200)	Sensitivity 100, specificity 100	High
Karim et al[92]; data from unknown origin, target population unclear; covid-19 diagnosis	Not applicable	Unknown	Unknown	Unknown	Severe inconsistencies in reported performance: data not extracted	High
Khan et al[93]; data from unknown origin, target population unclear; covid-19 diagnosis	Not applicable	1300 (284)	Training test split	221 (30)	Sensitivity 100, PPV 97	High
Kumar et al[94]; data from USA, China and Italy, target population unclear; covid-19 diagnosis; covid-19 diagnosis	Not applicable	Unknown	Apparent performance only	Not applicable	C index 0.997, sensitivity 100, specificity 100	High
Kumar et al[94]; data from USA, China and Italy, target population unclear; covid-19 diagnosis; covid-19 diagnosis	Not applicable	Unknown	Apparent performance only	Not applicable	C index 0.998, sensitivity 100, specificity 100	High
Moutounet-Cartan[103]; data from repositories, target population unclear; covid-19 pneumonia	Not applicable	325 (125)	Training test split	98 (unknown)	Sensitivity 88	High
Ozturk et al[104]; data from repositories, target population unclear; covid-19 pneumonia	Not applicable	1127 (127)	Fivefold cross validation	Not applicable	Sensitivity 85, specificity 92 PPV 90	High
Rahimzadeh et al[105]; data from repositories, target population unclear; covid-19 pneumonia	Not applicable	633 (149)	Fivefold cross validation	Not applicable	Sensitivity 81, specificity 100, PPV 35	High
Rehman et al[106]; data from unknown origin, target population unclear; covid-19 diagnosis	Not applicable	320 (160)	Training test split	80 (40)	Sensitivity 100, specificity 98, PPV 96¶¶	High
Rehman et al[106]; data from unknown origin, target population unclear; covid-19 diagnosis	Not applicable	320 (160)	Training test split	80 (40)	Sensitivity 100, specificity 98, PPV 96¶¶	High
Rehman et al[106]; data from unknown origin, target population unclear; covid-19 diagnosis	Not applicable	320 (160)	Training test split	80 (40)	Sensitivity 100, specificity 98, PPV 96¶¶	High
Rehman et al[106]; data from unknown origin, target population unclear; covid-19 diagnosis	Not applicable	480 (160)	Training test split	120 (40)	Sensitivity 98, specificity 99. PPV 96	High

(Continued)

Table 2 | Continued

Study; setting; and outcome	Predictors in final model	Sample size: total No of participants for model development set (No with outcome)	Predictive performance on validation: Type of validation*	Predictive performance on validation: Sample size: total No of participants for model validation (No with outcome)	Predictive performance on validation: Performance* (C index, sensitivity (%), specificity (%), PPV/NPV (%), calibration slope, other (95% CI, if reported))	Overall risk of bias using PROBAST
Rehman et al[106]; data from unknown origin, target population unclear; covid-19 diagnosis	Not applicable	640 (160)	Training test split	160 (40)	Sensitivity 82, specificity 93, PPV 96	High
Singh et al[107]; data from unknown origin, target population unclear; covid-19 diagnosis	Not applicable	Unknown	Twentyfold cross validation	Not applicable	Sensitivity 91, specificity 89	High
Ucar et al[111]; data from unknown origin, target population unclear; covid-19 diagnosis	Not applicable	Unknown	Training test split	Unknown	Sensitivity 100, specificity 100, PPV 99	High
Wu et al[115]; data from unknown origin, target population unclear; covid-19 diagnosis	Not applicable	300 (150)	Training test split	400 (200)	Sensitivity 95 (91 to 98), specificity 93 (89 to 97)	High

CHD=coronary heart disease; COPD=chronic obstructive pulmonary disease; covid-19=coronavirus disease 2019; CRP=C reactive protein; CT=computed tomography; GGO=ground glass opacities; NPV=negative predictive value; PPV=positive predictive value; PROBAST=prediction model risk of bias assessment tool; PCR=polymerase chain reaction; WBC=white blood cells.
*Performance is given for the strongest form of validation reported. This is indicated in the column "type of validation." When a training test split was used, performance on the test set is reported. Apparent performance is the performance observed in the development data.
‡Calibration plot presented, but unclear which data were used.
§The development set contains scans from Chinese patients, the testing set contains scans from Chinese patients, the testing set contains scans from Chinese cases and controls, and US controls.
¶Data contain mixed cases and controls. Chinese data and controls from US and Switzerland.
**Describes similarity between segmentation of the CT scan by a medical doctor and automated segmentation.
§§Pearson correlation between the predicted and ground truth scores for patients with lung abnormalities.
¶¶Performance similar for models with different non-cases (healthy, bacterial pneumonia, and viral pneumonia).

studies had small sample sizes (table 1, table 2, table 3), which led to an increased risk of overfitting, particularly if complex modelling strategies were used. Three studies did not report the predictive performance of the developed model, and four studies reported only the apparent performance (the performance with exactly the same data used to develop the model, without adjustment for optimism owing to potential overfitting). Only 13 studies assessed calibration,[7 12 22 43 50 67 69 78 83 108 116 117 119] but the method to check calibration was probably suboptimal in two studies.[12 119]

Twenty five models were developed and externally validated in the same study (in an independent dataset, excluding random training test splits and temporal splits).[7 12 26 42 43 51 52 59 67 77 81 83 84 91 95 100 102 110 112 113 116 119] However, in 11 of these models, the datasets used for the external validation were likely not representative of the target population,[7 12 26 42 59 91 100 102 116] and in one study, data from before the covid-19 crisis were used.[113] Consequently, predictive performance could differ if the models are applied in the targeted population. In one study, commonly used performance statistics for prognosis (discrimination, calibration) were not reported.[42] Gozes,[52] Fu,[51] Chassagnon,[77] Hu,[84] Kurstjens,[95] and Vaid[112] had satisfactory predictive performance on an external validation set, but it is unclear how the data for the external validation were collected (eg, whether the patients were consecutive), and whether they are representative. Wang,[43] Barda,[67] Guo,[83] Tordjman,[110] and Gong[119] obtained satisfactory discrimination on probably unbiased validation datasets, but each of these had fewer than the recommended number of events for external validation (100).[137 138] Diaz-Quijano externally validated a diagnostic model in a large registry with reasonable discrimination, but many patients had to be excluded because no polymerase chain reaction (PCR) testing was performed.[81]

One study presented a small external validation (27 participants) that reported satisfactory predictive performance of a model originally developed for avian influenza H7N9 pneumonia. However, patients who had not recovered at the end of the study period were excluded, which again led to a selection bias.[23] Another study was a small scale external validation study (78 participants) of an existing severity score for lung computed tomography images with satisfactory reported discrimination.[54] Three studies validated existing early warning or severity scores to predict in-hospital mortality or deterioration.[85 96 108] They had satisfactory discrimination but less than the recommended number of events for validation[137 138] or unclear sample sizes, excluded patients who remained in hospital at the end of the study period, or had an unclear study design.

Discussion

In this systematic review of prediction models related to the covid-19 pandemic, we identified and critically appraised 107 studies that described 145 models.

Table 3 | Overview of prediction models for prognosis of covid-19

Study; setting; and outcome	Predictors in final model	Sample size: total No of participants for model development set (No with outcome)	Predictive performance on validation: Type of validation*	Predictive performance on validation: Sample size: total No of participants for model validation (No with outcome)	Predictive performance on validation: Performance* (C index, sensitivity (%), specificity (%), PPV/NPV (%), calibration slope, other (95% CI, if reported))	Overall risk of bias using PROBAST
Prognosis						
Original review						
Bai et al[9]; data from China, inpatients at admission with mild confirmed covid-19; deterioration into severe/critical disease (period unspecified)	Combination of demographics, signs and symptoms, laboratory results and features derived from CT images	133 (54)	Unclear	Not applicable	C index 0.95 (0.94 to 0.97)	High
Caramelo et al[19]; data from China, target population unclear; mortality (period unspecified)††	Age, sex, presence of any comorbidity (hypertension, diabetes, cardiovascular disease, chronic respiratory disease, cancer)††	Unknown	Not reported	Not applicable	Not reported	High
Lu et al[20]; data from China, inpatients at admission with suspected or confirmed covid-19; mortality (within 12 days)	Age, CRP	577 (44)	Not reported	Not applicable	Not reported	High
Qi et al[21]; data from China, inpatients with confirmed covid-19 at admission; hospital stay >10 days	6 features derived from CT images‡‡ (logistic regression model)	26 (20)	Fivefold cross validation	Not applicable	C index 0.92	High
Qi et al[21]; data from China, inpatients with confirmed covid-19 at admission; hospital stay >10 days	6 features derived from CT images‡‡ (random forest)	26 (20)	5 fold cross validation	Not applicable	C index 0.96	High
Shi et al[37]; data from China, inpatients with confirmed covid-19 at admission; death or severe covid-19 (period unspecified)	Age (dichotomised), sex, hypertension	478 (49)	Validation in less severe cases	66 (15)	Not reported	High
Xie et al[7]; data from China, inpatients with confirmed covid-19 at admission; mortality (in hospital)	Age, LDH, lymphocyte count, SP_O2	299 (155)	External validation (other Chinese centre)	130 (69)	C index 0.98 (0.96 to 1.00), calibration slope 2.5 (1.7 to 3.7)	High
Yan et al[22]; data from China, inpatients suspected of covid-19; mortality (period unspecified)	LDH, lymphocyte count, high sensitivity CRP	375 (174)	Temporal validation, selecting only severe cases	29 (17)	Sensitivity 92, PPV 95	High
Yuan et al[23]; data from China, inpatients with confirmed covid-19 at admission; mortality (period unspecified)	Clinical scorings of CT images (zone, left/right, location, attenuation, distribution of affected parenchyma)	Not applicable	External validation of existing model	27 (10)	C index 0.90 (0.87 to 0.93)	High
Update 1						
Huang et al[60]; data from China, inpatients with confirmed covid-19 at admission; severe symptoms three days after admission	Underlying diseases, fast respiratory rate >24/min, elevated CRP level (>10 mg/dL), elevated LDH level (>250 U/L)	125 (32)	Apparent performance only	Not applicable	C index 0.99 (0.97 to 1.00), sensitivity 91, specificity 96	High
Pourhomayoun et al[61]; data from 76 countries, inpatients with confirmed covid-19; in-hospital mortality (period unspecified)	Unknown	Unknown	10-fold cross validation	Not applicable	C index 0.96, sensitivity 90, specificity 97	High
Sarkar et al[44]; data from several continents (Australia, Asia, Europe, North America), inpatients with covid-19 symptoms; death *v* recovery (period unspecified)	Age, days from symptom onset to hospitalisation, from Wuhan, sex, visit to Wuhan	80 (37)	Apparent performance only	Not applicable	C index 0.97	High
Wang et al[42]; data from China, inpatients with confirmed covid-19; length of hospital stay	Age and CT features	301 (not applicable)	Not reported	Not applicable	Not reported	High
Zeng et al[62]; data from China, inpatients with confirmed covid-19; severe disease progression (period unspecified)	CT features	338 (76)	Cross validation (number of folds unclear)	Not applicable	C index 0.88	High
Zeng et al[62]; data from China, inpatients with confirmed covid-19; severe disease progression (period unspecified)	CT features and laboratory markers	338 (76)	Cross validation (number of folds unclear)	Not applicable	C index 0.88	High
Update 2						
Al-Najjar et al[63]; data from South Korea, target population unclear; recovery from covid-19 (period unspecified)	Birth year (age), sex, country, group, infection reason, confirmed date	466 (40)	Training test split	193 (14)	Sensitivity 43, specificity 98	High

