FAST TRACK

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

Laure Wynants,[1,2] Ben Van Calster,[2,3] Marc M J Bonten,[4,5] Gary S Collins,[6,7] Thomas P A Debray,[4,8] Maarten De Vos,[2,9] Maria C Haller,[10,11] Georg Heinze,[10] Karel G M Moons,[4,8] Richard D Riley,[12] Ewoud Schuit,[4,8] Luc J M Smits,[1] Kym I E Snell,[12] Ewout W Steyerberg,[3] Christine Wallisch,[10,13,14] Maarten van Smeden[4]

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

Accepted: 31 March 2020

ABSTRACT

OBJECTIVE

To review and critically appraise 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 risk of being admitted to hospital for covid-19 pneumonia.

DESIGN

Rapid systematic review and critical appraisal.

DATA SOURCES

PubMed and Embase through Ovid, Arxiv, medRxiv, and bioRxiv up to 24 March 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

2696 titles were screened, and 27 studies describing 31 prediction models were included. Three models were identified for predicting hospital admission from pneumonia and other events (as proxy outcomes for covid-19 pneumonia) in the general population; 18 diagnostic models for detecting covid-19 infection (13 were machine learning based on computed tomography scans); and 10 prognostic models for predicting mortality risk, progression to severe disease, or length of hospital stay. Only one study used patient data from outside of China. The most reported predictors of presence of covid-19 in patients with suspected disease included age, body temperature, and signs and symptoms. The most reported predictors of severe prognosis in patients with covid-19 included age, sex, features derived from computed tomography scans, C reactive protein, lactic dehydrogenase, and lymphocyte count. C index estimates ranged from 0.73 to 0.81 in prediction models for the general population (reported for all three models), from 0.81 to more than 0.99 in diagnostic models (reported for 13 of the 18 models), and from 0.85 to 0.98 in prognostic models (reported for six of the 10 models). All studies 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, and high risk of model overfitting. Reporting quality varied substantially between studies. Most reports did not include a description of the study population or intended use of the models, and calibration of 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 performance is probably optimistic. Immediate sharing of well documented individual participant data from covid-19 studies is needed for collaborative efforts to develop more rigorous prediction models and validate existing ones. The predictors identified in included studies could 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.

WHAT IS ALREADY KNOWN ON THIS TOPIC

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

Viral nucleic acid testing and chest computed tomography (CT) 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 covid-19 infection

WHAT THIS STUDY ADDS

Three models were identified that predict hospital admission from pneumonia and other events (as proxy outcomes for covid-19 pneumonia) in the general population

Eighteen diagnostic models were identified for detecting covid-19 infection (13 were machine learning based on CT scans); and 10 prognostic models for predicting mortality risk, progression to severe disease, or length of hospital stay

Proposed models are poorly reported and at high risk of bias, raising concern that their predictions could be unreliable when applied in daily practice

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 775 000 in more than 160 countries, and the number of people infected is probably much higher. More than 36 000 people have died from covid-19 infection (up to 30 March 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 will almost certainly 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 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 currently available prediction models for covid-19, in particular diagnostic and prognostic models for the disease. This systematic review was carried out 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 the list to examine whether it is fit for purpose by comparing it to relevant hits from bioRxiv and medRxiv when combining covid-19 search terms (covid-19, sars-cov-2, novel corona, 2019-ncov) with methodological search terms (diagnostic, prognostic, prediction model, machine learning, artificial intelligence, algorithm, score, deep learning, regression). All relevant hits were found on the living systematic review list.[6] We supplemented this list with hits from PubMed by searching for "covid-19" because when we performed our intitial search this term was not included in the reported living systematic review[6] search terms for PubMed. We further supplemented the list 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 initially searched databases on 13 March 2020, with an update on 24 March 2020. All studies were considered, regardless of language or publication status (preprint or peer reviewed articles). 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 diagnostic and prognostic models for covid-19, or those aiming to identify people at increased risk of developing covid-19 pneumonia 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. Titles, abstracts, and full texts were screened in duplicate for eligibility by pairs of independent reviewers (from LW, BVC, and MvS), 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, 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[13] and PROBAST (prediction model risk of bias assessment tool).[14] 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.[15] 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).[16] We focus on performance statistics as estimated from the strongest available form of validation. Any discrepancies in data extraction 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)[17] and TRIPOD[15] in reporting our article.

