OPEN ACCESS

Check for updates

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 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] Michael Kammer,[7,19] Nina Kreuzberger,[20] Anna Lohmann,[21] Kim Luijken,[21] Jie Ma,[5] Glen P Martin,[22] David J McLernon,[23] Constanza L Andaur Navarro,[8,9] Johannes B Reitsma,[8,9] Jamie C Sergeant,[24,25] Chunhu Shi,[26] Nicole Skoetz,[19] Luc J M Smits,[1] Kym I E Snell,[6] Matthew Sperrin,[27] René Spijker,[8,9,28] Ewout W Steyerberg,[3] Toshihiko Takada,[8] Ioanna Tzoulaki,[29,30] Sander M J van Kuijk,[31] Bas C T van Bussel,[1,32] Iwan C C van der Horst,[32] Florien S van Royen,[8] Jan Y Verbakel,[33,34] Christine Wallisch,[7,35,36] Jack Wilkinson,[22] Robert Wolff,[37] 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: 12 January 2021

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 covid-19 infection 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, up to 1 July 2020, supplemented with 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

37 421 titles were screened, and 169 studies describing 232 prediction models were included. The review identified seven models for identifying people at risk in the general population; 118 diagnostic models for detecting covid-19 (75 were based on medical imaging, 10 to diagnose disease severity); and 107 prognostic models for predicting mortality risk, progression to severe disease, intensive care unit admission, ventilation, intubation, or length of hospital stay. The most frequent types of predictors included in the covid-19 prediction models are vital signs, age, comorbidities, and image features. Flu-like symptoms are frequently predictive in diagnostic models, while sex, C reactive protein, and lymphocyte counts are frequent prognostic factors. Reported C index estimates from the strongest form of validation available per model ranged from 0.71 to 0.99 in prediction models for the general population, from 0.65 to more than 0.99 in diagnostic models, and from 0.54 to 0.99 in prognostic models. All models were rated at high or unclear 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 unclear reporting. Many models did not include a description of the target population (n=27, 12%) or care setting (n=75, 32%), and only 11 (5%) were externally

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

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

Thirty three diagnostic models were identified for detecting covid-19, in addition to 75 diagnostic models based on medical images, 10 diagnostic models for severity classification, and 107 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

Two prediction models (one for diagnosis and one for prognosis) were identified as being of higher quality than others and efforts should be made to validate these in other datasets

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

validated by a calibration plot. The Jehi diagnostic model and the 4C mortality score were identified as promising models.

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 almost all pubished prediction models are poorly reported, and at high risk of bias such that their reported predictive performance is probably optimistic. However, we have identified two (one diagnostic and one prognostic) promising models that should soon be validated in multiple cohorts, preferably through collaborative efforts and data sharing to also allow an investigation of the stability and heterogeneity in their performance across populations and settings. Details on all reviewed models are publicly available at https://www.covprecise.org/. Methodological guidance as provided in this paper should be followed because unreliable predictions could cause more harm than benefit in guiding clinical decisions. Finally, prediction model authors 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 3 of the original article published on 7 April 2020 (*BMJ* 2020;369:m1328). Previous updates can be found as data supplements (https://www.bmj.com/content/369/bmj.m1328/related#datasupp). When citing this paper please consider adding the update number and date of access for clarity.

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 47 million as the disease spread globally, and the number of people infected is probably much higher. More than 1.2 million people have died from covid-19 (up to 3 November 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 can also become infected. Several regions have had or are experiencing second waves, and despite improvements in testing and tracing, several regions are again facing the limits of their test capacity, hospital resources and healthcare staff.[5 6]

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 are 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.[7]

We aimed to systematically review and critically appraise all currently available prediction models for covid-19, in particular models to predict the risk of covid-19 infection or being admitted to hospital with the disease, 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 international COVID-PRECISE (Precise Risk Estimation to optimise covid-19 Care for Infected or Suspected patients in diverse sEttings; https://www.covprecise.org/) group in collaboration with the Cochrane Prognosis Methods Group.

Methods

We searched the publicly available, continuously updated publication list of the covid-19 living systematic review.[8] 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. 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 to predict the presence or severity of covid-19 in patients with suspected infection; prognostic models to predict the course of infection in patients with covid-19; and prediction models to identify people in the general population at risk of covid-19 infection or at risk of being admitted to hospital with the disease.

We searched the database repeatedly up to 1 July 2020 (supplementary table 1). As of the third update (search date 1 July), we only include peer reviewed articles (indexed in PubMed and Embase through Ovid). Preprints (from bioRxiv, medRxiv, and arXiv) that were already included in previous updates of the systematic review remain included in the analysis. Reassessment takes place after publication of a preprint in a peer reviewed journal. No restrictions were made on the setting (eg, inpatients, outpatients, or general population), prediction horizon (how far ahead the model predicts), included predictors,

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

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

or outcomes. Epidemiological studies that aimed to model disease transmission or fatality rates, diagnostic test accuracy, and predictor finding studies were excluded. We focus on studies published in English. Starting with the second update, retrieved records were initially screened by a text analysis tool developed using 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,[9] 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, BCTvB, ICCvdH, DJM, MK, 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[10] and PROBAST (prediction model risk of bias assessment tool; www.probast.org) for assessing the reported prediction models.[11] 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; www.tripod-statement.org) statement.[12] 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).[13] 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 or MvS. The online supplementary material provides details on data extraction. Some studies investigated multiple models and some models were investigated in multiple studies (that is, in external validation studies). The unit of analysis was a model within a study, unless stated otherwise. We considered aspects of PRISMA (preferred reporting items for systematic reviews and meta-analyses)[14] and TRIPOD[12] in reporting our study. Details on all reviewed studies and prediction models are publicly available at https://www.covprecise.org/.

Patient and public involvement

It was not possible to involve patients or the public in the design, conduct, or reporting of our research. A lay summary of the project's aims is available on https://www.covprecise.org/project/. The study protocol and preliminary results are publicly available on https://osf.io/ehc47/, medRxiv and https://www.covprecise.org/living-review/.

Results

We retrieved 37 412 titles through our systematic search (of which 23 203 were included in the present update; supplementary table 1, fig 1). We included a further nine studies that were publicly available but were not detected by our search. Of 37 421 titles, 444 studies were retained for abstract and full text screening (of which 169 are included in the present update). One hundred sixty nine studies describing 232 prediction models met the inclusion criteria (of which 62 studies and 87 models added since the present update, supplementary table 1).[15-183] These studies were selected for data extraction and critical appraisal. The unit of analysis was the model within a study: of these 232 models, 208 were unique, newly developed models for covid-19. The remaining 24 analyses were external validations of existing models (in a study other than the model development study). Some models were validated more than once (in different studies, as described below). Many models are publicly available (box 1). A database with the description of each model and its risk of bias assessment can be found on https://www.covprecise.org/.