(*Continued*)

Table 3 | Continued

Study; setting; and outcome	Predictors in final model	Sample size: total No of participants for model development set (No with outcome)	Predictive performance on validation: Type of validation*	Predictive performance on validation: Sample size: total No of participants for model validation (No with outcome)	Predictive performance on validation: Performance* (C index, sensitivity (%), specificity (%), PPV/NPV (%), calibration slope, other (95% CI, if reported))	Overall risk of bias using PROBAST
Al-Najjar et al[63]; data from South Korea, target population unclear; mortality (period unspecified)	Age, sex, country, region, infection reason, confirmed date	463 (25)	Training test split	191 (7)	Sensitivity 86, specificity 100	High
Barda et al[67]; data from Israel, patients with confiremed covid-19; mortality (period unspecified)	Age, sex, pack years, COPD, number of wheezing/dyspnea diagnoses, albumin, red cell distribution width, C reactive peptide, urea, lymphocyte, chloride, creatinine, high density lipoprotein, duration of hospital admissions, count of hospital admissions, count of ambulance rides, count of sulfonamide dispenses, count of anticholinergic dispenses, count of glucocorticoid dispenses, chronic respiratory disease, cardiovascular disease, diabetes, malignancy, hypertension	735,000 (8251)	Other (specify in column CL) January 29 to April 8 inclusion, follow-up until April 22 2020 (covid cases for external validation)	3176 (87)	C index 0.94 (0.92 to 0.96) , sensitivity 90 (83 to 96), PPV 17 (14 to 21)	High
Bello-Chavolla et al[70]; data from Mexico, confirmed covid-19 patients presenting at GP; 30-day mortality	Age, pregnancy, diabetes, obesity, pneumonia, CKD, COPD, immunosuppression	12424 (1137)	Training test split	3105 (297)	C index 0.80, Somer's D 0.60	High
Carr et al[75]; data from United Kingdom, inpatients with confirmed covid-19; progression to severe covid-19 (period unspecified)	Age, National Early Warning Score (NEWS) 2, CRP, neutrophil, eGFR, albumin	452 (159)	Temporal validation	256 (59)	C index 0.73, sensitivity 46, specificity 87	High
Chassagnon et al[77]; data from France, inpatients with confirmed covid-19; composite, 4-day intubation or mortality	Unclear	383 (84)	External validation (new centres, France)	95 (26)	Sensitivity 88, specificity 74	High
Colombi et al[79]; data from Italy, inpatients with confirmed covid-19; ICU admission or in-hospital (period unspecified)	Age, cardiovascular comorbidities, median platelet count, CRP, visual assessment of well aerated lung %	236 (108)	Apparent performance only	Not applicable	C index 0.86 (0.81 to 0.90), sensitivity 72 (63 to 80), specificity 81 (73 to 88) PPV 70 (61 to 78), NPV 78 (72 to 83)	High
Colombi et al[79]; data from Italy, inpatients with confirmed covid-19; ICU admission or in-hospital mortality (period unspecified)	Age, cardiovascular comorbidities, median platelet count, LDH, CRP, software assessment of well aerated lung absolute volume, adipose tissue	236 (108)	Apparent performance only	Not applicable	C index 0.86 (0.81 to 0.90), sensitivity 75 (66 to 83), specificity 81 (73 to 88), PPV 70 (61 to 78), NPV 78 (72 to 83)	High
Das et al[80]; data from South Korea, inpatients with confirmed covid-19; ICU admission or in-hospital mortality (period unspecified)	Age, sex, province, date of diagnosis, place of exposure to covid-19	3022 (61)	Training test split	604 (12)	C index 0.97	High
Gong et al[119]; data from China, target population unclear; 15-day progression to severe covid-19	Age, direct bilirubin, red cell distribution width, blood urea nitrogen, CRP, lactate dehydrogenase, albumin	189 (28)	External validation (new centres, China)	165 (40)	C index 0.85 (0.79 to 0.92), sensitivity 78, specificity 78	High
Guo et al[83]; data from China, inpatients with confirmed covid-19; 14-day progression to severe covid-19	Age, chronic illness, neutrophil to lymphocyte ratio, CRP, D-dimer	818 (24)	External validation (new centres, China)	320 (38)	C index 0.78 (0.70 to 0.87)	High
Hu et al[84]; data from China inpatients with confirmed covid-19;in-hospital mortality (period unspecified)	Age, high-sensitivity CRP, lymphocyte count, D-dimer	183 (68)	External validation (new centres, China)	64 (31)	C index 0.88, sensitivity 84, specificity 79	High
Hu et al[85]; data from China, inpatients with confirmed covid-19; in-hospital mortality (period unspecified)	Modified Early Warning Score (MEWS): heart rate, systolic blood pressure, respiratory rate, body temperature, consciousness	Not applicable	External validation only	105 (19)	C index 0.68 (0.58 to 0.77), sensitivity 68, specificity 65, PPV 30, NPV 90	High
Hu et al[85]; data from China, inpatients with confirmed covid-19; in-hospital mortality (period unspecified)	Rapid Emergency Medicine Score (REMS): mean arterial pressure, pulse rate, respiratory rate, oxygen saturation, GCS, age	Not applicable	External validation only	105 (19)	C index 0.84 (0.76 to 0.91), sensitivity 89, specificity 70, PPV 40, NPV 97	High
Ji et al[88]; data from China, inpatients with confirmed covid-19; 10-day progression to severe COVID-19	Comorbidity, age, lymphocyte count, lactate dehydrogenase	208 (40)	Internal validation by resampling (bootstrap)	Not applicable	C index 0.91 (0.86 to 0.94), sensitivity 95 (83 to 99), specificity 78 (71 to 84)	High

(Continued)

Table 3 | Continued

Study; setting; and outcome	Predictors in final model	Sample size: total No of participants for model development set (No with outcome)	Predictive performance on validation: Type of validation*	Sample size: total No of participants for model validation (No with outcome)	Performance* (C index, sensitivity (%), specificity (%), PPV/NPV (%), calibration slope, other (95% CI, if reported))	Overall risk of bias using PROBAST
Jiang et al[89]; data from China, inpatients with confirmed covid-19; acute respiratory distress syndrome***	Alanine aminotransferase, myalgias, haemoglobin, sex, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	53 (5)	10-fold cross validation	Not applicable	Classification accuracy 50%	High
Jiang et al[89]; data from China, inpatients with confirmed covid-19; acute respiratory distress syndrome***	Alanine aminotransferase, myalgias, haemoglobin, sex, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	53 (5)	10-fold cross validation	Not applicable	Classification accuracy 80%	High
Jiang et al[89]; data from China, inpatients with confirmed covid-19; acute respiratory distress syndrome***	Alanine aminotransferase, myalgias, haemoglobin, sex, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	53 (5)	10-fold cross validation	Not applicable	Classification accuracy 70%	High
Jiang et al[89]; data from China, inpatients with confirmed covid-19; acute respiratory distress syndrome***	Alanine aminotransferase, myalgias, haemoglobin, sex, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	53 (5)	10-fold cross validation	Not applicable	Classification accuracy 70%	High
Jiang et al[89]; data from China, inpatients with confirmed covid-19; acute respiratory distress syndrome***	Alanine aminotransferase, myalgias, haemoglobin, sex, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	53 (5)	10-fold cross validation	Not applicable	Classification accuracy 70%	High
Jiang et al[89]; data from China, inpatients with confirmed covid-19; acute respiratory distress syndrome***	Alanine aminotransferase, myalgias, haemoglobin, sex, temp, Na+, K+, lymphocyte count, creatinine, age, white blood count	53 (5)	10-fold cross validation	Not applicable	Classification accuracy 80%	High
Levy et al[96]; data from USA, inpatients with confirmed covid-19; in-hospital mortality (period unspecified)	Age, serum blood urea nitrogen, emergency severity index, red cell distribution width, absolute neutrophil count, serum bicarbonate, glucose	Unknown	Leave-one-out cross validation	Not applicable	C index 0.83	High
Levy et al[96]; data from USA, inpatients with confirmed covid-19; in-hospital mortality (period unspecified)	SOFA score	Not applicable	External validation only	Unclear	C index 0.73	High
Levy et al[96]; data from USA, inpatients with confirmed covid-19; in-hospital mortality (period unspecified)	CURB-65 score	Not applicable	External validation only	Unclear	C index 0.74	High
Levy et al[96]; data from USA, inpatients with confirmed covid-19; in-hospital mortality (period unspecified)	SOFA+ score	Not applicable	External validation only	Unclear	C index 0.83	High
Liu et al[98]; data from China, inpatients with confirmed covid-19; in-hospital mortality (period unspecified)	Age, underlying disease status, helper T cells, helper T cells and suppressor T cells ratio	340 (30)	Apparent performance only	Not applicable	McFadden pseudo R-squared 0.35	High
McRae et al[100]; data from China, inpatients with confirmed covid-19; in-hospital mortality (period unspecified)	Age, sex, cardiac troponin I, CRP, procalcitonin, myoglobin	160 (43)	New centres in China, case series	12 (unknown)	C index 0.94 (0.89 to 0.99)	High
Singh et al[108]; data from USA, inpatients with confirmed covid-19; ICU-level care, mechanical ventilation or in-hospital mortality (period unspecified)	Epic Deterioration Index	Unknown	External validation only	174 (61)	C index 0.76 (0.68 to 0.84), sensitivity 39 PPV 80	High