doi: 10.1136/bmj.m1328 | BMJ 2020;369:m1328 | the bmj

Patient and public involvement

It was not appropriate or 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 2690 titles through our systematic search (fig 1; 1916 on 13 March 2020 and 774 during an update on 24 March 2020). Two additional unpublished studies were made available on request (after a call on social media). We included a further four studies that were publicly available but were not detected by our search. Of 2696 titles, 85 studies were retained for abstract and full text screening. Twenty seven studies describing 31 prediction models met the inclusion criteria and were selected for data extraction and critical appraisal.[7-12 18-38]

Primary datasets

Twenty five studies used data on patients with covid-19 from China (supplementary table 1), one study used data on patients from Italy,[31] and one study used international data (United States, United Kingdom, and China, among others).[35] Based on 18 of the 25 studies that reported study dates, data were collected between 8 December 2019 and 15 March 2020. The duration of follow-up was unclear in most studies, although one reported a median follow-up of 8.4 days,[19] while another reported a median follow-up of 15 days.[37] Some Chinese centres provided data to multiple studies, but it was unclear how much these datasets overlapped across our 25 identified studies. One study used US Medicare claims data from 2015 to 2016 to estimate vulnerability to covid-19,[8] two studies used control CT (computed tomography) scans from the US or Switzerland,[11 25] and one study used simulated data.[18]All but one study[24] developed prediction models for use in adults. The median age varied between studies (from 34 to 65 years; see supplementary table 1), as did the proportion of men (from 41% to 61%).

Among the six studies that developed prognostic models to predict mortality risk in people with confirmed or suspected covid-19 infection, the percentage of deaths varied between 8% and 59% (table 1). This wide variation is partly because of severe 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 20-22] 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 18 diagnostic model studies, only one reported on prevalence of covid-19 infection in people with suspected covid-19; the prevalence was 19% (development dataset) and 24% (validation dataset).[30] One study reported that 8% of patients had severe disease among confirmed paediatric patients with covid-19 infection.[24] Because 16 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 gives an overview of the 31 prediction models reported in the 27 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 risk of hospital admission for covid-19 pneumonia in general population

We identified three models that predicted risk of hospital admission for covid-19 pneumonia in the general population, but used admission for non-tuberculosis pneumonia, influenza, acute bronchitis, or upper respiratory tract infections as outcomes in a dataset without any patients with covid-19 (table 1).[8] Among the predictors were age, sex, previous hospital admissions, comorbidity data, and social determinants of health. The study estimated C indices of 0.73, 0.81, and 0.81 for the three models.

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

We identified one study that developed a model to detect covid-19 pneumonia in fever clinic patients (estimated C index 0.94)[10]; one to diagnose covid-19 in patients with suspected disease (estimated C index 0.97)[30]; one to diagnose covid-19 in patients with suspected disease and asymptomatic patients (estimated C index 0.87)[12]; and one to diagnose covid-19 by using deep learning of genomic sequences (estimated C index 0.98).[35] A further study was developed to diagnose severe disease in paediatric inpatients with symptoms, based on direct bilirubin and alanine transaminase (reporting an F1 score of 1.00, indicating 100% observed sensitivity and specificity).[24] Only one study reported assessing calibration, but it was unclear how this was done.[12] Predictors used in more than one model were age (n=3), body temperature or fever (n=2), and signs and symptoms (such as shortness of breath, headache, shiver, sore throat, and fatigue, n=2; table 1).

Thirteen prediction models were proposed to support the diagnosis of covid-19 or covid-19 pneumonia (and monitor progression) based on CT images. The predictive performance varied widely, with estimated C index values ranging from 0.81 to nearly 1.