Primary datasets

One hundred seventy four (75%) models used data from a single country (table 1), 42 (18%) models used international data, and for 16 (7%) models it was unclear how many (and which) countries contributed data. Two (1%) models used simulated data and 12 (5%) used proxy data to estimate covid-19 related risks (eg, Medicare claims data from 2015 to 2016). Most models were intended for use in confirmed covid-19 cases (47%) and a hospital setting (51%). The average patient age ranged from 39 to 71 years, and the proportion of men ranged from 35% to 75%, although this information was often not reported. One study developed a prediction model for use in paediatric patients.[27]

Based on the studies that reported study dates, data were collected from December 2019 to June 2020. Some centres provided data to multiple studies and several studies used open Github[184] or Kaggle[185] data repositories (version or date of access often unspecified), and so it was unclear how much these datasets overlapped across our identified studies.

Among the diagnostic model studies, the reported prevalence of covid-19 varied between 7% and 71% (if a cross sectional or cohort design was used). Because 75 diagnostic studies used either case-control sampling or an unclear method of data collection, the

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

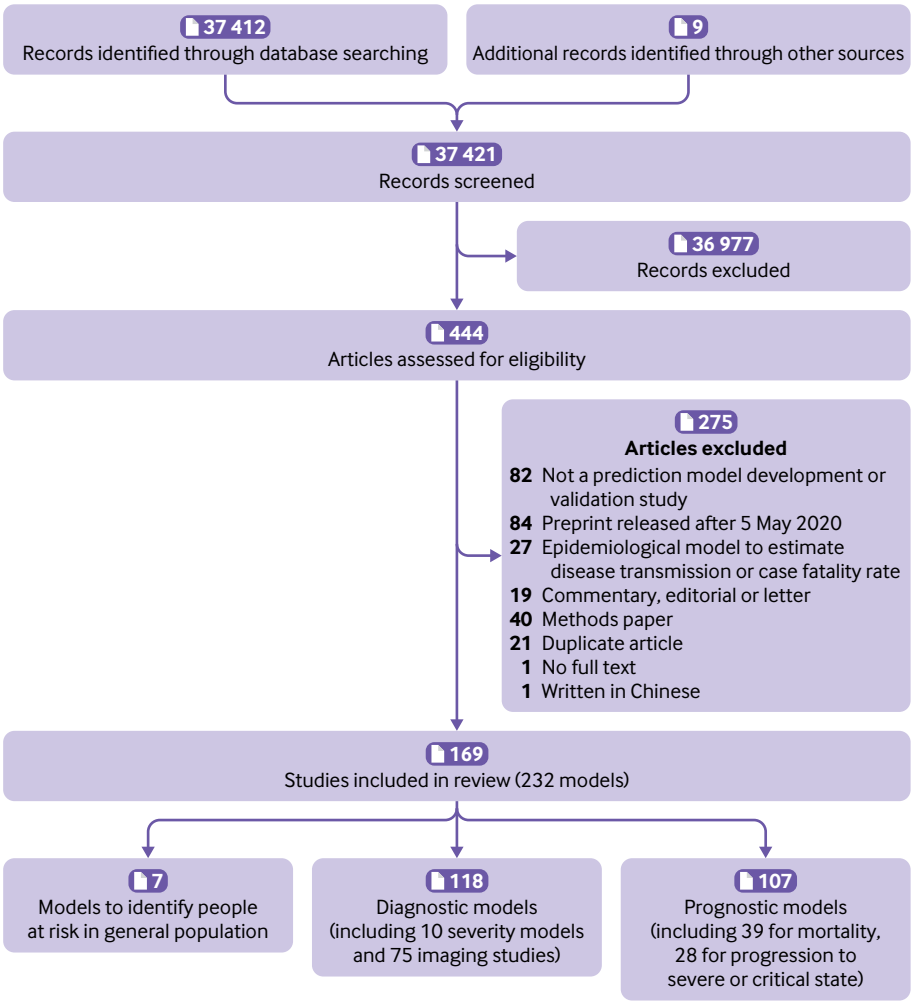


Fig 1 | PRISMA (preferred reporting items for systematic reviews and meta-analyses) flowchart of study inclusions and exclusions

prevalence in these diagnostic studies might not be representative of their target population.

Among the studies that developed prognostic models to predict mortality risk in people with confirmed or suspected infection, the percentage of deaths ranged from 1% to 52%. 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). Additionally, length of follow-up varied between studies (but was often not reported), and there is likely to 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).

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

Two hundred and eight unique models were developed in the included studies. Thirty (14%) of these models were presented as a model equation including intercept and regression coefficients. Eight (4%) models were only partially presented (eg, intercept or baseline hazard were missing). The remaining did not provide the underlying model equation.

Seventy two models (35%) are available as a tool for use in clinical practice (in addition to or instead of a published equation). Twenty seven models were presented as a web calculator (13%), 12 as a sum score (6%), 11 as a nomogram (5%), 8 as a software object (4%), 5 as a decision tree or set of predictions for subgroups (2%), 3 as a chart score (1%), and 6 in other usable formats (3%).

All these presentation formats make predictions readily available for use in the clinic. However, because all models were at high or uncertain risk of bias, we do not recommend their routine use before they are externally validated, ideally by independent investigators.

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

We identified seven models that predicted risk of covid-19 in the general population. Three models from one study used hospital admission for non-tuberculosis pneumonia, influenza, acute bronchitis, or upper respiratory tract infections as proxy outcomes in a dataset without any patients with covid-19.[16] Among the predictors were age, sex, previous hospital admission, comorbidities, 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

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

Table 1 | Characteristics of reviewed prediction models for diagnosis and prognosis of coronavirus disease 2019 (covid-19)

	No (%) of models* or median (interquartile range)
Country†	
Single country data	174 (75)
China	97 (42)
Italy	23 (10)
United States	17 (7)
South Korea	10 (4)
France	5 (2)
Singapore	4 (2)
Turkey	4 (2)
Brazil	3 (1)
Spain	2 (1)
United Kingdom	2 (1)
Other single country	8 (3)
International (combined) data	42 (18)
Unknown origin of data	16 (7)
Type of data used	
Proxy (non-covid-19) data	12 (5)
Simulated data	2 (1)
Target setting	
Patients admitted to hospital	119 (51)
Patient at triage centre or fever clinic	12 (5)
Patients in general practice	3 (1)
Other	23 (10)
Unclear	75 (32)
Target population	
Confirmed covid-19	108 (47)
Suspected covid-19	84 (36)
Other	13 (6)
Unclear	27 (12)
Type of model	
Predict risks of covid-19 in the general population	7 (3)
Diagnostic (covid-19 *v* not covid-19)	33 (14)
Diagnostic classification of covid-19 severity	10 (4)
Diagnostic, imaging data only	75 (32)
Prognostic	107 (46)
Study type	
Developed in reviewed study	50 (22)
Developed and internally validated in reviewed study	112 (48)
Developed and externally validated in reviewed study	46 (20)
Externally validated in reviewed study	24 (10)
Sample size	
Sample size (development)	338 (134-707)
No of events (development)	69 (37-160)
Sample size (external validation)	189 (76-312)
No of events (external validation)	40 (24-122)

*Analysis unit is a model within a study. Some studies investigated multiple models and some models were investigated in multiple studies (that is, in external validation studies).
†A study that uses development data from one country and validation data from another is classified as international.

videos from the faces of people wearing facemasks to determine abnormal breathing (not covid related) with a reported sensitivity of 80%.[92] A fifth model used demographics, symptoms, and contact history in a mobile app to assist general practitioners in collecting data and to risk-stratify patients. It was contrasted with two further models that included additional blood values and blood values plus computed tomography (CT) images. The authors reported a C index of 0.71 with demographics only, which rose to 0.97 and 0.99 as blood values and imaging characteristics were added.[151] Calibration was not assessed in any of the general population models.