(*Continued*)

Table 3 | Continued

Study; setting; and outcome	Predictors in final model	Sample size: total No of participants for model development set (No with outcome)	Predictive performance on validation: Type of validation*	Sample size: total No of participants for model validation (No with outcome)	Performance* (C index, sensitivity (%), specificity (%), PPV/NPV (%), calibration slope, other (95% CI, if reported))	Overall risk of bias using PROBAST
Vaid et al[112]; data from USA, inpatients with confirmed covid-19; intubation, discharge to hospice care or mortalit (period unspecified)	Sex, race, ethnicity, age, hypertension, atrial fibrillation, coronary artery disease, heart failure, stroke, chronic kidney disease, diabetes, asthma, COPD, cancer, heart rate, pulse, oximetry, respiration rate, temperature, systolic blood pressure, diastolic blood pressure, body weight, sodium, potassium, creatinine, lactate, white blood cells, lymphocyte percentage, haemoglobin, red blood cell distribution width, platelets, alanine, aminotransferase, aspartate, aminotransferase, albumin, total bilirubin, prothrombin time, partial thromboplastin time, PCO_2, pH, CRP, ferritin, D-dimer, creatinine phosphokinase, lactate dehydrogenase, procalcitonin, troponin I	1225 (37)	External validation, new centres (USA)	1830 (unknown)	C index 0.84, sensitivity 86, specificity 82	High
Vazquez Guillamet et al[113]; data from USA, target population unclear; in-hospital mortality (period unspecified)	Age, immunosuppression, COPD, congestive heart failure, BMI, sex, time to mechanical ventilation (days), length of hospital stay prior to hospital admission, PaO_2/FiO_2, Glasgow coma scale, maximum heart rate, maximum respiratory rate, minimum mean arterial blood pressure, maximum temperature, minimum albumin, minimum pH	2122 (429)	External validation, new centres (USA)	1175 (154)	C index 0.81, PPV 55, NPV 89	High
Vazquez Guillamet et al[113]; data from USA, target population unclear; mechanical ventilation ›96 hours	Age, immunosuppression, COPD, congestive heart failure, BMI, sex, time to mechanical ventilation (days), length of hospital stay before hospital admission, PaO_2/FiO_2, Glasgow coma scale, maximum heart rate, maximum respiratory rate, minimum mean arterial blood pressure, maximum temperature, minimum albumin, minimum pH	2167 (158)	Training test split	1063 (96)	C index 0.81	High
Vazquez Guillamet et al[113]; data from USA, target population unclear; mechanical ventilation ›96 hours	Age, immunosuppression, COPD, congestive heart failure, BMI, sex, time to mechanical ventilation (days), length of hospital stay prior to hospital admission, PaO_2/FiO_2, Glasgow coma scale, maximum heart rate, maximum respiratory rate, minimum mean arterial blood pressure, maximum temperature, minimum albumin, minimum pH	1169 (141)	Training test split	619 (90)	C index 0.78	High
Zhang et al[116]; data from China and United Kingdom, inpatients with confirmed covid-19; in hospital mortality (period unspecified)	Age, sex, neutrophil count, lymphocyte count, platelet count, CRP, creatinine	653 (20)	External validation (new centres, different country)	226 (77)	C index 0.75, sensitivity 23, specificity 95, PPV 69, NPV 71	High
Zhang et al[116]; data from China, inpatients with confirmed covid-19; ARDS, intubation or ECMO, ICU admission, in hospital mortality (period unspecified)	Age, sex, chronic lung disease, diabetes mellitus, malignancy, cough, dyspnoea, immunocompromised, hypertension, heart disease, chronic renal disease, fever, fatigue, diarrhoea	768 (72)	Repeated five-fold cross validation	Not applicable	C index 0.80, sensitivity 9, specificity 99 PPV 53, NPV 91	High
Zhang et al[116]; data from China and United Kingdom, inpatients with confirmed covid-19; ARDS, intubation or ECMO, ICU admission, in hospital mortality (period unspecified)	Age, sex, neutrophil count, lymphocyte count, platelet count, CRP, creatinine	653 (58)	External validation (new centres, different country)	226 (97)	C index 0.72, sensitivity 40, specificity 85, PPV 67, NPV 65	High

ARDS=acute respiratory distress syndrome; BMI=body mass index; COPD=chronic obstructive pulmonary disease; covid-19=coronavirus disease 2019; CRP=C reactive protein; CT=computed tomography; ECMO=extracorporal membrane oxygenation; ICU=intensive care unit; LDH=lactate dehydrogenase; NPV=negative predictive value; PaO_2/FiO_2=the ratio of arterial oxygen partial pressure to fractional inspired oxygen; PCO_2=partial pressure of carbon dioxide; PPV=positive predictive value; PROBAST=prediction model risk of bias assessment tool; SOFA=sequential organ failure assessment score; SPO_2=oxygen saturation; Na+=sodium; K+=potassium.

*Performance is given for the strongest form of validation reported. This is indicated in the column "type of validation." When a training test split was used, performance on the test set is reported. Apparent performance is the performance observed in the development data.

††Outcome and predictor data were simulated.

‡‡Wavelet-HLH_gldm_SmallDependenceLowGrayLevelEmphasis, wavelet-LHH_glcm_Correlation, wavelet-LHL_glszm_GrayLevelVariance, wavelet-LLH_glszm_SizeZoneNonUniformityNormalized, wavelet-LLH_glszm_SmallAreaEmphasis, wavelet-LLH_glcm_Correlation.

***Each model uses a different predictive algorithm.

Table 4 | Risk of bias assessment (using PROBAST) based on four domains across 107 studies that created prediction models for coronavirus disease 2019

	Risk of bias			
Authors	Participants	Predictors	Outcome	Analysis
Hospital admission in general population				
Original review				
DeCaprio et al[8]	High	Low	High	High
Update 2				
Jiang et al[90]	High	Unclear	High	High
Diagnosis				
Original review				
Feng et al[10]	Low	Unclear	High	High
Lopez-Rincon et al[35]	Unclear	Low	Low	High
Meng et al[12]	High	Low	High	High
Song et al[31]	High	Unclear	Low	High
Update 1				
Martin et al[41]	High	High	High	High
Sun et al[40]	Low	Low	Unclear	High
Wang et al[43]	Low	Unclear	Unclear	High
Wu et al[45]	High	Unclear	Low	High
Update 2				
Batista et al[69]	Unclear	Unclear	Low	High
Brinati et al[74]	Unclear	Unclear	Low	High
Chen et al[78]	High	High	Low	High
Diaz-Quijano et al[81]	High	High	Low	High
Kurstjens et al[95]	Unclear	Low	High	High
Mei et al[101]	High	Unclear	Unclear	High
Menni et al[102]	High	Unclear	Unclear	High
Soares et al[109]	Unclear	Unclear	Low	High
Tordjman et al[110]	Low	Unclear	Unclear	High
Zhao et al[117]	High	High	Unclear	High
Diagnosis of severity				
Original review				
Yu et al[25]	Unclear	Unclear	Unclear	High
Update 1				
Zhou et al[46]	Unclear	Low	High	High
Update 2				
Benchoufi et al[71]	High	Low	Low	High
Chassagnon et al[77]	Low	Low	Low	High
Li et al[97]	Unclear	Unclear	Unclear	High
Lyu et al[99]	Low	Unclear	Unclear	High
Wang et al[114]	Unclear	High	Low	High
Zhu et al[118]	Low	Low	High	High
Diagnostic imaging				
Original review				
Barstugan et al[32]	Unclear	Unclear	Unclear	High
Chen et al[27]	High	Unclear	Low	High*
Gozes et al[26]	Unclear	Unclear	High	High
Jin et al[11]	High	Unclear	Unclear	High†
Jin et al[33]	High	Unclear	High	High*
Li et al[34]	Low	Unclear	Low	High
Shan et al[29]	Unclear	Unclear	High	High†
Shi et al[36]	High	Unclear	Low	High
Wang et al[30]	High	Unclear	Low	High
Xu et al[28]	High	Unclear	High	High
Song et al[24]	Unclear	Unclear	Low	High
Zheng et al[38]	Unclear	Unclear	High	High
Update 1				
Abbas et al[47]	High	Unclear	Unclear	High
Apostolopoulos et al[48]	High	Unclear	High	High
Bukhari et al[49]	Unclear	Unclear	Unclear	High
Chaganti et al[50]	High	Unclear	Low	Unclear
Chowdhury et al[39]	High	Unclear	Unclear	High
Fu et al[51]	High	Unclear	Unclear	High
Gozes et al[52]	High	Unclear	Unclear	High
Imran et al[53]	High	Unclear	Unclear	High*
Li et al[54]	Low	Low	Unclear	High
Li et al[55]	High	Unclear	High	High*
Hassanien et al[56]	Unclear	Unclear	Unclear	High*
Tang et al[57]	Unclear	Unclear	High	High

(*Continued*)

These prediction models can be divided into three categories: models for the general population to predict the risk of having covid-19 or being admitted to hospital for covid-19; models to support the diagnosis of covid-19 in patients with suspected infection; and models to support the prognostication of patients with covid-19. All models reported moderate to excellent predictive performance, but all were appraised to have high risk of bias owing to a combination of poor reporting and poor methodological conduct for participant selection, predictor description, and statistical methods used. Models were developed on data from different countries, but the majority used data from China or public international data repositories. With few exceptions, the available sample sizes and number of events for the outcomes of interest were limited. This is a well known problem when building prediction models and increases the risk of overfitting the model.[139] A high risk of bias implies that the performance of these models in new samples will probably be worse than that reported by the researchers. Therefore, the estimated C indices, often close to 1 and indicating near perfect discrimination, are probably optimistic. The majority of studies developed new models, only 27 carried out an external validation, and calibration was rarely assessed.