Prognostic models for patients with a diagnosis of covid-19 infection

We identified 10 prognostic models (table 1). Of these, six estimated mortality risk in patients with suspected or confirmed covid-19.[7 18 19 21 22 37] The intended use of these models (that is, when to use them, in whom to use them, and the prediction horizon, eg, mortality by what time) was not clearly described. Two models aimed to predict a hospital stay of more than 10 days

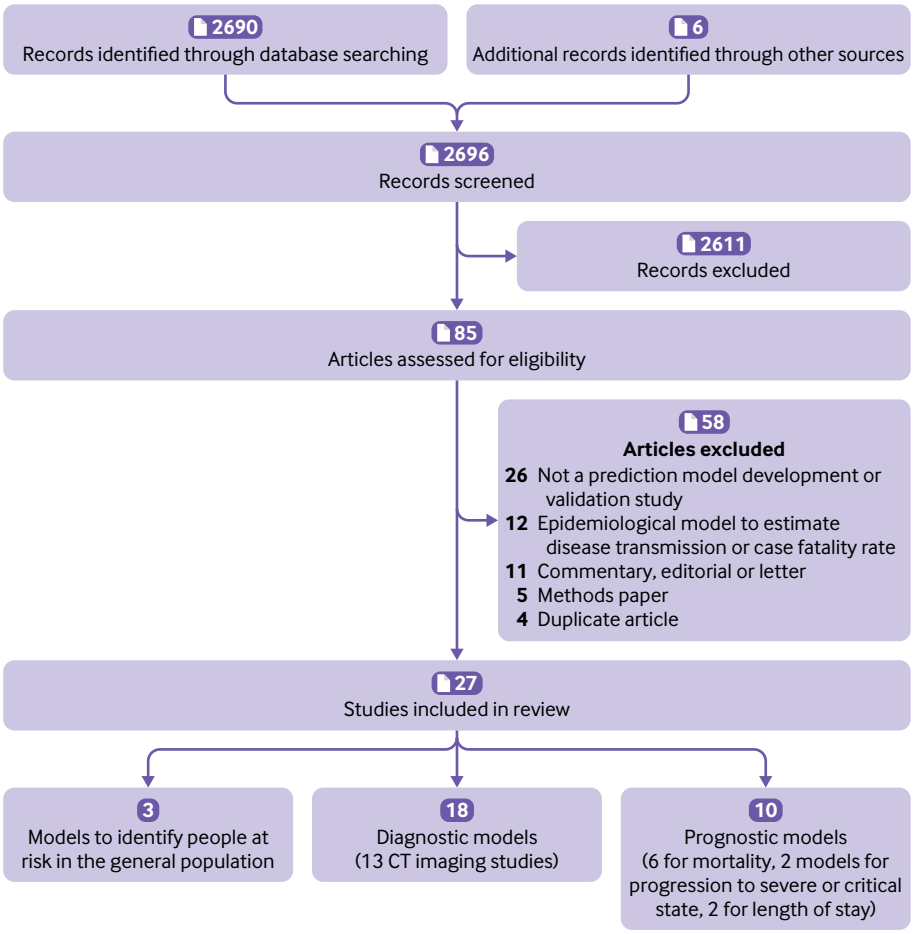

Fig 1 | PRISMA (preferred reporting items for systematic reviews and meta-analyses) flowchart of study inclusions and exclusions. CT=computed tomography

from admission.[20] Two models aimed to predict progression to a severe or critical state.[9 32] Predictors included in more than one prognostic model were age (n=5), sex (n=2), features derived from CT scoring (n=5), C reactive protein (n=3), lactic dehydrogenase (n=3), and lymphocyte count (n=2; table 1).

Only two studies that predicted mortality reported a C index; these studies obtained estimates of 0.90[22] and 0.98.[7] One study also evaluated calibration.[7] When applied to new patients, their model 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] One study developed two models to predict a hospital stay of more than 10 days and estimated C indices of 0.92 and 0.96.[20] The two studies that developed models to predict progression to a severe or critical state estimated C indices of 0.95 and 0.85.[9 32] One of these studies also reported perfect calibration, but it was unclear how this was evaluated.[32]