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

We identified 33 multivariable models to distinguish between patients with and without covid-19. Most models targeted patients with suspected covid-19. Reported C index values ranged between 0.65 and 0.99. Calibration was assessed for seven models using calibration plots (including two at external validation), with mixed results. The most frequently included predictors (≥10 times) were vital signs (eg, temperature, heart rate, respiratory rate, oxygen saturation, blood pressure), flu-like signs and symptoms (eg, shiver, fatigue), age, electrolytes, image features (eg, pneumonia signs on CT scan), contact with individuals with confirmed covid-19, lymphocyte count, neutrophil count, cough or sputum, sex, leukocytes, liver enzymes, and red cell distribution width.

Ten studies aimed to diagnose severe disease in patients with covid-19: nine in adults with reported C indices between value of 0.80 and 0.99, and one in children that reported perfect classification of severe disease.[27] Calibration was not assessed in any of the models. Predictors of severe covid-19 used more than once were comorbidities, liver enzymes, C reactive protein, imaging features, lymphocyte count, and neutrophil count.

Seventy five 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 CT images or chest radiographs. Others used spectrograms of cough sounds[55] and lung ultrasound.[75] The predictive performance varied considerably, with reported C index values ranging from 0.70 to more than 0.99. Only one model based on imaging was evaluated by use of a calibration plot, and it appeared to be well calibrated at external validation.[186]

Prognostic models for patients with diagnosis of covid-19

We identified 107 prognostic models 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 37 days, but were often unspecified.

Of these models, 39 estimated mortality risk and 28 aimed to predict progression to a severe or critical disease. 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, acute respiratory distress syndrome, cardiac injury and thrombotic complication. One study used data from 2015 to 2019 to predict mortality and prolonged assisted mechanical ventilation (as a non-covid-19 proxy outcome).[115] The most frequently used categories of prognostic factors (for any outcome, included at least 20 times) included age, comorbidities, vital signs, image features, sex, lymphocyte count, and C reactive protein.

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

Studies that predicted mortality reported C indices between 0.68 and 0.98. Four studies also presented calibration plots (including at external validation for three models), all indicating miscalibration[15 69 118] or showing plots for integer scores without clearly explaining how these were translated into predicted risks.[143] The studies that developed models to predict progression to a severe or critical disease reported C indices between 0.58 and 0.99. Five of these models also were evaluated by calibration plots, two of them at external validation. Even though calibration appeared good, plots were constructed in an unclear way.[85 121] Reported C indices for other outcomes varied between 0.54 (admission to intensive care) and 0.99 (severe symptoms three days after admission), and five models had calibration plots (of which three at external validation), with mixed results.

Risk of bias

All models were at high (n=226, 97%) or unclear (n=6, 3%) risk of bias according to assessment with PROBAST, which suggests that their predictive performance when used in practice is probably lower than that reported (fig 2). Therefore, we have cause for concern that the predictions of the proposed models are unreliable when used in other people. Figure 2 and box 2 gives details on common causes for risk of bias for each type of model.

Ninety eight models (42%) had a high risk of bias for the participants domain, 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 led to an unclear risk of bias assessment in 58 models (25%), and 76 (33%) had a low risk of bias for the participants domain. Fifteen models (6%) 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 hundred and thirty five (58%) models were rated unclear and 82 (35%) rated at low risk of bias for the predictor domain. Most studies used outcomes that are easy to assess (eg, death, presence of covid-19 by laboratory confirmation), and hence 95 (41%) were rated at low risk of bias. Nonetheless, there was cause for concern about bias induced by the outcome measurement in 50 models (22%), for example, due to the use of subjective or proxy outcomes (eg, non-covid-19 severe respiratory infections). Eighty seven models (38%) had an unclear risk of bias due to opaque or ambiguous reporting. Two hundred and eighteen (94%) models were at high risk of bias for the analysis domain. The reporting was insufficiently clear to assess risk of bias in the analysis in 13 studies (6%). Only one model had a low risk of bias for the analysis domain (<1%). Twenty nine (13%) models had low risk of bias on all domains except analysis, indicating