We reviewed 57 studies that used advanced machine learning methodology on medical images to diagnose covid-19, covid-19 related pneumonia, or to assist in segmentation of lung images. The predictive performance measures showed a high to almost perfect ability to identify covid-19, although these models and their evaluations also had a high risk of bias, notably because of poor reporting and an artificial mix of patients with and without covid-19. Therefore, we do not recommend any of the 145 identified prediction models to be used in practice.

Challenges and opportunities

The main aim of prediction models is to support medical decision making. Therefore, it is vital to identify a target population in which predictions serve a clinical need, and a representative dataset (preferably comprising consecutive patients) on which the prediction model can be developed and validated. This target population must also be carefully described so that the performance of the developed or validated model can be appraised in context, and users know which people the model applies to when making predictions. Unfortunately, the studies included in our systematic review often lacked an adequate description of the study population, which leaves users of these models in doubt about the models' applicability. Although we recognise that all studies were done under severe time constraints, we recommend that any studies currently in preprint and all future studies should adhere to the TRIPOD reporting guideline[16] to improve the description of their study population and their modelling choices. TRIPOD translations (eg, in Chinese and Japanese) are also available at https://www.tripod-statement.org.

Table 4 | Continued

Authors	Risk of bias: Participants	Predictors	Outcome	Analysis
Wang et al[42]	Low	Unclear	Unclear	High
Zhang et al[58]	High	Unclear	High	High
Zhou et al[59]	High	Unclear	High	High*
Update 2				
Angelov et al[64]	High	Unclear	High	High
Arpan et al[65]	Unclear	Unclear	Unclear	High
Bai et al[66]	High	Unclear	High	High
Bassi et al[68]	High	Unclear	High	High
Borghesi et al[72]	High	Unclear	Unclear	High
Born et al[73]	High	Unclear	Unclear	High
Castiglioni et al[76]	Unclear	Unclear	Low	High
Guiot et al[82]	High	Unclear	Low	High
Hu et al[86]	High	Unclear	High	High
Islam et al[87]	High	Unclear	High	High
Kana et al[91]	High	Unclear	High	High*
Karim et al[92]	High	Unclear	High	High
Khan et al[93]	High	Unclear	High	High*
Kumar et al[94]	High	Unclear	Unclear	High*
Moutounet-Cartan[103]	Unclear	Unclear	Unclear	High
Ozturk et al[104]	High	Unclear	Unclear	High
Rahimzadeh et al[105]	High	Unclear	Unclear	High
Rehman et al[106]	High	Unclear	Unclear	High
Singh et al[107]	High	Unclear	Unclear	High
Ucar et al[107]	High	Unclear	Unclear	High
Wu et al[115]	High	Unclear	Unclear	High
Prognosis				
Original review				
Bai et al[9]	Low	Unclear	Unclear	High
Caramelo et al[19]	High	High	High	High
Lu et al[20]	Low	Low	Low	High
Qi et al[21]	Unclear	Low	Low	High
Shi et al[37]	High	High	High	High
Xie et al[7]	Low	Low	Low	High
Yan et al[22]	Low	High	Low	High
Yuan et al[23]	Low	High	Low	High
Update 1				
Huang et al[60]	Unclear	Unclear	Unclear	High
Pourhomayoun et al[61]	Low	Low	Unclear	High
Sarkar et al[44]	High	High	High	High
Wang et al[42]	Low	Low	Low	High
Zeng et al[62]	Low	Low	Low	High
Update 2				
Al-Najjar et al[63]	Unclear	Unclear	Unclear	High
Barda et al[67]	Low	Low	High	High
Bello-Chavolla et al[70]	Unclear	Unclear	Low	High
Carr et al[75]	Low	Low	Low	High
Chassagnon et al[77]	Low	Low	Low	High
Colombi et al[79]	High	Unclear	Unclear	High
Das et al[80]	Low	Low	Low	High
Gong et al[119]	Low	Low	high	High
Guo et al[83]	Low	High	Unclear	High
Hu et al[84]	High	Low	Low	High
Hu et al[85]	Low	Unclear	Low	High
Ji et al[88]	Low	Low	Low	High
Jiang et al[89]	Unclear	Unclear	Unclear	High
Levy et al[96]	Low	Low	Low	High
Liu et al[98]	Low	Low	Low	High
McRae et al[100]	High	High	High	High
Singh et al[108]	low	Unclear	High	High
Vaid et al[112]	Unclear	High	High	High
Vazquez Guillamet et al[113]	High	Low	Unclear	High
Zhang et al[116]	Low	Unclear	Unclear/low‡	High

PROBAST=prediction model risk of bias assessment tool.
*Risk of bias high owing to calibration not being evaluated. If this criterion is not taken into account, analysis risk of bias would have been unclear.
†Risk of bias high owing to calibration not being evaluated. If this criterion is not taken into account, analysis risk of bias would have been low.
‡Zhang et al evaluated two outcomes: death (low risk of bias) and a composite poor outcome (unclear risk of bias).

A better description of the study population could also help us understand the observed variability in the reported outcomes across studies, such as covid-19 related mortality and covid-19 prevalence. The variability in prevalence could in part be reflective of different diagnostic standards across studies. Note that the majority of diagnostic models use viral nucleic acid test results as the gold standard, which may have unacceptable false negative rates.

Covid-19 prediction problems will often not present as a simple binary classification task. Complexities in the data should be handled appropriately. For example, a prediction horizon should be specified for prognostic outcomes (eg, 30 day mortality). If study participants have neither recovered nor died within that time period, their data should not be excluded from analysis, which most reviewed studies have done. Instead, an appropriate time to event analysis should be considered to allow for administrative censoring.[17] Censoring for other reasons, for instance because of quick recovery and loss to follow-up of patients who are no longer at risk of death from covid-19, could necessitate analysis in a competing risk framework.[140]

A prediction model applied in a new healthcare setting or country often produces predictions that are miscalibrated[141] and might need to be updated before it can safely be applied in that new setting.[17] This requires data from patients with covid-19 to be available from that system. Instead of developing and updating predictions in their local setting, individual participant data from multiple countries and healthcare systems might allow better understanding of the generalisability and implementation of prediction models across different settings and populations. This approach could greatly improve the applicability and robustness of prediction models in routine care.[142-146]

The evidence base for the development and validation of prediction models related to covid-19 will quickly increase over the coming months. Together with the increasing evidence from predictor finding studies[147-153] and open peer review initiatives for covid-19 related publications,[154] data registries[120 121 155-157] are being set up. To maximise the new opportunities and to facilitate individual participant data meta-analyses, the World Health Organization has released a new data platform to encourage sharing of anonymised covid-19 clinical data.[158] To leverage the full potential of these evolutions, international and interdisciplinary collaboration in terms of data acquisition, model building and validation is crucial.

Study limitations

With new publications on covid-19 related prediction models rapidly entering the medical literature, this systematic review cannot be viewed as an up-to-date list of all currently available covid-19 related prediction models. Also, 87 of the studies we reviewed were only available as preprints. These studies might improve after peer review, when they enter the official medical literature; we will reassess these peer reviewed publications in future updates. We also found other prediction models that are

Box 1: Availability of models in format for use in clinical practice

Several studies presented their models in a format for use in clinical practice. However, because all models were at high risk of bias, we do not recommend their routine use before they are properly externally validated.

Models to predict risk of developing coronavirus disease 2019 (covid-19) or of hospital admission for covid-19 in general population

The "COVID-19 Vulnerability Index" to detect hospital admission for covid-19 pneumonia from other respiratory infections (eg, pneumonia, influenza) is available as an online tool.[8122]

Diagnostic models

Several sum scores,[31 95 110 117] and model equations[81 102] are available to support the diagnosis. Graphical diagnostic aids include nomograms[43 78 117] and a decision tree.[74] The "COVID-19 diagnosis aid" app is available on iOS and android devices to diagnose covid-19 in asymptomatic patients and those with suspected disease.[12] Additionally, online tools are available.[10 45 74 95 123-125] Classification in terms of disease severity can be done using a published equation.[114] A decision tree to detect severe disease for paediatric patients with confirmed covid-19 is also available in an article.[25]

Diagnostic models based on images

Five artificial intelligence models to assist with diagnosis based on medical images are available through web applications.[24 27 30 73 91 126-130] One model is deployed in 16 hospitals, but the authors do not provide any usable tools in their study.[33] Two papers includes a severity scoring system to classify patients based on images.[5472]

Prognostic models

To assist in the prognosis of mortality, a nomogram,[7] a decision tree,[22] a score system,[70] online tools,[80 84 96 98 131-134] and a computed tomography based scoring rule are available in the articles.[23] Other online tools predict in-hospital death and the need for prolonged mechanical ventilation,[113 135] or in-hospital death and a composite of poor outcomes.[116 136] Additionally nomograms,[88 119] sumscores[83 88] and a model equation[60] are available to predict progression to severe covid-19.

Several studies made their code available on GitHub.[8 11 34 35 38 47 55 65-68 70 73 86 92 98 101 104 105 109] Seventy four studies did not include any usable equation, format, code, or reference for use or validation of their prediction model.

currently being used in clinical practice without scientific publications,[159] and web risk calculators launched for use while the scientific manuscript is still under review (and unavailable on request). These unpublished models naturally fall outside the scope of this review of the literature.[160] As we have argued extensively elsewhere,[161] transparent reporting that enables validation by independent researchers is key for predictive analytics, and clinical guidelines should only recommend publicly available and verifiable algorithms.