Risk of bias

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

Eleven of the 27 studies had a high risk of bias for the participants domain (table 2), 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 eight studies. Four of the 27 studies had a high risk of bias for the predictors domain, which indicates that predictors were not available at the models' intended time of use, not clearly defined, or influenced by the outcome measurement. The diagnostic model studies that used CT imaging predictors were all scored as unclear on the predictors domain. The publications often lacked clear information on the preprocessing steps (eg, cropping of images). Moreover, complex machine learning algorithms transform CT images into predictors in an untransparent 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 reason to be concerned about bias induced by the outcome measurement in 10 studies

Table 1 | Overview of prediction models for diagnosis and prognosis of covid-19 infection

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			Overall risk of bias using PROBAST
			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))	
Hospital admission in general population						
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
Diagnosis						
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, sensivitity 100, 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[30]; 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
Yu et al[24]; 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
Diagnostic imaging						
Barstugan et al[31]; 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[26]; 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[25]; 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[28]; 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 pneunomia	5 categories of location features from imaging: volume, number, histogram, surface, radiomics	2685 (1658)	5-fold cross validation	Not applicable	C index 0.94	High

(*Continued*)

Table 1 | 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			Overall risk of bias using PROBAST
			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))	
Wang et al[29]; 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[27]; 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[23]; 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[23]; 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
Prognosis						
Bai et al[9]; data from China, inpatients at admission with mild confirmed covid-19 infection; 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[18]; 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
Gong et al[32]; data from China, inpatients with confirmed covid-19 at admission; severe covid-19 infection (within minimum 15 days)	Age, serum LDH, CRP, variation of red blood cell distribution width, blood urea nitrogen, albumin, direct bilirubin	189 (28)	External validation (two centres)	165 (40) and 18 (4)	Centre 1: C index 0.85 (0.79 to 0.92), sensitivity 78, specificity 78; centre 2: sensitivity 75, specificity 100	High
Lu et al[19]; 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[20]; 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)	5 fold cross validation	Not applicable	C index 0.92	High
Qi et al[20]; 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, SPO_2	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[21]; 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[22]; 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

Covid-19=coronavirus disease 2019; CRP=C reactive protein; CT=computed tomography; LDH=lactate dehydrogenase; NPV=negative predictive value; PPV=positive predictive value; PROBAST=prediction model risk of bias assessment tool; SD=standard deviation; SPO_2=oxygen saturation.
*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).
‡Calibration plot presented, but unclear on which data were used.
§Development set contains scans from Chinese patients, the testing set contained 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.
††Outcome and predictor data were simulated.
‡‡Wavelet-HLH_gldm_SmallDependenceLowGrayLevelEmphasis, wavelet-LHH_glcm_Correlation, wavelet-LHL_glszm_GrayLevelV ariance, wavelet-LLH_glszm_SizeZoneNonUniformityNormalized, wavelet-LLH_glszm_SmallAreaEmphasis, wavelet-LLH_glcm_Correlation.

doi: 10.1136/bmj.m1328 | BMJ 2020;369:m1328 | the bmj

because of the use of subjective or proxy outcomes (eg, non covid-19 severe respiratory infections).

All studies were at high risk of bias for the analysis domain (table 2). Many studies had small sample sizes (table 1), 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 one study reported only the apparent performance (the performance in exactly the same data used to develop the model, without adjustment for optimism owing to potential overfitting).

Four models were externally validated in the model development study (in an independent dataset, excluding random training test splits and temporal splits).[7 12 25 32] However, in three of these studies, the external validation datasets are probably not representative of the target population (box 2).[7 12 25] Consequently, predictive performance could differ if the models were applied in the target population. Gong and colleagues had a satisfactory predictive performance on two unbiased but small external validation datasets.[32] One study was a small (n=27) external validation 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 led to selection bias.[22] Only three studies assessed calibration,[7 12 32] but the method to check calibration was probably suboptimal in two studies.[12 32]

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

Twelve 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 hospital admission for coronavirus disease 2019 (covid-19) pneumonia 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.[8 39]

Diagnostic models

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] The "suspected COVID-19 pneumonia Diagnosis Aid System" is available as an online tool.[10 40] The "COVID-19 early warning score" to detect covid-19 infection in adults is available as a score chart in an article.[30] A decision tree to detect severe disease for paediatric patients with confirmed covid-19 is also available in an article.[24]