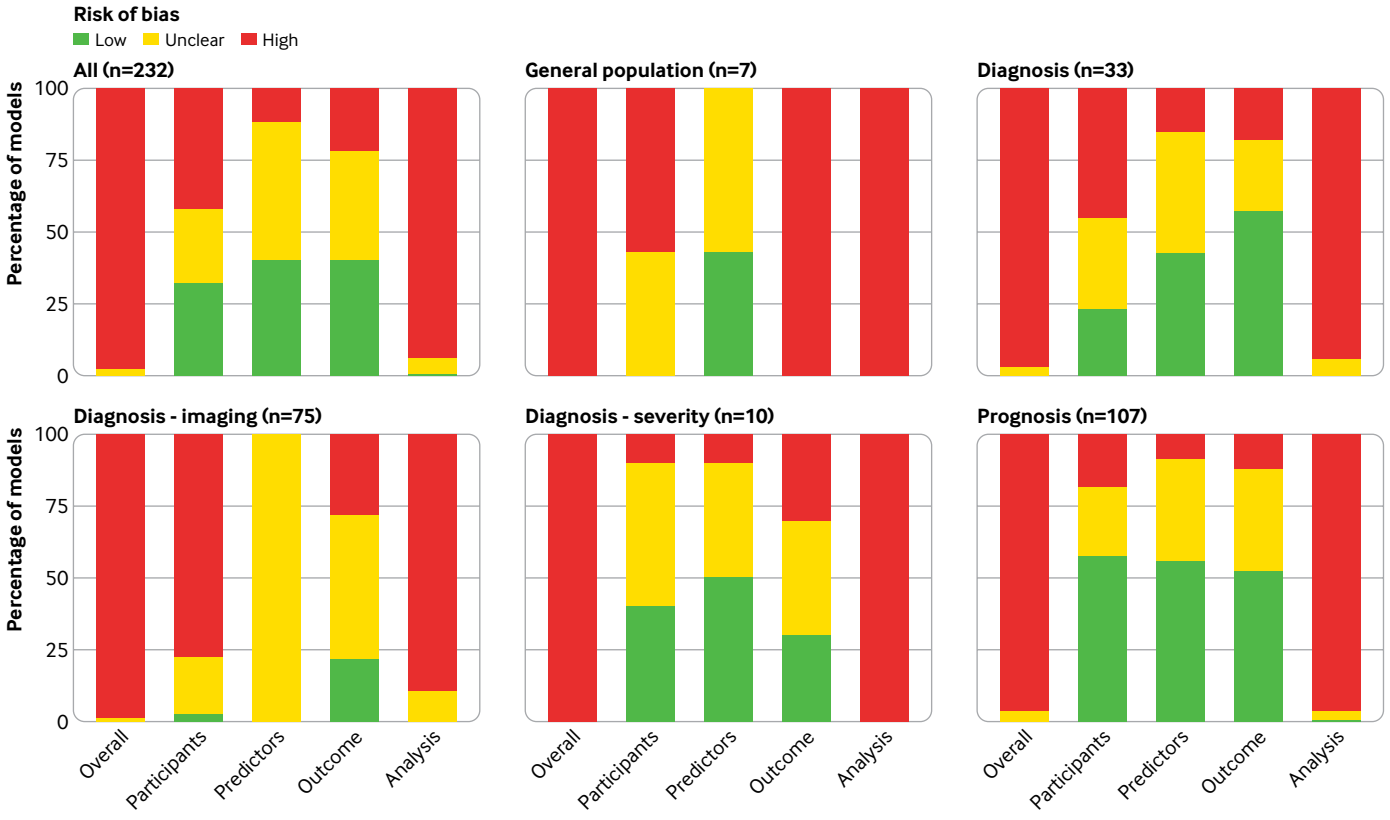


Fig 2 | PROBAST (prediction model risk of bias assessment tool) risk of bias for all included models combined (n=232) and broken down per type of model

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

adequate data collection and study design, but issues that could have been avoided by conducting a better statistical analysis. Many studies had small to modest sample sizes (table 1), which led to an increased risk of overfitting, particularly if complex modelling strategies were used. In addition, 50 models (22%) were neither internally nor externally validated. Performance statistics calculated on the development data from these models are likely optimistic. Calibration was only assessed for 22 models using calibration plots (10%), of which 11 on external validation data.

We found two models that were generally of good quality, built on large datasets, and had been rated low risk of bias on most domains but with an overall rating of unclear risk of bias, owing to unclear details on one signalling question within the analysis domain (table 2 provides a summary). Jehi and colleagues presented findings from developing a diagnostic model, however, there was substantial missing data and it remains unclear whether the use of median imputation influenced results, and there are unexplained discrepancies between the online calculator, nomogram, and published logistic regression model.[141] Hence, the calculator should not be used without further validation. Knight and colleagues developed a prognostic model for in-hospital mortality, however, continuous predictors were dichotomised, which reduces granularity of predicted risks (even though the model had a C index comparable with that of a generalised additive model).[143] The model was also converted into an sum score, but it was unclear how the scores were translated to the predicted mortality risks that were used to evaluate calibration.

External validation

Forty six models were developed and externally validated in the same study (in an independent dataset, excluding random training test splits and temporal splits). In addition, 24 external validations of models were developed for covid-19 or before the covid-19 pandemic in separate studies. However, none of the external validations was scored as low risk of bias, three were rated as unclear risk of bias, and 67 were rated as high risk of bias. One common concern is that datasets used for the external validation were likely not representative of the target population (eg, patients not being recruited consecutively, use of an inappropriate study design, use of unrepresentative controls, exclusion of patients still in follow-up). Consequently, predictive performance could differ if the models are applied in the targeted population. Moreover, only 15 (21%) external validations had

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

All of these models had unclear or high risk of bias for the participant, outcome, and analysis domain. All 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.[16 92 151]

Diagnostic models

Ten models (30%) used inappropriate data sources (eg, due to a non-nested case-control design), nine (27%) used inappropriate inclusion or exclusion criteria such that the study data was not representative of the target population, and eight (24%) selected controls that were not representative of the target population for a diagnostic model (eg, controls for a screening model had viral pneumonia). Other frequent problems were dichotomisation of predictors (nine models, 27%), and tests used to determine the outcome (eight models, 24%) or predictor definitions or measurement procedures (seven models, 21%) that varied between participants.

Diagnostic models based for severity classification

Two models (20%) used predictor data that was assessed while the severity (the outcome) was known. Other concerns include non-standard or lack of a prespecified outcome definition (two models, 20%), predictor measurements (eg, fever) being part of the outcome definition (two models, 20%) and outcomes being assessed with knowledge of predictor measurements (two models, 20%).

Diagnostic models based on medical imaging

Generally, studies did not clearly report which patients had imaging during clinical routine. Fifty five (73%) used an inappropriate or unclear study design to collect data (eg, a non-nested case-control). It was often unclear (39 models, 52%) whether the selection of controls was made from the target population (that is, patients with suspected covid-19). Outcome definitions were often not defined or determined in the same way in all participants (18 models, 24%). Diagnostic model studies that used medical images as predictors were all scored as unclear on the predictor domain. These 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. However, a more favourable assessment of the predictor domain does not lead to better overall judgment regarding risk of bias for the included models. Careful description of model specification and subsequent estimation were frequently 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

Dichotomisation of predictors was a frequent concern (22 models, 21%). Other problems include inappropriate inclusions or exclusions of study participants (18 models, 17%). 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 selected study sample (12 models, 11%). Additionally, many models (16 models, 15%) did not account for censoring or competing risks.

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

100 or more events, which is the recommended minumum.[187 188] Only 11 (16%) external validations presented a calibration plot.

Table 3 shows the results of external validations that had at most an unclear risk of bias and at least 100 events in the external validation set. The model by Jehi et al has been discussed above.[141] Luo and colleagues performed a validation of the CURB-65 score, originally developed to predict mortality of community acquired pneumonia, to assess its abilty to predict in-hospital mortality in patients with confirmed covid-19. This validation was conducted in a large retrospective cohort of patients admitted to two Chinese designated hospitals to treat patients with pneumonia from SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2).[155] It was unclear whether all consecutive patients were included (although this is likely given the retrospective design), no calibration plot was used because the score gives an integer as output rather than estimates risks, and the score uses dichotomised predictors. Overall, the external validation by Luo et al was performed well. Studies that validated CURB-65 in patients with covid-19 obtained C indexes of 0.58, 0.74, 0.75, 0.84, and 0.88.[130 148 155 164 189] These observed differences might be due to differences in risk of bias (all except Luo et al were rated high risk of bias), heterogeneity in study populations (South Korea, China, Turkey, and the United States), outcome definitions (progression to severe covid-19 *v* mortality), and sampling variability (number of events were 36, 55, 131, 201, and unclear).