Implications for practice

All 145 reviewed prediction models were found to have a high risk of bias, and evidence from independent external validation of the newly developed models is currently lacking. However, the urgency of diagnostic and prognostic models to assist in quick and efficient triage of patients in the covid-19 pandemic might encourage clinicians and policymakers to prematurely implement prediction models without sufficient documentation and validation. Earlier studies have shown that models were of limited use in the context of a pandemic,[162] and they could even cause more harm than good.[163] Therefore, we cannot recommend any model for use in practice at this point.

The current oversupply of insufficiently validated models is not useful for clinical practice. Future studies should focus on validating, comparing, improving, and updating promising available prediction models, rather than developing new ones.[17] For example, Diaz-Quijano developed and externally validated a diagnostic model using Brazilian surveillance data with reasonable discrimination, but many patients had to be excluded because no PCR testing was performed, hence this model needs further validation.[17] Two other models to diagnose covid-19 also showed promising discrimination at external validation in small unselected cohorts.[43 110] An externally validated model that used computed tomography based total severity scores showed good discrimination between patients with mild, common, and severe-critical disease.[54] Two models to predict progression to severe covid-19 within two weeks showed promising discrimination when validated externally on unselected cohorts.[83 119] Another model discriminated well between survivors and non-survivors among confirmed cases, but the prediction horizon was not specified, and the study had many missing values for key parameters.[67] Because reporting in each of these studies was insufficiently detailed and the validation was in datasets with fewer than 100 events in the smallest outcome category, validation in larger, international datasets is needed. Such external validations should assess not only discrimination, but also calibration and clinical utility (net benefit).[141 146 163] Owing to differences between healthcare systems (eg, Chinese and European) in when patients are admitted to and discharged from hospital, as well as the testing criteria for patients with suspected covid-19, we anticipate most existing models will be miscalibrated, but this can usually be solved by updating and adjustment to the local setting.

When creating a new prediction model, we recommend building on previous literature and expert opinion to select predictors, rather than selecting predictors in a purely data driven way.[17] This is especially important for datasets with limited sample size.[164] Based on the predictors included in multiple models identified by our review, we encourage researchers to consider incorporating several candidate predictors. Common predictors include age, body temperature, lymphocyte count, and lung imaging features. Flu-like signs and symptoms and neutrophil count are frequently predictive in diagnostic models, while comorbidities, sex, C reactive protein, and creatinine are frequently reported prognostic factors. By pointing to the most important methodological challenges and issues in design and reporting of the currently available models, we hope to have provided a useful starting point for further studies aiming to develop new models, or to validate and update existing ones.

This living systematic review has been conducted in collaboration with the Cochrane Prognosis Methods Group. We will update this review and appraisal continuously to provide up-to-date information for healthcare decision makers and professionals as more international research emerges over time.

Box 2: Common causes of risk of bias in the reported prediction models

Models to predict coronavirus disease 2019 (covid-19) risk in general population
These models were based on proxy outcomes to predict covid-19 related risks, such as presence of or hospital admission due to severe respiratory disease, in the absence of data of patients with covid-19.[8 90]

Diagnostic models
Controls are probably not representative of the target population for a diagnostic model (eg, controls for a screening model had viral pneumonia).[12 41 45 78 102] The test used to determine the outcome varied between participants,[12 41 95] or one of the predictors (eg, fever) was part of the outcome definition.[10]

Diagnostic models based on medical imaging
Generally, studies did not clearly report which patients had imaging during clinical routine, and it was unclear whether the selection of controls was made from the target population (that is, patients with suspected covid-19). Often studies did not clearly report how regions of interest were annotated. Images were sometimes annotated by only one scorer without quality control.[26 28 47 52 55 91-93] Careful description of model specification and subsequent estimation were lacking, challenging the transparency and reproducibility of the models. Studies used different deep learning architectures, some were established and others specifically designed, without benchmarking the used architecture against others.

Prognostic models
Study participants were often excluded because they did not develop the outcome at the end of the study period but were still in follow-up (that is, they were in hospital but had not recovered or died), yielding a highly selected study sample.[7 21-23 44 96 98 100] Additionally, only six studies accounted for censoring by using Cox regression[20 42 70 83 88] or competing risk models.[62] Some studies used the last available predictor measurement from electronic health records (rather than measuring the predictor value at the time when the model was intended for use).[22 67 100]

Conclusion

Several diagnostic and prognostic models for covid-19 are currently available and they all report moderate to excellent discrimination. However, these models are all at high risk of bias, mainly because of non-representative selection of control patients, exclusion of patients who had not experienced the event of interest by the end of the study, and model overfitting. Therefore, their performance estimates are probably optimistic and misleading. The COVID-PRECISE group does not recommend any of the current prediction models to be used in practice. Future studies aimed at developing and validating diagnostic or prognostic models for covid-19 should explicitly address the concerns raised. Sharing data and expertise for the validation and updating of covid-19 related prediction models is urgently needed.

AUTHOR AFFILIATIONS

[1]Department of Epidemiology, CAPHRI Care and Public Health Research Institute, Maastricht University, Peter Debyeplein 1, 6229 HA Maastricht, Netherlands

[2]Department of Development and Regeneration, KU Leuven, Leuven, Belgium

[3]Department of Biomedical Data Sciences, Leiden University Medical Centre, Leiden, Netherlands

[4]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Musculoskeletal Sciences, University of Oxford, Oxford, UK

[5]NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK

[6]Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Keele, UK

[7]Section for Clinical Biometrics, Centre for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria

[8]Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht University, Utrecht, Netherlands

[9]Cochrane Netherlands, University Medical Centre Utrecht, Utrecht University, Utrecht, Netherlands

[10]Department of Medical Microbiology, University Medical Centre Utrecht, Utrecht, Netherlands

[11]HRB Clinical Research Facility, Cork, Ireland

[12]School of Public Health, University College Cork, Cork, Ireland

[13]Department of Electrical Engineering, ESAT Stadius, KU Leuven, Leuven, Belgium

[14]Ordensklinikum Linz, Hospital Elisabethinen, Department of Nephrology, Linz, Austria

[15]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[16]Palliative and Advanced Illness Research Center and Division of Pulmonary and Critical Care Medicine, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[17]Department of Microbiology, Immunology and Transplantation, KU Leuven-University of Leuven, Leuven, Belgium

[18]Department of General Internal Medicine, KU Leuven-University Hospitals Leuven, Leuven, Belgium

[19]Evidence-Based Oncology, Department I of Internal Medicine and Centre for Integrated Oncology Aachen Bonn Cologne Dusseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

[20]Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, Netherlands

[21]Division of Informatics, Imaging and Data Science, Faculty of Biology, Medicine and Health, Manchester Academic Heath Science Centre, University of Manchester, Manchester, UK

[22]Centre for Biostatistics, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK

[23]Centre for Epidemiology Versus Arthritis, Centre for Musculoskeletal Research, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK

[24]Division of Nursing, Midwifery and Social Work, School of Health Sciences, University of Manchester, Manchester, UK

[25]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

[26]Amsterdam UMC, University of Amsterdam, Amsterdam Public Health, Medical Library, Netherlands

[27]Department of Epidemiology and Biostatistics, Imperial College London School of Public Health, London, UK

[28]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece

[29]Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre+, Maastricht, Netherlands

[30]EPI-Centre, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

[31]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

[32]Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany

[33]Berlin Institute of Health, Berlin, Germany

[34]Kleijnen Systematic Reviews, York, UK

We thank the authors who made their work available by posting it on public registries or sharing it confidentially. A preprint version of the study is publicly available on medRxiv.

Contributors: LW conceived the study. LW and MvS designed the study. LW, MvS, and BVC screened titles and abstracts for inclusion. LW, BVC, GSC, TPAD, MCH, GH, KGMM, RDR, ES, LJMS, EWS, KIES, CW, JAAD, PD, MCH, NK, AL, KL, JM, CLAN, JBR, JCS, CS, NS, MS, RS, TT, SMJvK, FSvR, LH, RW, GPM, IT, JYV, DLD, JW, FSvR, PH, VMTdJ, and MvS extracted and analysed data. MDV helped interpret the findings on deep learning studies and MMJB, LH, and MCH assisted in the interpretation from a clinical viewpoint. RS and FSvR offered technical and administrative support. LW and MvS wrote the first draft, which all authors revised for critical content. All authors approved the final manuscript. LW and MvS are the guarantors. The guarantors had full access to all the data in the study, take responsibility for the integrity of the data and the accuracy of the data analysis, and had final responsibility for the decision to submit for publication. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: LW, BVC, LH, and MDV acknowledge specific funding for this work from Internal Funds KU Leuven, KOOR, and the COVID-19 Fund. LW is a postdoctoral fellow of Research Foundation-Flanders (FWO). BVC received support from FWO (grant G0B4716N) and Internal Funds KU Leuven (grant C24/15/037). TPAD acknowledges financial support from the Netherlands Organisation for Health Research and Development (grant 91617050). VMTdJ was supported by the European Union Horizon 2020 Research and Innovation Programme under ReCoDID grant agreement 825746. KGMM and JAAD acknowledge financial support from Cochrane Collaboration (SMF 2018). KIES is funded by the National Institute for Health Research (NIHR) School for Primary Care Research. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. GSC was supported by the NIHR Biomedical Research Centre, Oxford, and Cancer Research UK (programme grant C49297/A27294). JM was supported by the Cancer Research UK (programme grant C49297/A27294). PD was supported by the NIHR Biomedical Research Centre, Oxford. MOH is supported by the National Heart, Lung, and Blood Institute of the United States National Institutes of Health (grant R00 HL141678). The funders played no role in study design, data collection, data analysis, data interpretation, or reporting.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: support from Internal Funds KU Leuven, KOOR, and the COVID-19 Fund for the submitted work; no competing interests with regards to the submitted work; LW discloses support from Research Foundation-Flanders; RDR reports personal fees as a statistics editor for *The BMJ* (since 2009), consultancy fees for Roche for giving meta-analysis teaching and advice in October 2018, and personal fees for delivering in-house training courses at Barts and the London School of Medicine and Dentistry, and the Universities of Aberdeen, Exeter, and Leeds, all outside the submitted work; MS coauthored the editorial on the original article.

Ethical approval: Not required.

Data sharing: The study protocol is available online at https://osf.io/ehc47/. Most included studies are publicly available. Additional data are available upon reasonable request.