Diagnostic models based on computed tomography (CT) imaging

Three of the seven artificial intelligence models to assist with diagnosis based on CT images are available through web applications.[23 26 29 41-43] One model is deployed in 16 hospitals, but the authors do not provide any usable tools in their study.[33]

Prognostic models

To assist in the prognosis of mortality, a nomogram (a graphical aid to calculate mortality risk),[7] a decision tree,[21] and a CT based scoring rule are available in the articles.[22] Additionally a nomogram exists to predict progression to severe covid-19 disease.[32]

Five studies made their source code available on GitHub.[8 11 34 35 38] Ten studies did not include any usable equation, format, or reference for use or validation of their prediction model.

Discussion

In this systematic review of prediction models related to the covid-19 pandemic, we identified and critically appraised 27 studies that described 31 models. These prediction models were developed for detecting people in the general population at risk of being admitted to hospital for covid-19 pneumonia, for diagnosis of covid-19 in patients with symptoms, and for prognosis of patients with covid-19 infection. All models reported good 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. As expected, in these early covid-19 related prediction model studies, clinical data from patients with covid-19 are still scarce and limited to data from China, Italy, and international registries. 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.[44] A high risk of bias implies that these models will probably perform worse in practice than the performance reported by the researchers. Therefore, the estimated C indices, often close to 1 and indicating near perfect discrimination, are probably optimistic. Five studies carried out an external validation,[7 12 22 25 32] and only one study assessed calibration correctly.[7]

We reviewed 13 studies that used advanced machine learning methodology on chest CT scans to diagnose covid-19 disease, 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.

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. However, the included studies 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 caused by urgency, we recommend that any studies currently in preprint and all future studies

should adhere to the TRIPOD reporting guideline[15] 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.

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. The variability in the relative frequencies of the predicted outcomes presents an important challenge to the prediction modeller. A prediction model applied in a setting with a different relative frequency of the outcome might produce predictions that are miscalibrated[45] and might need to be updated before it can safely be applied in that new setting.[16 46] Such an update might often be required when prediction models are transported to different healthcare systems, which requires data from patients with covid-19 to be available from that system.

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.[16] 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.[47]

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.[48-52]

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[53-59] and open peer review initiatives for covid-19 related publications,[60] data registries[61-65] are being set up. To maximise the new opportunities and to facilitate individual participant data meta-analyses, the World Health Organization has recently released a new data platform to encourage sharing of anonymised covid-19 clinical data.[66] To leverage the full potential of these evolutions, international and interdisciplinary collaboration in terms of data acquisition and model building is crucial.

Box 2: Common causes of risk of bias in the 19 reported prediction models

Models to predict hospital admission for coronavirus disease 2019 (covid-19) pneumonia in general population

These models were based on Medicare claims data, and used proxy outcomes to predict hospital admission for covid-19 pneumonia, in the absence of patients with covid-19.[8]

Diagnostic models

People without covid-19 (or a proportion of them) were excluded, altering the disease prevalence.[30] Controls had viral pneumonia, which is not representative of the target population for a screening model.[12] The test used to determine the outcome varied between participants,[12] or one of the predictors (fever) was part of the outcome definition.[10] Predictors were dichotomised, which led to a loss of information.[24 30 36]

Diagnostic models based on computed tomography (CT) imaging

Generally, studies did not clearly report which patients had CT scans 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).[11 23 29 33 36] Often studies did not clearly report how regions of interest were annotated. Images were sometimes annotated by only one scorer without quality control,[25 27] the model output influenced annotation,[28] or the "ground truth" that was used to build the model was a composite outcome based on the same CT images used to make the prediction, among other factors.[38] Careful description of model specification and subsequent estimation were lacking, challenging the transparency and reproducibility of the models. Every study used a different deep learning architecture, 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 20-22] Additionally, only one study accounted for censoring by using Cox regression.[19] One study developed a model to predict future severity using cross sectional data (some participants were severely ill at inclusion)[37]; this implies that the timing of the measurement of the predictors is not appropriate and the (unclearly defined) outcome might have been influenced by the predictor values. Other studies used highly subjective predictors,[22] or 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).[21]

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, 24 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 also found other prediction models that are currently being used in clinical practice but without scientific publications,[67] and web risk calculators launched for use while the scientific manuscript is still under review (and unavailable on request).[68] These unpublished models naturally fall outside the scope of this review of the literature.