Discussion

In this systematic review of prediction models related to the covid-19 pandemic, we identified and critically appraised 232 models described in 169 studies. 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 or uncertain 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 a single country. Often, the available sample sizes and number of events for the outcomes of interest were limited. This problem is well known when building prediction models and increases the risk of overfitting the model.[190] 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 specifically for covid-19, but only 46 carried out an external validation, and calibration was rarely assessed. We cannot yet recommend any of the identified prediction models for widespread use in clinical practice, although a few diagnostic and prognostic models originated from studies that were clearly of better quality. We suggest that these models should be further validated in other data sets, and ideally by independent investigators.[141 143]

Challenges and opportunities

The main aim of prediction models is to support medical decision making in individual patients. Therefore, it is vital to identify a target setting in which predictions serve a clinical need (eg, emergency department, intensive care unit, general practice, symptom monitoring app in the general population), and a representative dataset from that setting (preferably comprising consecutive patients) on which the prediction model can be developed and validated. This clinical setting and patient characteristics should be described in detail (including timing within the disease course, the severity of disease at the moment of prediction, and the comorbidity), so that readers and clinicians are able to understand if the proposed model could be suited for their population. Unfortunately, the studies included in our systematic review often lacked an adequate description of the target setting and study population, which leaves users of these models in doubt about the models' applicability. Although we recognise that the earlier studies were done under severe time constraints, we recommend that any studies currently in preprint and all future studies

Table 2 | Prediction models with unclear risk of bias overall and large development samples

Study; setting; and outcome	Model	Sample size (total No of participants (No with outcome))*	Predictive performance		Overall risk of bias using PROBAST
			Strongest type of validation reported	Performance†	
Diagnostic models					
Jehi et al[141]; data from US, patients with suspected covid-19; covid-19 diagnosis	Jehi model	Development 11 672 (818); external validation 2295 (290)	External validation, same country, new centres, and later period	C index 0.84 (95% CI 0.82 to 0.86)	Unclear
Prognostic models					
Knight et al[143]; data from UK, suspected or confirmed symptomatic inpatients; in-hospital mortality	4C Mortality Score	Development 35 463 (11 426); temporal validation 22 361 (6729)	Temporal validation	C index 0.77 (95% CI 0.76 to 0.77)	Unclear

PROBAST=prediction model risk of bias assessment tool; covid-19=coronavirus disease 2019.
*According to PROBAST, a large dataset is at least 10 events per candidate variable (EPV) for model development and at least 100 events for validation. If EPV could not be extracted or calculated from the study report, 100 events for model development was the lower limit to be included in this table.
†Performance from strongest type of validation reported.

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

Table 3 | External validations with unclear risk of bias and large validation samples

Study; setting; and outcome	Model	Sample size (total No of participants for model validation set (No with outcome))*	Predictive performance		Overall risk of bias using PROBAST
			Type of validation	Performance	
Diagnostic models					
Jehi et al[141]; data from US, patients with suspected covid-19; covid-19 diagnosis	Jehi model	Development 11 672 (818); external validation 2295 (290)	External validation, same country, new centres and later period	C index 0.84 (95% CI 0.82 to 0.86)	Unclear
Prognostic models					
Luo et al[155]; data from China, in-patients with confirmed covid-19; in-hospital mortality	CURB-65	1018 (201)	Independent external validation	C index 0.84 (95% CI 0.82 to 0 .93)	Unclear

PROBAST=prediction model risk of bias assessment tool; CURB-65=confusion, urea, respiratory rate, blood pressure plus age of at least 65 years.
*According to PROBAST, a large dataset is at least 10 events per candidate variable for model development and at least 100 events for validation.

should adhere to the TRIPOD reporting guideline[12] to improve the description of their study population and guide 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 and covid-19 prevalence. The variability in mortality could be related to differences in included patients (eg, age, comorbidities) and interventions for covid-19. The variability in prevalence could in part be reflective of different diagnostic standards across studies.

Covid-19 prediction 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 some reviewed studies have done. Instead, an appropriate time to event analysis should be considered to allow for administrative censoring.[13] 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.[191]

We reviewed 75 studies that used only medical images to diagnose covid-19, covid-19 related pneumonia, or to assist in segmentation of lung images, the majority using advanced machine learning methodology. 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. Currently, none of these models is recommended to be used in clinical practice. An independent systematic review and critical appraisal (using PROBAST[12]) of machine learning models for covid-19 using chest radiographs and CT scans came to the same conclusions, even though they focused on models that met a minimum requirement of study quality based on specialised quality metrics for the assessment of radiomics and deep-learning based diagnostic models in radiology.[192]

A prediction model applied in a new healthcare setting or country often produces predictions that are miscalibrated[193] and might need to be updated before it can safely be applied in that new setting.[13] 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.[194-198]

The evidence base for the development and validation of prediction models related to covid-19 will continue to increase over the coming months. 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, 80 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 currently being used in clinical practice without scientific publications,[199] and web risk calculators launched for use while the scientific manuscript is still under review (and unavailable on request).[200] These unpublished models naturally fall outside the scope of this review of the literature. As we have argued extensively elsewhere,[201] 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 reviewed prediction models were found to have an unclear or high risk of bias, and evidence from independent external validations of the newly

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

developed models is still scarce. 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,[202] and they could even cause more harm than good.[203] 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. Moreover, predictive performance estimates obtained from different populations, settings, and types of validation (internal *v* external) are not directly comparable. Future studies should focus on validating, comparing, improving, and updating promising available prediction models.[13] The models by Knight and colleagues[143] and Jehi and colleagues[141] are good candidates for validation studies in other data. We advise Jehi and colleagues to make all model equations available for independent validation.[141] Such external validations should assess not only discrimination, but also calibration and clinical utility (net benefit),[193 198 203] in large datasets[187 188] collected using an appropriate study design. In addition, these models' transportability to other countries or settings remains to be investigated. Owing to differences between healthcare systems (eg, Chinese and European) and over time 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 researchers could attempt to update and adjust the model to the local setting.

Most reviewed models used data from a hospital setting, but few are available for primary care and the general population. Additional research is needed, including validation of any recently proposed models not yet included in the current update of the living review (eg, Clift et al[204]). The models reviewed to date predicted the covid-19 diagnosis or assess the risk of mortality or deterioration, whereas long term morbidity and functional outcomes remain understudied and could be a target outcome of interest in future studies developing prediction models.[205 206]

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.[13] This is especially important for datasets with limited sample size.[207] Frequently used predictors included in multiple models identified by our review are vital signs, age, comorbidities, and image features, and these should be considered when appropriate. Flu-like symptoms should be considered in diagnostic models, and sex, C reactive protein, and lymphocyte counts could be considered as 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, which should preferably 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.

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 or unclear risk of bias, mainly because of model overfitting, inappropriate model evaluation (eg, calibration ignored), use of inappropriate data sources and unclear reporting. Therefore, their performance estimates are probably optimistic and not representative for the target population. The COVID-PRECISE group does not recommend any of the current prediction models to be used in practice, but one diagnostic and one prognostic model originated from higher quality studies and should be (independently) validated in other datasets. For details of the reviewed models, see https://www.covprecise.org/. Future studies aimed at developing and validating diagnostic or prognostic models for covid-19 should explicitly describe the concerns raised and follow existing methodological guidance for prediction modeling studies, because unreliable predictions could cause more harm than benefit in guiding clinical decisions. Prediction model authors should adhere to the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) reporting guideline. Finally, 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

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

[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]Department of Nephrology, Medical University of Vienna, Vienna, Austria
[20]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
[21]Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, Netherlands
[22]Division of Informatics, Imaging and Data Science, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK
[23]Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK
[24]Centre for Biostatistics, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
[25]Centre for Epidemiology Versus Arthritis, Centre for Musculoskeletal Research, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
[26]Division of Nursing, Midwifery and Social Work, School of Health Sciences, University of Manchester, Manchester, UK
[27]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK
[28]Amsterdam UMC, University of Amsterdam, Amsterdam Public Health, Medical Library, Netherlands
[29]Department of Epidemiology and Biostatistics, Imperial College London School of Public Health, London, UK
[30]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece
[31]Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre+, Maastricht, Netherlands
[32]Department of Intensive Care, Maastricht University Medical Centre+, Maastricht University, Maastricht, Netherlands
[33]EPI-Centre, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium
[34]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
[35]Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany
[36]Berlin Institute of Health, Berlin, Germany
[37]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, MK, ICCvdH, BCTvB, DJM, 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) and receives support from ZonMw (grant 10430012010001). 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). ICCvDH and BCTvB received funding from Euregio Meuse-Rhine (grant Covid Data Platform (coDaP) interref EMR-187). 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/. Detailed extracted data on all included studies are available on https://www.covprecise.org/.