The lead authors affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Dissemination to participants and related patient and public communities: The study protocol is available online at https://osf.io/ehc47/.

1 Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020:S1473-3099(20)30120-1. doi:10.1016/S1473-3099(20)30120-1
2 Arabi YM, Murthy S, Webb S. COVID-19: a novel coronavirus and a novel challenge for critical care. *Intensive Care Med* 2020. doi:10.1007/s00134-020-05955-1
3 Grasselli G, Pesenti A, Cecconi M. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. *JAMA* 2020. doi:10.1001/jama.2020.4031
4 Xie J, Tong Z, Guan X, Du B, Qiu H, Slutsky AS. Critical care crisis and some recommendations during the COVID-19 epidemic in China. *Intensive Care Med* 2020. doi:10.1007/s00134-020-05979-7
5 Wellcome Trust. Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak 2020. https://wellcome.ac.uk/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-covid-19-outbreak.
6 Institute of Social and Preventive Medicine. Living evidence on COVID-19 2020. https://ispmbern.github.io/covid-19/living-review/index.html.
7 Xie J, Hungerford D, Chen H, et al. Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19. medRxiv [Preprint] 2020. doi:10.1101/2020.03.28.20045997
8 DeCaprio D, Gartner J, Burgess T, et al. Building a COVID-19 vulnerability index. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200307347D.
9 Bai X, Fang C, Zhou Y, et al. Predicting COVID-19 malignant progression with AI techniques. medRxiv [Preprint] 2020. doi:10.1101/2020.03.20.20037325
10 Feng C, Huang Z, Wang L, et al. A novel triage tool of artificial intelligence assisted diagnosis aid system for suspected covid-19 pneumonia in fever clinics. medRxiv [Preprint] 2020. doi:10.1101/2020.03.19.20039099
11 Jin C, Chen W, Cao Y, et al. Development and evaluation of an AI system for covid-19 diagnosis. medRxiv [Preprint] 2020. doi:10.1101/2020.03.20.20039834
12 Meng Z, Wang M, Song H, et al. Development and utilization of an intelligent application for aiding COVID-19 diagnosis. medRxiv [Preprint] 2020. doi:10.1101/2020.03.18.20035816
13 Thomas J, Brunton J, Graziosi S. EPPI-Reviewer 4.0: software for research synthesis [program]. EPPI-Centre Software. London: Social Science Research Unit, Institute of Education, University of London, 2010.
14 Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014;11:e1001744. doi:10.1371/journal.pmed.1001744
15 Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med* 2019;170:W1-33. doi:10.7326/M18-1377
16 Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1-73. doi:10.7326/M14-0698
17 Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating.* Springer US, 2019. doi:10.1007/978-3-030-16399-0
18 Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009;6:e1000100. doi:10.1371/journal.pmed.1000100
19 Caramelo F, Ferreira N, Oliveiros B. Estimation of risk factors for COVID-19 mortality - preliminary results. medRxiv [Preprint] 2020. doi:10.1101/2020.02.24.20027268
20 Lu J, Hu S, Fan R, et al. ACP risk grade: a simple mortality index for patients with confirmed or suspected severe acute respiratory syndrome coronavirus 2 disease (COVID-19) during the early stage of outbreak in Wuhan, China. medRxiv [Preprint] 2020. doi:10.1101/2020.02.20.20025510
21 Qi X, Jiang Z, YU Q, et al. Machine learning-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study. medRxiv [Preprint] 2020. doi:10.1101/2020.02.29.20029603
22 Yan L, Zhang H-T, Xiao Y, et al. Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. medRxiv [Preprint] 2020. doi:10.1101/2020.02.27.20028027
23 Yuan M, Yin W, Tao Z, Tan W, Hu Y. Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. *PLoS One* 2020;15:e0230548. doi:10.1371/journal.pone.0230548
24 Song Y, Zheng S, Li L, et al. Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with CT images. medRxiv [Preprint] 2020. doi:10.1101/2020.02.23.20026930
25 Yu H, Shao J, Guo Y, et al. Data-driven discovery of clinical routes for severity detection in covid-19 pediatric cases. medRxiv [Preprint] 2020. doi:10.1101/2020.03.09.20032219
26 Gozes O, Frid-Adar M, Greenspan H, et al. Rapid AI development cycle for the coronavirus (covid-19) pandemic: initial results for automated detection & patient monitoring using deep learning CT image analysis. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200305037G
27 Chen J, Wu L, Zhang J, et al. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. medRxiv [Preprint] 2020. doi:10.1101/2020.02.25.20021568
28 Xu X, Jiang X, Ma C, et al. Deep learning system to screen coronavirus disease 2019 pneumonia. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200209334X
29 Shan F, Gao Y, Wang J, et al. Lung infection quantification of covid-19 in CT images with deep learning. arXiv e-prints 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200304655S

30 Wang S, Kang B, Ma J, et al. A deep learning algorithm using CT images to screen for corona virus disease (covid-19). medRxiv [Preprint] 2020. doi:10.1101/2020.02.14.20023028
31 Song C-Y, Xu J, He J-Q, et al. COVID-19 early warning score: a multi-parameter screening tool to identify highly suspected patients. medRxiv [Preprint] 2020. doi:10.1101/2020.03.05.20031906
32 Barstugan M, Ozkaya U, Ozturk S. Coronavirus (COVID-19) classification using CT images by machine learning methods. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200309424B
33 Jin S, Wang B, Xu H, et al. AI-assisted CT imaging analysis for COVID-19 screening: building and deploying a medical AI system in four weeks. medRxiv [Preprint] 2020. doi:10.1101/2020.03.19.20039354
34 Li L, Qin L, Xu Z, et al. Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest CT. *Radiology* 2020:200905. doi:10.1148/radiol.2020200905.
35 Lopez-Rincon A, Tonda A, Mendoza-Maldonado L, et al. Accurate identification of SARS-CoV-2 from viral genome sequences using deep learning. bioRxiv [Preprint] 2020. doi:10.1101/2020.03.13.990242
36 Shi F, Xia L, Shan F, et al. Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200309860S
37 Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. *Crit Care* 2020;24:108. doi:10.1186/s13054-020-2833-7
38 Zheng C, Deng X, Fu Q, et al. Deep learning-based detection for covid-19 from chest CT using weak label. medRxiv [Preprint] 2020. doi:10.1101/2020.03.12.20027185
39 Chowdhury MEH, Rahman T, Khandakar A, et al. Can AI help in screening Viral and COVID-19 pneumonia? arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200313145C.
40 Sun Y, Koh V, Marimuthu K, et al. Epidemiological and clinical predictors of covid-19. *Clin Infect Dis* 2020;ciaa322. doi:10.1093/cid/ciaa322
41 Martin A, Nateqi J, Gruarin S, et al. An artificial intelligence-based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot. bioRxiv [Preprint] 2020. doi:10.1101/2020.03.25.008805
42 Wang S, Zha Y, Li W, et al. A fully automatic deep learning system for covid-19 diagnostic and prognostic analysis. medRxiv [Preprint] 2020. doi:10.1101/2020.03.24.20042317
43 Wang Z, Weng J, Li Z, et al. Development and validation of a diagnostic nomogram to predict covid-19 pneumonia. medRxiv [Preprint] 2020. doi:10.1101/2020.04.03.20052068
44 Sarkar J, Chakrabarti P. A machine learning model reveals older age and delayed hospitalization as predictors of mortality in patients with covid-19. medRxiv [Preprint] 2020. doi:10.1101/2020.03.25.20043331
45 Wu J, Zhang P, Zhang L, et al. Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. medRxiv [Preprint] 2020. doi:10.1101/2020.04.02.20051136
46 Zhou Y, Yang Z, Guo Y, et al. A new predictor of disease severity in patients with covid-19 in Wuhan, China. medRxiv [Preprint] 2020. doi:10.1101/2020.03.24.20042119
47 Abbas A, Abdelsamea M, Gaber M. Classification of covid-19 in chest x-ray images using DeTraC deep convolutional neural network. medRxiv [Preprint] 2020. doi:10.1101/2020.03.30.20047456
48 Apostolopoulos ID, Mpesiana TA. *Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks.* Physical and Engineering Sciences in Medicine, 2020. doi:10.1007/s13246-020-00865-4
49 Bukhari SUK, Bukhari SSK, Syed A, et al. The diagnostic evaluation of Convolutional Neural Network (CNN) for the assessment of chest X-ray of patients infected with COVID-19. medRxiv [Preprint] 2020. doi:10.1101/2020.03.26.20044610
50 Chaganti S, Balachandran A, Chabin G, et al. Quantification of tomographic patterns associated with covid-19 from chest CT. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200401279C.
51 Fu M, Yi S-L, Zeng Y, et al. Deep learning-based recognizing covid-19 and other common infectious diseases of the lung by chest CT scan images. medRxiv [Preprint] 2020. doi:10.1101/2020.03.28.20046045
52 Gozes O, Frid-Adar M, Sagie N, et al. Coronavirus detection and analysis on chest CT with deep learning. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200402640G.
53 Imran A, Posokhova I, Qureshi HN, et al. AI4COVID-19: AI enabled preliminary diagnosis for covid-19 from cough samples via an app. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200401275I.
54 Li K, Fang Y, Li W, et al. CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). *Eur Radiol* 2020; doi:10.1007/s00330-020-06817-6
55 Li X, Li C, Zhu D. COVID-MobileXpert: on-device covid-19 screening using snapshots of chest x-ray. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200403042L.
56 Hassanien AE, Mahdy LN, Ezzat KA, et al. Automatic x-ray covid-19 lung image classification system based on multi-level thresholding and support vector machine. medRxiv [Preprint] 2020. doi:10.1101/2020.03.30.20047787
57 Tang Z, Zhao W, Xie X, et al. Severity assessment of coronavirus disease 2019 (covid-19) using quantitative features from chest CT images. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200311988T.
58 Zhang J, Xie Y, Li Y, et al. COVID-19 Screening on Chest X-ray Images Using Deep Learning based Anomaly Detection. arXiv e-prints 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200312338Z.
59 Zhou M, Chen Y, Wang D, et al. Improved deep learning model for differentiating novel coronavirus pneumonia and influenza pneumonia. medRxiv [Preprint] 2020. doi:10.1101/2020.03.24.20043117
60 Huang H, Cai S, Li Y, et al. Prognostic factors for COVID-19 pneumonia progression to severe symptom based on the earlier clinical features: a retrospective analysis. medRxiv [Preprint] 2020. doi:10.1101/2020.03.28.20045989
61 Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with covid-19 using artificial intelligence to help medical decision-making. medRxiv [Preprint] 2020. doi:10.1101/2020.03.30.20047308
62 Zeng L, Li J, Liao M, et al. Risk assessment of progression to severe conditions for patients with COVID-19 pneumonia: a single-center retrospective study. medRxiv [Preprint] 2020. doi:10.1101/2020.03.25.20043166
63 Al-Najjar H, Al-Rousan N. A classifier prediction model to predict the status of coronavirus covid-19 patients in South Korea. *Eur Rev Med Pharmacol Sci* 2020;24:3400-3. doi:10.26355/eurrev_202003_20709
64 Angelov P, Soares E. Explainable-by-design approach for covid-19 classification via CT-scan. medRxiv [Preprint] 2020. doi:10.1101/2020.04.24.20078584
65 Arpan M, Surya K, Harish R, et al. CovidAID: COVID-19 Detection Using Chest X-Ray. ArXiv e-prints [Preprint] 2020
66 Bai HX, Wang R, Xiong Z, et al. AI augmentation of radiologist performance in distinguishing covid-19 from pneumonia of other etiology on chest CT. *Radiology* 2020;201491. doi:10.1148/radiol.2020201491
67 Barda N, Riesel D, Akriv A, et al. Performing risk stratification for COVID-19 when individual level data is not available, the experience of a large healthcare organization. medRxiv [Preprint] 2020. doi:10.1101/2020.04.23.20076976
68 Bassi PRAS, Attux R. A deep convolutional neural network for covid-19 detection using chest x-rays. ArXiv e-prints [Preprint] 2020
69 Batista AFdM. Miraglia JL, Donato THR, et al. COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. medRxiv [Preprint] 2020. doi:10.1101/2020.04.04.20052092
70 Bello-Chavolla OY, Bahena-Lopez JP, Antonio-Villa NE, et al. Predicting mortality attributable to SARS-CoV-2: A mechanistic score relating obesity and diabetes to COVID-19 outcomes in Mexico. medRxiv [Preprint] 2020. doi:10.1101/2020.04.20.20072223
71 Benchoufi M, Bokobza J, Anthony C, et al. Lung injury in patients with or suspected COVID-19 : a comparison between lung ultrasound and chest CT-scanner severity assessments, an observational study. MedRxiv [Preprint] 2020. doi:10.1101/2020.04.24.20069633
72 Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. *Radiol Med* 2020;125:509-13. doi:10.1007/s11547-020-01200-3
73 Born J, Brandle G, Cossio M, et al. POCOVID-Net: Automatic detection of covid-19 from a new lung ultrasound imaging dataset (POCUS). ArXiv e-prints [Preprint] 2020.
74 Brinati D, Campagner A, Ferrari D, et al. Detection of covid-19 infection from routine blood exams with machine learning: a feasibility study. medRxiv [Preprint] 2020. doi:10.1101/2020.04.22.20075143
75 Carr E, Bendayan R, O'Gallagher K, et al. Supplementing the National Early Warning Score (NEWS2) for anticipating early deterioration among patients with covid-19 infection. medRxiv [Preprint] 2020. doi:10.1101/2020.04.24.20078006
76 Castiglioni I, Ippolito D, Interlenghi M, et al. Artificial intelligence applied on chest X-ray can aid in the diagnosis of COVID-19 infection: a first experience from Lombardy. medRxiv [Preprint] 2020. doi:10.1101/2020.04.08.20040907
77 Chassagnon G, Vakalopoulou M, Battistella E, et al. AI-driven CT-based quantification, staging and short-term outcome prediction of covid-19 pneumonia. medRxiv [Preprint] 2020. doi:10.1101/2020.04.17.20069187