Implications for practice

All 31 reviewed prediction models were found to have a high risk of bias, and evidence from independent external validation of these 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 to implement prediction models without sufficient documentation and validation. Although we cannot let perfect be the enemy of good, earlier studies have

Table 2 | Risk of bias assessment (using PROBAST) based on four domains across 27 studies that created prediction models for coronavirus disease 2019

	Risk of bias			
Authors	Participants	Predictors	Outcome	Analysis
Hospital admission in general population				
DeCaprio et al[8]	High	Low	High	High
Diagnosis				
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[30]	High	Unclear	Low	High
Yu et al[24]	Unclear	Unclear	Unclear	High
Diagnostic imaging				
Barstugan et al[31]	Unclear	Unclear	Unclear	High
Chen et al[26]	High	Unclear	Low	High*
Gozes et al[25]	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[28]	Unclear	Unclear	High	High†
Shi et al[36]	High	Unclear	Low	High
Wang et al[29]	High	Unclear	Low	High
Xu et al[27]	High	Unclear	High	High
Song et al[23]	Unclear	Unclear	Low	High
Zheng et al[38]	Unclear	Unclear	High	High
Prognosis				
Bai et al[9]	Low	Unclear	Unclear	High
Caramelo et al[18]	High	High	High	High
Gong et al[32]	Low	Unclear	Unclear	High
Lu et al[19]	Low	Low	Low	High
Qi et al[20]	Unclear	Low	Low	High
Shi et al[37]	High	High	High	High
Xie et al[7]	Low	Low	Low	High
Yan et al[21]	Low	High	Low	High
Yuan et al[22]	Low	High	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.

shown that models were of limited use in the context of a pandemic,[69] and they could even cause more harm than good.[70] Therefore, we cannot recommend any model for use in practice at this point.

We anticipate that more covid-19 data at the individual participant level will soon become available. These data could be used to validate and update currently available prediction models.[16] For example, one model that predicted progression to severe covid-19 disease within 15 days of admission to hospital showed promising discrimination when validated externally on two small but unselected cohorts.[32] Because reporting in this study was insufficiently detailed and the validation was in small Chinese datasets, validation in larger, international datasets is needed. Owing to differences between healthcare systems (eg, Chinese and European) on when patients are admitted to and discharged from hospital, and testing criteria for patients with covid-19, we anticipate most existing models will need to be updated (that is, adjusted to the local setting).

When building 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[16]; this is especially true for datasets with limited sample size.[71] Based on the predictors included in multiple models identified by our review, we encourage researchers to consider incorporating several candidate predictors: for diagnostic models, these include age, body temperature, and (respiratory) signs and symptoms; for prognostic models, age, sex, C reactive protein, lactic dehydrogenase, lymphocyte count, and potentially features derived from CT scoring. Predictors that were included in both diagnostic and prognostic models were albumin (or albumin/globin), direct bilirubin, and red blood cell distribution width; these predictors could be considered as well. 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 systematic review aims to be the first stage of a living review of this field, 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.

Conclusion

Diagnostic and prognostic models for covid-19 are available and they all appear to show good to excellent discriminative performance. However, these models are 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 likely to be optimistic and misleading. Future studies should address these concerns. Sharing data and expertise for development, 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]Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht University, Utrecht, Netherlands
[5]Department of Medical Microbiology, University Medical Centre Utrecht, Utrecht, Netherlands
[6]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Musculoskeletal Sciences, University of Oxford, Oxford, UK
[7]NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK
[8]Cochrane Netherlands, University Medical Centre Utrecht, Utrecht University, Utrecht, Netherlands
[9]Department of Electrical Engineering, ESAT Stadius, KU Leuven, Leuven, Belgium
[10]Section for Clinical Biometrics, Centre for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria
[11]Ordensklinikum Linz, Hospital Elisabethinen, Department of Nephrology, Linz, Austria

[12]Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Keele, UK
[13]Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany
[14]Berlin Institute of Health, Berlin, Germany

We thank the authors who made their work available by posting it on public registries or sharing it confidentially.