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/.

Provenance and peer review: Not commissioned; externally peer reviewed.

This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/.

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 Looi M-K. Covid-19: Is a second wave hitting Europe?*BMJ* 2020;371:m4113. doi:10.1136/bmj.m4113
6 Woolf SH, Chapman DA, Lee JH. COVID-19 as the Leading Cause of Death in the United States. *JAMA* 2021;325:123-4.
7 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.

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

8 Institute of Social and Preventive Medicine. Living evidence on COVID-19 2020. https://ispmbern.github.io/covid-19/living-review/index.html.
9 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.
10 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
11 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
12 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
13 Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating*. Springer US, 2019. doi:10.1007/978-3-030-16399-0
14 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
15 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
16 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.
17 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
18 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
19 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
20 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
21 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
22 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
23 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
24 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
25 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
26 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
27 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
28 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
29 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
30 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
31 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
32 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
33 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
34 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
35 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
36 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
37 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
38 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
39 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
40 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
41 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.
42 Sun Y, Koh V, Marimuthu K, et al. Epidemiological and clinical predictors of covid-19. *Clin Infect Dis* 2020;ciaa322.
43 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
44 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
45 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
46 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
47 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
48 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
49 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
50 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
51 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
52 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.
53 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
54 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.
55 Imran A, Posokhova I, Qureshi HN, et al. AI4COVID-19: AI enabled preliminary diagnosis for covid-19 from cough samples via an

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

app. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200401275I.
56 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
57 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.
58 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
59 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.
60 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.
61 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
62 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
63 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
64 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
65 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
66 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
67 Arpan M, Surya K, Harish R, et al. CovidAID: COVID-19 Detection Using Chest X-Ray. ArXiv e-prints [Preprint] 2020
68 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.
69 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
70 Bassi PRAS, Attux R. A deep convolutional neural network for covid-19 detection using chest x-rays. ArXiv e-prints [Preprint] 2020
71 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
72 Bello-Chavolla OY, Bahena-López JP, Antonio-Villa NE, et al. Predicting mortality due to SARS-CoV-2: A mechanistic score relating obesity and diabetes to COVID-19 outcomes in Mexico. *J Clin Endocrinol Metab* 2020;105:dgaa346. doi:10.1210/clinem/dgaa346
73 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
74 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
75 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.
76 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
77 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
78 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
79 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
80 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
81 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
82 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
83 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:10.1101/2020.04.05.20047944
84 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
85 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
86 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
87 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
88 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
89 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
90 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.
91 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
92 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.
93 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
94 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.
95 Khan AI, Shah JL, Bhat MM. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Comput Methods Programs Biomed* 2020;196:105581. doi:10.1016/j.cmpb.2020.105581
96 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
97 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
98 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
99 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
100 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
101 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.
102 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. *Lab Chip* 2020;20:2075-85. doi:10.1039/D0LC00373E
103 Mei X, Lee HC, Diao KY, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nat Med* 2020;26:1224-8. doi:10.1038/s41591-020-0931-3