78 Chen X, Tang Y, Mo Y, et al. A diagnostic model for coronavirus disease 2019 (COVID-19) based on radiological semantic and clinical features: a multi-center study. *Eur Radiol* 2020. doi:10.1007/s00330-020-06829-2.
79 Colombi D, Bodini FC, Petrini M, et al. Well-aerated lung on admitting chest CT to predict adverse outcome in covid-19 pneumonia. *Radiology* 2020;201433. doi:10.1148/radiol.2020201433
80 Das A, Mishra S, Gopalan SS. Predicting community mortality risk due to CoVID-19 using machine learning and development of a prediction tool. medRxiv [Preprint] 2020. doi:10.1101/2020.04.27.20081794
81 Diaz-Quijano FA, Silva JMNd, Ganem F, et al. A model to predict SARS-CoV-2 infection based on the first three-month surveillance data in Brazil. medRxiv [Preprint] 2020. doi:10.1101/2020.04.05.2004794 4.
82 Guiot J, Vaidyanathan A, Deprez L, et al. Development and validation of an automated radiomic CT signature for detecting covid-19. medRxiv [Preprint] 2020. doi:10.1101/2020.04.28.20082966
83 Guo Y, Liu Y, Lu J, et al. Development and validation of an early warning score (EWAS) for predicting clinical deterioration in patients with coronavirus disease 2019. medRxiv [Preprint] 2020. doi:10.1101/2020.04.17.20064691
84 Hu C, Liu Z, Jiang Y, et al. Early prediction of mortality risk among severe covid-19 patients using machine learning. medRxiv [Preprint] 2020. doi:10.1101/2020.04.13.20064329
85 Hu H, Yao N, Qiu Y. Comparing rapid scoring systems in mortality prediction of critical ill patients with novel coronavirus disease. *Acad Emerg Med* 2020;27:461-8. doi:10.1111/acem.13992
86 Hu R, Ruan G, Xiang S, et al. Automated diagnosis of covid-19 using deep learning and data augmentation on chest CT. medRxiv [Preprint] 2020. doi:10.1101/2020.04.24.20078998
87 Islam MT, Fleischer JW. Distinguishing L and H phenotypes of covid-19 using a single x-ray image. medRxiv [Preprint] 2020. doi:10.1101/2020.04.27.20081984
88 Ji D, Zhang D, Xu J, et al. Prediction for progression risk in patients with covid-19 pneumonia: the CALL score. *Clin Infect Dis* 2020;ciaa414. doi:10.1093/cid/ciaa414
89 Jiang X, Coffee M, Bari A, et al. Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. Computers. *Materials & Continua* 2020;63:537-51. doi:10.32604/cmc.2020.010691
90 Jiang Z, Hu M, Fan L, et al. Combining visible light and infrared imaging for efficient detection of respiratory infections such as covid-19 on portable device. ArXiv e-prints [Preprint] 2020.
91 Kana GEB, Kana ZMG, Kana DAF, et al. A web-based diagnostic tool for covid-19 using machine learning on chest radiographs (CXR). medRxiv [Preprint] 2020. doi:10.1101/2020.04.21.20063263
92 Rezaul KM, Döhmen T, Rebholz-Schuhmann D, et al. DeepCOVIDExplainer: explainable covid-19 predictions based on chest x-ray images. ArXiv e-prints [Preprint] 2020.
93 Khan AI, Shah JL, Bhat M. CoroNet: a deep neural network for detection and diagnosis of covid-19 from chest x-ray images. ArXiv e-prints [Preprint] 2020.
94 Kumar R, Arora R, Bansal V, et al. Accurate prediction of covid-19 using chest x-ray images through deep feature learning model with SMOTE and machine learning classifiers. medRxiv [Preprint] 2020. doi:10.1101/2020.04.13.20063461
95 Kurstjens S, van der Horst A, Herpers R, et al. Rapid identification of SARS-CoV-2-infected patients at the emergency department using routine testing. medRxiv [Preprint] 2020. doi:10.1101/2020.04.20.20067512
96 Levy TJ, Richardson S, Coppa K, et al. Estimating survival of hospitalized covid-19 patients from admission information. medRxiv [Preprint] 2020. doi:10.1101/2020.04.22.20075416
97 Li Z, Zhong Z, Li Y, et al. From community acquired pneumonia to covid-19: a deep learning based method for quantitative analysis of covid-19 on thick-section CT scans. medRxiv [Preprint] 2020. doi:10.1101/2020.04.17.20070219
98 Liu Q, Fang X, Tokuno S, et al. Prediction of the clinical outcome of COVID-19 patients using T lymphocyte subsets with 340 cases from Wuhan, China: a retrospective cohort study and a web visualization tool. medRxiv [Preprint] 2020. doi:10.1101/2020.04.06.20056127
99 Lyu P, Liu X, Zhang R, Shi L, Gao J. The performance of chest CT in evaluating the clinical severity of COVID-19 pneumonia: identifying critical cases based on CT characteristics. *Invest Radiol* 2020;55:412-21. doi:10.1097/RLI.0000000000000689
100 McRae MP, Simmons GW, Christodoulides NJ, et al. Clinical decision support tool and rapid point-of-care platform for determining disease severity in patients with covid-19. medRxiv [Preprint] 2020. doi:10.1101/2020.04.16.20068411
101 Mei X, Lee HC, Diao K, et al. Artificial intelligence for rapid identification of the coronavirus disease 2019 (covid-19). medRxiv [Preprint] 2020. doi:10.1101/2020.04.12.20062661
102 Menni C, Valdes A, Freydin MB, et al. Loss of smell and taste in combination with other symptoms is a strong predictor of COVID-19 infection. medRxiv [Preprint] 2020. doi:10.1101/2020.04.05.2004 8421.
103 Moutounet-Cartan PGB. Deep convolutional neural networks to diagnose covid-19 and other pneumonia diseases from posteroanterior chest x-rays. ArXiv e-prints [Preprint] 2020
104 Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput Biol Med* 2020;121:103792. doi:10.1016/j.compbiomed.2020.103792.
105 Rahimzadeh M, Attar A. A new modified deep convolutional neural network for detecting covid-19 from x-ray images. ArXiv e-prints [Preprint] 2020
106 Rehman A, Naz S, Khan A, et al. Improving coronavirus (covid-19) diagnosis using deep transfer learning. medRxiv [Preprint] 2020. doi: 10.1101/2020.04.11.20054643.
107 Singh D, Kumar V, Vaishali NA, Kaur M. Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks. *Eur J Clin Microbiol Infect Dis* 2020;39:1379-89. doi:10.1007/s10096-020-03901-z.
108 Singh K, Valley TS, Tang S, et al. Validating a widely implemented deterioration index model among hospitalized covid-19 patients. medRxiv [Preprint] 2020. doi:10.1101/2020.04.24.20079012.
109 Soares F, Villavicencio A, Anzanello MJ, et al. A novel high specificity COVID-19 screening method based on simple blood exams and artificial intelligence. medRxiv [Preprint] 2020. doi:10.1101/2020.0 4.10.20061036.
110 Tordjman M, Mekki A, Mali RD, et al. Pre-test probability for SARS-Cov-2-related Infection Score: the PARIS score. medRxiv [Preprint] 2020. doi:10.1101/2020.04.28.20081687.
111 Ucar F, Korkmaz D. COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Med Hypotheses* 2020;140:109761. doi:10.1016/j.mehy.2020.109761
112 Vaid A, Somani S, Russak AJ, et al. Machine learning to predict mortality and critical events in covid-19 positive New York City patients. medRxiv [Preprint] 2020. doi:10.1101/2020.04.26.20073411
113 Vazquez Guillamet C, Vazquez Guillamet R, Kramer AA, et al. Toward a covid-19 score-risk assessments and registry. medRxiv [Preprint] 2020. doi:10.1101/2020.04.15.20066860
114 Wang c, Deng R, Gou L, et al. Preliminary study to identify severe from moderate cases of COVID-19 using NLR&RDW-SD combination parameter. medRxiv [Preprint] 2020. doi:10.1101/2020.04.09.20058594
115 Wu Y-H, Gao S-H, Mei J, et al. JCS: an explainable covid-19 diagnosis system by joint classification and segmentation. ArXiv e-prints [Preprint] 2020.
116 Zhang H, Shi T, Wu X, et al. Risk prediction for poor outcome and death in hospital in-patients with COVID-19: derivation in Wuhan, China and external validation in London. medRxiv [Preprint] 2020. doi:10.1101/2020.04.28.20082222
117 Zhao B, Wei Y, Sun W, et al. Distinguish coronavirus disease 2019 patients in general surgery emergency by CIAAD scale: development and validation of a prediction model based on 822 cases in China. medRxiv [Preprint] 2020. doi:10.1101/2020.04.18.20071019
118 Zhu Z, Cai T, Fan L, et al. Clinical value of immune-inflammatory parameters to assess the severity of coronavirus disease 2019. *Int J Infect Dis* 2020;95:332-9. doi:10.1016/j.ijid.2020.04.041
119 Gong J, Ou J, Qiu X, et al. A tool to early predict severe corona virus disease 2019 (covid-19) : a multicenter study using the risk nomogram in Wuhan and Guangdong, China. *Clin Infect Dis* 2020;ciaa443. doi:10.1093/cid/ciaa443
120 Cohen JP, Morrison P, Dao L. COVID-19 image data collection. arXiv e-prints [Preprint] 2020. https://github.com/ieee8023/covid-chestxray-dataset.
121 Kaggle. COVID-19 Kaggle community contributions 2020. https://www.kaggle.com/covid-19-contributions.
122 ClosedLoop.ai. Covid-19 vulnerability index (CV19 index) 2020. https://closedloop.ai/cv19index/.
123 Chinese PLA General Hospital. Suspected covid-19 pneumonia diagnosis aid system 2020. https://intensivecare.shinyapps.io/COVID19/.
124 Brinati D. ML-based covid-19 test from routine blood test. 2020. https://covid19-blood-ml.herokuapp.com/.
125 Kurstjens S. Corona Score. 2020. https://corona-score.nvkc.nl/
126 Renmin Hospital of Wuhan University & Wuhan EndoAngel Medical Technology. AI diagnostic system for 2019-nCoV 2020. http://121.40.75.149/znyx-ncov/index.
127 National Supercomputing Center of Tianjin. Peunomnia CT 2020. https://ai.nscc-tj.cn/thai/deploy/public/pneumonia_ct.
128 Sun Yat-sen University. Discriminating covid-19 pneumonia from CT images 2020. http://biomed.nscc-gz.cn/server/Ncov2019.
129 Born J. POCIVIDSCREEN. 2020. https://pocovidscreen.org/.
130 AI SOCKET. Medics image inference socket. 2020. https://medics-inference.onrender.com/.

131 Das A. CoVID-19 Community Mortality Risk Prediction (CoCoMoRP) tool. 2020. https://ashis-das.shinyapps.io/CoCoMoRP/.
132 Hu C. Risk scores. 2020. https://phenomics.fudan.edu.cn/risk_scores/.
133 Health N. Northwell covid19 survival (NOCOS) calculator. 2020. https://cbmi.northwell.edu/nocos/.
134 Fang M. Covid-19 patient condition monitor 2020. https://rpubs.com/mindyfang/covid19.
135 Washington University School of Medicine. Towards a covid-19 score calculator. 2020. https://covid19score.azurewebsites.net/.
136 Zhang H. Covid-19 prediction. 2020. https://covid.datahelps.life/about/.
137 Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016;35:214-26. doi:10.1002/sim.6787.
138 Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005;58:475-83. doi:10.1016/j.jclinepi.2004.06.017.
139 Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:m441. doi:10.1136/bmj.m441.
140 Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. *Circulation* 2016;133:601-9. doi:10.1161/CIRCULATIONAHA.115.017719.
141 Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019;17:230. doi:10.1186/s12916-019-1466-7
142 Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges [correction: *BMJ* 2019;365:l4379]. *BMJ* 2016;353:i3140. doi:10.1136/bmj.i3140.
143 Debray TP, Riley RD, Rovers MM, Reitsma JB, Moons KG, Cochrane IPD Meta-analysis Methods group. Individual participant data (IPD) meta-analyses of diagnostic and prognostic modeling studies: guidance on their use. *PLoS Med* 2015;12:e1001886. doi:10.1371/journal.pmed.1001886.
144 Steyerberg EW, Harrell FEJr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* 2016;69:245-7. doi:10.1016/j.jclinepi.2015.04.005.
145 Wynants L, Kent DM, Timmerman D, Lundquist CM, Van Calster B. Untapped potential of multicenter studies: a review of cardiovascular risk prediction models revealed inappropriate analyses and wide variation in reporting. *Diagn Progn Res* 2019;3:6. doi:10.1186/s41512-019-0046-9.
146 Wynants L, Riley RD, Timmerman D, Van Calster B. Random-effects meta-analysis of the clinical utility of tests and prediction models. *Stat Med* 2018;37:2034-52. doi:10.1002/sim.7653.
147 Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395:1054-62. doi:10.1016/S0140-6736(20)30566-3.
148 Li K, Wu J, Wu F, et al. The clinical and chest CT features associated with severe and critical covid-19 pneumonia. *Invest Radiol* 2020. doi:10.1097/RLI.0000000000000672.
149 Li B, Yang J, Zhao F, et al. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. *Clin Res Cardiol* 2020. doi:10.1007/s00392-020-01626-9.
150 Jain V, Yuan J-M. Systematic review and meta-analysis of predictive symptoms and comorbidities for severe COVID-19 infection. medRxiv [Preprint] 2020. doi:10.1101/2020.03.15.20035360
151 Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, et al, Latin American Network of Coronavirus Disease 2019-COVID-19 Research (LANCOVID-19).https://www.lancovid.org. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. *Travel Med Infect Dis* 2020:101623. doi:10.1016/j.tmaid.2020.101623.
152 Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: a meta-analysis. *Clin Chim Acta* 2020;506:145-8. doi:10.1016/j.cca.2020.03.022.
153 Zhao X, Zhang B, Li P, et al. Incidence, clinical characteristics and prognostic factor of patients with covid-19: a systematic review and meta-analysis. medRxiv [Preprint] 2020. doi:10.1101/2020.03.17.20037572
154 Johansson MA, Saderi D. Open peer-review platform for COVID-19 preprints. *Nature* 2020;579:29. doi:10.1038/d41586-020-00613-4
155 Xu B, Kraemer MU, Gutierrez B, et al. Open access epidemiological data from the COVID-19 outbreak. *Lancet Infect Dis* 2020. doi:10.1016/s1473-3099(20)30119-5
156 Società Italiana di Radiologia Medica e Interventistica. COVID-19 database 2020. https://www.sirm.org/category/senza-categoria/covid-19/.
157 Dutch CardioVascular Alliance. European registry of patients with covid-19 including cardiovascular risk and complications 2020. https://capacity-covid.eu/.
158 World Health Organization. Coronavirus disease (COVID-19) technical guidance: early investigations protocols 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/early-investigations.
159 Infervision. Infervision launches hashtag#AI-based hashtag#Covid-19 solution in Europe 2020. https://www.linkedin.com/posts/infervision_ai-covid-medicine-activity-6650772755031613440-TqLJ.
160 Surgisphere Corporation. COVID-19 response center 2020. https://surgisphere.com/covid-19-response-center/.
161 Van Calster B, Wynants L, Timmerman D, Steyerberg EW, Collins GS, et al. Predictive analytics in health care: how can we know it works?*J Am Med Inform Assoc* 2019;26:1651-4. doi:10.1093/jamia/ocz130.
162 Enfield K, Miller R, Rice T. Limited utility of SOFA and APACHE II prediction models for ICU triage in pandemic Influenza. *Chest* 2011;140:913A. doi:10.1378/chest.1118087.
163 Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. *Med Decis Making* 2015;35:162-9. doi:10.1177/0272989X14547233.
164 van Smeden M, Moons KG, de Groot JA, et al. Sample size for binary logistic prediction models: beyond events per variable criteria. *Stat Methods Med Res* 2019;28:2455-74. doi:10.1177/0962280218784726.

Web appendix: Supplementary material