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, and MvS extracted and analysed data. MDV helped interpret the findings on deep learning studies and MMJB and MCH assisted in the interpretation from a clinical viewpoint. 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 corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: 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 No 91617050). KGMM gratefully acknowledges financial support from Cochrane Collaboration (SMF 2018). KIES is funded by the National Institute for Health Research School for Primary Care Research (NIHR SPCR). 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). The funders played no role in study design, data collection, data analysis, data interpretation, or reporting. 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.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no competing interests with regards to the submitted work; LW discloses support from Research Foundation–Flanders (FWO); 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 also the Universities of Aberdeen, Exeter, and Leeds, all outside the submitted work.

Ethical approval: Not required.

Data sharing: The study protocol is available online at https://osf.io/ehc47/. Most included studies are publically 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/. A preprint version of the study is publicly available on medRxiv.

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 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
14 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
15 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
16 Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating.* Springer US, 2019. doi:10.1007/978-3-030-16399-0
17 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
18 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
19 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
20 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
21 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
22 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
23 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
24 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
25 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
26 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
27 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
28 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

29 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
30 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
31 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
32 Gong J, Ou J, Qiu X, et al. A tool to early predict severe 2019-novel coronavirus pneumonia (covid-19): a multicenter study using the risk nomogram in Wuhan and Guangdong, China. medRxiv [Preprint] 2020. doi:10.1101/2020.03.17.20037515
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 ClosedLoop.ai. Covid-19 vulnerability index (CV19 index) 2020. https://closedloop.ai/cv19index/.
40 Chinese PLA General Hospital. Suspected covid-19 pneumonia diagnosis aid system 2020. https://intensivecare.shinyapps.io/COVID19/.
41 Renmin Hospital of Wuhan University & Wuhan EndoAngel Medical Technology Co. AI diagnostic system for 2019-nCoV 2020. http://121.40.75.149/znyx-ncov/index.
42 National Supercomputing Center of Tianjin. Peunomnia CT 2020. https://ai.nscc-tj.cn/thai/deploy/public/pneumonia_ct.
43 Sun Yat-sen University. Discriminating covid-19 pneumonia from CT images 2020. http://biomed.nscc-gz.cn/server/Ncov2019.
44 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
45 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
46 Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating*. Springer US, 2009. doi:10.1007/978-0-387-77244-8
47 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
48 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
49 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
50 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
51 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
52 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
53 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
54 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
55 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
56 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
57 Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, et al, Latin American Network of Coronavirus Disease 2019-COVID-19 Research (LANCOVID-19). Electronic address: 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
58 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
59 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
60 Johansson MA, Saderi D. Open peer-review platform for COVID-19 preprints. *Nature* 2020;579:29. doi:10.1038/d41586-020-00613-4
61 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
62 Società Italiana di Radiologia Medica e Interventistica. COVID-19 database 2020. https://www.sirm.org/category/senza-categoria/covid-19/.
63 Kaggle. COVID-19 Kaggle community contributions 2020. https://www.kaggle.com/covid-19-contributions.
64 Cohen JP, Morrison P, Dao L. COVID-19 image data collection. arXiv [Preprint] 2020. doi:2003.11597, https://github.com/ieee8023/covid-chestxray-dataset.
65 Dutch CardioVascular Alliance. European registry of patients with covid-19 including cardiovascular risk and complications 2020. https://capacity-covid.eu/.
66 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.
67 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.
68 Surgisphere Corporation. COVID-19 response center 2020. https://surgisphere.com/covid-19-response-center/.
69 Enfield K, Miller R, Rice T, et al. Limited utility of SOFA and APACHE II prediction models for ICU triage in pandemic Influenza. *Chest* 2011;140:913A. doi:10.1378/chest.1118087
70 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
71 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