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

104 Menni C, Valdes AM, Freidin MB, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020;26:1037-40. doi:10.1038/s41591-020-0916-2
105 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
106 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
107 Rahimzadeh M, Attar A. A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. *Inform Med Unlocked* 2020;19:100360. doi:10.1016/j.imu.2020.100360
108 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
109 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
110 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
111 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:10.1101/2020.04.10.20061036
112 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
113 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
114 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
115 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
116 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
117 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.
118 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
119 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
120 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
121 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.
122 Apostolopoulos ID, Aznaouridis SI, Tzani MA. Extracting possibly representative covid-19 biomarkers from x-ray images with deep learning approach and image data related to pulmonary diseases. *J Med Biol Eng* 2020;01:1-8. doi:10.1007/s40846-020-00529-4
123 Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks. *Comput Biol Med* 2020;121:103795. doi:10.1016/j.compbiomed.2020.103795
124 Bar S, Lecourtois A, Diouf M, et al. The association of lung ultrasound images with COVID-19 infection in an emergency room cohort. *Anaesthesia* 2020;75:1620-5. doi:10.1111/anae.15175
125 Bi X, Su Z, Yan H, et al. Prediction of severe illness due to covid-19 based on an analysis of initial fibrinogen to albumin ratio and platelet count. *Platelets* 2020;31:674-9. doi:10.1080/09537104.2020.1760230
126 Borghesi A, Zigliani A, Golemi S, et al. Chest x-ray severity index as a predictor of in-hospital mortality in coronavirus disease 2019: a study of 302 patients from Italy. *Int J Infect Dis* 2020;96:291-3. doi:10.1016/j.ijid.2020.05.021
127 Burian E, Jungmann F, Kaissis GA, et al. Intensive care risk estimation in covid-19 pneumonia based on clinical and imaging parameters: experiences from the Munich cohort. *J Clin Med* 2020;9:E1514. doi:10.3390/jcm9051514
128 Cecconi M, Piovani D, Brunetta E, et al. Early predictors of clinical deterioration in a cohort of 239 patients hospitalized for covid-19 infection in Lombardy, Italy. *J Clin Med* 2020;9:E1548. doi:10.3390/jcm9051548
129 Cheng FY, Joshi H, Tandon P, et al. Using machine learning to predict ICU transfer in hospitalized covid-19 patients. *J Clin Med* 2020;9:E1668. doi:10.3390/jcm9061668
130 Choi MH, Ahn H, Ryu HS, et al. Clinical characteristics and disease progression in early-stage covid-19 patients in South Korea. *J Clin Med* 2020;9:E1959. doi:10.3390/jcm9061959
131 Clemency BM, Varughese R, Scheafer DK, et al. Symptom criteria for covid-19 testing of heath care workers. *Acad Emerg Med* 2020;27:469-74. doi:10.1111/acem.14009
132 Dong Y, Zhou H, Li M, et al. A novel simple scoring model for predicting severity of patients with SARS-CoV-2 infection. *Transbound Emerg Dis* 2020;67:2823-9. doi:10.1111/tbed.13651
133 El Asnaoui K, Chawki Y. Using X-ray images and deep learning for automated detection of coronavirus disease. *J Biomol Struct Dyn* 2020;1-12. doi:10.1080/07391102.2020.1767212
134 Fu L, Li Y, Cheng A, Pang P, Shu Z. A novel machine learning-derived radiomic signature of the whole lung differentiates stable from progressive covid-19 infection: a retrospective cohort study. *J Thorac Imaging* 2020. doi:10.1097/RTI.0000000000000544
135 Galloway JB, Norton S, Barker RD, et al. A clinical risk score to identify patients with COVID-19 at high risk of critical care admission or death: An observational cohort study. *J Infect* 2020;81:282-8. doi:10.1016/j.jinf.2020.05.064
136 Gezer NS, Ergan B, Barış MM, et al. COVID-19 S: A new proposal for diagnosis and structured reporting of COVID-19 on computed tomography imaging. *Diagn Interv Radiol* 2020;26:315-22. doi:10.5152/dir.2020.20351
137 Gidari A, De Socio GV, Sabbatini S, Francisci D. Predictive value of National Early Warning Score 2 (NEWS2) for intensive care unit admission in patients with SARS-CoV-2 infection. *Infect Dis (Lond)* 2020;52:698-704. doi:10.1080/23744235.2020.1784457
138 Hong Y, Wu X, Qu J, Gao Y, Chen H, Zhang Z. Clinical characteristics of Coronavirus Disease 2019 and development of a prediction model for prolonged hospital length of stay. *Ann Transl Med* 2020;8:443. doi:10.21037/atm.2020.03.147
139 Huang D, Wang T, Chen Z, Yang H, Yao R, Liang Z. A novel risk score to predict diagnosis with coronavirus disease 2019 (COVID-19) in suspected patients: a retrospective, multicenter, and observational study. *J Med Virol* 2020;92:2709-17. doi:10.1002/jmv.26143
140 Huang J, Cheng A, Lin S, Zhu Y, Chen G. Individualized prediction nomograms for disease progression in mild COVID-19. *J Med Virol* 2020;92:2074-80. doi:10.1002/jmv.25969
141 Jehi L, Ji X, Milinovich A, et al. Individualizing risk prediction for positive coronavirus disease 2019 testing: results from 11,672 patients. *Chest* 2020;158:1364-75. doi:10.1016/j.chest.2020.05.580
142 Joshi RP, Pejaver V, Hammarlund NE, et al. A predictive tool for identification of SARS-CoV-2 PCR-negative emergency department patients using routine test results. *J Clin Virol* 2020;129:104502. doi:10.1016/j.jcv.2020.104502
143 Knight SR, Ho A, Pius R, et al, ISARIC4C investigators. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. *BMJ* 2020;370:m3339. doi:10.1136/bmj.m3339
144 Ko H, Chung H, Kang WS, et al. COVID-19 pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT image: model development and validation. *J Med Internet Res* 2020;22:e19569. doi:10.2196/19569
145 Li Q, Zhang J, Ling Y, et al. A simple algorithm helps early identification of SARS-CoV-2 infection patients with severe progression tendency. *Infection* 2020;48:577-84. doi:10.1007/s15010-020-01446-z
146 Li Y, Yang Z, Ai T, Wu S, Xia L. Association of "initial CT" findings with mortality in older patients with coronavirus disease 2019 (COVID-19). *Eur Radiol* 2020;30:6186-93. doi:10.1007/s00330-020-06969-5
147 Li Z, Zeng B, Lei P, et al. Differentiating pneumonia with and without COVID-19 using chest CT images: from qualitative to quantitative. *J Xray Sci Technol* 2020;28:583-9. doi:10.3233/XST-200689
148 Liang W, Liang H, Ou L, et al, China Medical Treatment Expert Group for COVID-19. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. *JAMA Intern Med* 2020;180:1081-9. doi:10.1001/jamainternmed.2020.2033
149 Liu F, Zhang Q, Huang C, et al. CT quantification of pneumonia lesions in early days predicts progression to severe illness in a cohort of

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

COVID-19 patients. *Theranostics* 2020;10:5613-22. doi:10.7150/thno.45985
150 Liu X, Shi S, Xiao J, et al. Prediction of the severity of coronavirus disease 2019 and its adverse clinical outcomes. *Jpn J Infect Dis* 2020;73:404-10. doi:10.7883/yoken.JJID.2020.194
151 Liu Y, Wang Z, Ren J, et al. A covid-19 risk assessment decision support system for general practitioners: design and development study. *J Med Internet Res* 2020;22:e19786. doi:10.2196/19786.
152 Liu YP, Li GM, He J, et al. Combined use of the neutrophil-to-lymphocyte ratio and CRP to predict 7-day disease severity in 84 hospitalized patients with COVID-19 pneumonia: a retrospective cohort study. *Ann Transl Med* 2020;8:635. doi:10.21037/atm-20-2372
153 Lorente-Ros A, Monteagudo Ruiz JM, Rincón LM, et al. Myocardial injury determination improves risk stratification and predicts mortality in COVID-19 patients. *Cardiol J* 2020;27:489-96.
154 Luo L, Luo Z, Jia Y, et al. CT differential diagnosis of COVID-19 and non-COVID-19 in symptomatic suspects: a practical scoring method. *BMC Pulm Med* 2020;20:129. doi:10.1186/s12890-020-1170-6
155 Luo M, Liu J, Jiang W, Yue S, Liu H, Wei S. IL-6 and CD8+ T cell counts combined are an early predictor of in-hospital mortality of patients with COVID-19. *JCI Insight* 2020;5:139024. doi:10.1172/jci.insight.139024
156 Luo Y, Yuan X, Xue Y, et al. Using a diagnostic model based on routine laboratory tests to distinguish patients infected with SARS-CoV-2 from those infected with influenza virus. *Int J Infect Dis* 2020;95:436-40. doi:10.1016/j.ijid.2020.04.078
157 Matos J, Paparo F, Mussetto I, et al. Evaluation of novel coronavirus disease (COVID-19) using quantitative lung CT and clinical data: prediction of short-term outcome. *Eur Radiol Exp* 2020;4:39. doi:10.1186/s41747-020-00167-0
158 Mazzaccaro D, Giacomazzi F, Giannetta M, et al. Non-overt coagulopathy in non-ICU patients with mild to moderate covid-19 pneumonia. *J Clin Med* 2020;9:E1781. doi:10.3390/jcm9061781
159 Murphy K, Smits H, Knoops AJG, et al. Covid-19 on the chest radiograph: a multireader evaluation of an artificial intelligence system. *Radiology* 2020;296:E166-72. doi:10.1148/radiol.2020201874
160 Obeid JS, Davis M, Turner M, et al. An artificial intelligence approach to COVID-19 infection risk assessment in virtual visits: A case report. *J Am Med Inform Assoc* 2020;27:1321-5. doi:10.1093/jamia/ocaa105
161 Pu J, Leader J, Bandos A, et al. Any unique image biomarkers associated with COVID-19?*Eur Radiol* 2020;30:6221-7. doi:10.1007/s00330-020-06956-w
162 Rajaraman S, Antani S. Weakly labeled data augmentation for deep learning: a study on covid-19 detection in chest x-rays. *Diagnostics (Basel)* 2020;10:E358. doi:10.3390/diagnostics10060358
163 Roland LT, Gurrola JG2nd, Loftus PA, Cheung SW, Chang JL. Smell and taste symptom-based predictive model for COVID-19 diagnosis. *Int Forum Allergy Rhinol* 2020;10:832-8. doi:10.1002/alr.22602
164 Satici C, Demirkol MA, Sargin Altunok E, et al. Performance of pneumonia severity index and CURB-65 in predicting 30-day mortality in patients with COVID-19. *Int J Infect Dis* 2020;98:84-9. doi:10.1016/j.ijid.2020.06.038
165 Song J, Wang H, Liu Y, et al. End-to-end automatic differentiation of the coronavirus disease 2019 (COVID-19) from viral pneumonia based on chest CT. *Eur J Nucl Med Mol Imaging* 2020;47:2516-24. doi:10.1007/s00259-020-04929-1
166 Sun L, Song F, Shi N, et al. Combination of four clinical indicators predicts the severe/critical symptom of patients infected COVID-19. *J Clin Virol* 2020;128:104431. doi:10.1016/j.jcv.2020.104431
167 Toraih EA, Elshazli RM, Hussein MH, et al. Association of cardiac biomarkers and comorbidities with increased mortality, severity, and cardiac injury in COVID-19 patients: A meta-regression and decision tree analysis. *J Med Virol* 2020;92:2473-88. doi:10.1002/jmv.26166
168 Tuncer T, Dogan S, Ozyurt F. An automated residual exemplar local binary pattern and iterative relieff based covid-19 detection method using chest x-ray image. *Chemometr Intell Lab Syst* 2020;203:104054. doi:10.1016/j.chemolab.2020.104054
169 Vaid S, Kalantar R, Bhandari M. Deep learning COVID-19 detection bias: accuracy through artificial intelligence. *Int Orthop* 2020;44:1539-42. doi:10.1007/s00264-020-04609-7
170 Vultaggio A, Vivarelli E, Virgili G, et al. Prompt predicting of early clinical deterioration of moderate-to-severe COVID-19 patients: usefulness of a combined score using IL-6 in a preliminary study. *J Allergy Clin Immunol Pract* 2020;8:2575-2581.e2. doi:10.1016/j.jaip.2020.06.013
171 Wang F, Hou H, Wang T, et al. Establishing a model for predicting the outcome of COVID-19 based on combination of laboratory tests. *Travel Med Infect Dis* 2020;36:101782. doi:10.1016/j.tmaid.2020.101782
172 Wang K, Zuo P, Liu Y, et al. Clinical and laboratory predictors of in-hospital mortality in patients with coronavirus disease-2019: a cohort study in Wuhan, China. *Clin Infect Dis* 2020;71:2079-88. doi:10.1093/cid/ciaa538
173 Wang L, Liu Y, Zhang T, et al. Differentiating between 2019 novel coronavirus pneumonia and influenza using a nonspecific laboratory marker-based dynamic nomogram. *Open Forum Infect Dis* 2020;7:a169. doi:10.1093/ofid/ofaa169
174 Wu S, Du Z, Shen S, et al. Identification and validation of a novel clinical signature to predict the prognosis in confirmed COVID-19 patients. *Clin Infect Dis* 2020;ciaa793. doi:10.1093/cid/ciaa793
175 Wu X, Hui H, Niu M, et al. Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: a multicentre study. *Eur J Radiol* 2020;128:109041. doi:10.1016/j.ejrad.2020.109041
176 Yang P, Wang P, Song Y, Zhang A, Yuan G, Cui Y. A retrospective study on the epidemiological characteristics and establishment of an early warning system of severe COVID-19 patients. *J Med Virol* 2020;92:2173-80. doi:10.1002/jmv.26022
177 Yang Y, Shen C, Li J, et al. Plasma IP-10 and MCP-3 levels are highly associated with disease severity and predict the progression of COVID-19. *J Allergy Clin Immunol* 2020;146:119-127.e4. doi:10.1016/j.jaci.2020.04.027
178 Yu C, Lei Q, Li W, et al. Clinical characteristics, associated factors, and predicting covid-19 mortality risk: a retrospective study in Wuhan, China. *Am J Prev Med* 2020;59:168-75. doi:10.1016/j.amepre.2020.05.002
179 Zhang C, Qin L, Li K, et al. A novel scoring system for prediction of disease severity in covid-19. *Front Cell Infect Microbiol* 2020;10:318. doi:10.3389/fcimb.2020.00318
180 Zhang K, Liu X, Shen J, et al. Clinically applicable AI System for accurate diagnosis, quantitative measurements, and prognosis of covid-19 pneumonia using computed tomography. *Cell* 2020;181:1423-1433.e11. doi:10.1016/j.cell.2020.04.045
181 Zheng QN, Xu MY, Zheng YL, Wang XY, Zhao H. Prediction of the rehabilitation duration and risk management for mild-moderate COVID-19. *Disaster Med Public Health Prep* 2020;14:652-7. doi:10.1017/dmp.2020.214
182 Zhou Y, He Y, Yang H, et al. Development and validation a nomogram for predicting the risk of severe COVID-19: a multi-center study in Sichuan, China. *PLoS One* 2020;15:e0233328. doi:10.1371/journal.pone.0233328
183 Zou X, Li S, Fang M, et al. Acute physiology and chronic health evaluation II score as a predictor of hospital mortality in patients of coronavirus disease 2019. *Crit Care Med* 2020;48:e657-65. doi:10.1097/CCM.0000000000004411
184 Cohen JP, Morrison P, Dao L. COVID-19 image data collection. arXiv e-prints [Preprint] 2020. https://github.com/ieee8023/covid-chestxray-dataset.
185 Kaggle. COVID-19 Kaggle community contributions 2020. https://www.kaggle.com/covid-19-contributions.
186 Wang S, Zha Y, Li W, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J* 2020;56:2000775. doi:10.1183/13993003.00775-2020
187 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
188 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
189 Levy TJ, Richardson S, Coppa K, et al. Estimating survival of hospitalized covid-19 patients from admission information. *MedRxiv* 2020. doi:10.1101/2020.04.22.20075416
190 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
191 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
192 Roberts M, Driggs D, Thorpe M, et al. Machine learning for covid-19 detection and prognostication using chest radiographs and CT scans: a systematic methodological review. arXiv 2020:200806388.
193 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
194 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
195 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

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

196 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
197 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
198 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
199 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.
200 Surgisphere Corporation. COVID-19 response center 2020. https://surgisphere.com/covid-19-response-center/.
201 Van Calster B, Wynants L, Timmerman D, Steyerberg EW, Collins GS. 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
202 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
203 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
204 Clift AK, Coupland CAC, Keogh RH, et al. Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. *BMJ* 2020;371:m3731. doi:10.1136/bmj.m3731
205 Mahase E. Covid-19: What do we know about "long covid"?*BMJ* 2020;370:m2815. doi:10.1136/bmj.m2815
206 Klok FA, Boon GJAM, Barco S, et al. The post-covid-19 functional status scale: a tool to measure functional status over time after covid-19. *Eur Respir J* 2020;56:2001494. doi:10.1183/13993003.01494-2020
207 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

BMJ: first published as 10.1136/bmj.m1328 on 7 April 2020. Downloaded from http://www.bmj.com/ on 10 February 2021 by guest. Protected by copyright.

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe