Prognostic scores developed or validated for coronavirus disease 2019: Protocol for a rapid systematic review

Version 1.0, 1 May 2020

Mark G Pritchard,* Louise Sigfrid, Blake Thomson, John Amuasi, J Kenneth Baillie, Mike Christian, Annemarie Docherty, Natalie Elkheir, Yeetey Enuameh, Ewen Harrison, Antonia Ho, Karl Holden, Peter Horby, Stephen Knight, Laura Merson, Matthew Muller, Piero Olliaro, Clark D Russell, Calum Semple, Olivia Swann, Gail Carson;
On behalf of the International Severe Acute Respiratory and emerging Infection Consortium (ISARIC)

* mark.pritchard@ndm.ox.ac.uk

Introduction

Prognostic scores allow synthesis of clinical information to identify patients who are likely to experience a poor outcome from their disease. They can be used to identify individuals who will require a greater level of medical intervention. Application of prognostic scores can ensure severely unwell patients receive appropriate treatment, thereby reducing mortality.[1-4] Scores can also identify patients who can safely be discharged from hospital to be treated at home.[5] In addition to benefits for individual patients, accurate triage during a pandemic helps to maximize public health benefit from available resources.[6-8] Triage protocols may be based on severity scores,[9,10] for which it is necessary to identify reliable scoring systems. However, it is important to note that the performance of a triage protocol additionally depends on population level outcomes which are influenced by systems factors. Often thresholds score values used in a decision aid will have a significant impact on the performance of the decision aid overall.

In December 2019, a cluster of cases of pneumonia occurred in the city of Wuhan in the People's Republic of China.[11] This has been named Coronavirus disease 2019 (COVID-19), and the causative virus is severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). By 23 April 2020, 2.5 million people had been recorded as infected, of whom 175,000 have died.[12] There is an urgent need for prognostic scores that can be applied clinically when patients present with COVID-19. Wynants and colleagues[13] reviewed prognostic models to predict outcomes for patients with COVID-19. However, we are not aware of any systematic reviews of prognostic scores that can be applied in a clinical setting when patients present with COVID-19.

Prognostic scores can be applied at different points in a patient's journey. For example in the United Kingdom, CRB-65[14] is recommended for use in primary care to guide a decision whether to refer a patient with community-acquired pneumonia to hospital.[5] Scores that include laboratory results, such as CURB-65[14] or pneumonia severity index (PSI),[15] may be used once in hospital. On arrival in an intensive care unit (ICU), even more detailed scores, such as the acute physiology and chronic health evaluation (APACHE-II),[16] are commonly calculated.[17] Scores can also be designed to predict different outcomes: CURB-65, PSI and APACHE-II predict mortality,[14-16] CURXO-80[18] predicts either mortality or ICU admission, and SMART-COP[19] is designed to predict the need for intensive respiratory or vasopressor support.

The primary objective of this review is to evaluate the effectiveness of prognostic scores for COVID-19. This will include both novel prognostic scores developed specifically for COVID-19

patients, and the application of existing scores (for example pneumonia severity scores[1,14,15,18-20] or the quick sepsis-related organ failure assessment [qSOFA])[21] to patients with COVID-19. We will include scores that are relevant at the time of presentation to a healthcare service (such as described above), but not scores that are intended for frequent re-calculation throughout a patient's stay (such as the National Early Warning Score)[2,3]. We will include any clinically relevant outcome, including mortality, or increased healthcare requirements (such as prolonged hospital stay or ICU admission). Secondarily, we will evaluate prognostic scores that have been used for patients with severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS).

Methods

We will follow the guidance and CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) proposed by Moons and colleagues.[22,23] To meet the current clinical need, this is planned as a rapid review. We will therefore limit our search to a single database of published material, Ovid Medline, and one pre-print server, medRxiv, and we will use more restrictive search terms than would otherwise have been the case. A minimum of 100 to 200 events are required to reliably validate a prognostic score.[24,25] However, in this first review we will not exclude studies based on size as studies with smaller cohorts may demonstrate an initial result that needs to be replicated in a subsequent study.

Study selection

Inclusion criteria

- Studies that report on prognostic scores calculated using data available at the time of presentation to a healthcare setting or soon after, including validation studies of existing prognostic scores and development of novel scores.
- Studies including participants with clinical COVID-19, SARS or MERS, or laboratory-confirmed SARS-CoV-2, SARS-CoV or MERS-CoV.
- Studies including participants of any age.
- Studies will be included for first assessment in any healthcare setting, including assessment by paramedics or other pre-hospital healthcare providers, presentation to primary care, presentation to an emergency medicine department, and arrival at an intensive care unit
- Studies reporting associations between prognostic scores and clinically relevant outcomes. Accepted outcomes will include mortality, admission to an ICU, higher-dependency unit or requirement for invasive mechanical ventilation, or prolonged hospital stay. Accepted associations will include areas under receiver operating characteristic curves, odds ratios and hazard ratios, and positive- and negative-predictive values for score thresholds.
- Observational and interventional studies will be included.

Exclusion criteria

- Prognostic scores that consist entirely of radiological findings.
- Studies that use a single parameter as a marker of severity.
- Studies that report results of a regression or other model but do not provide a prognostic score that could be calculated in a clinical setting.
- Studies including people with illnesses other than COVID-19, SARS or MERS.
- Studies that only correlate scores with other parameters such as radiological findings.
- Studies in which score was being used to determine, rather than predict an outcome, for example if a severity score threshold were used to determine who to admit to an ICU.

If studies that otherwise do not meet the inclusion criteria report that clinical decisions were made on the basis of a severity score, this information will be recorded. There will be no restriction on language of publication.

Search strategies

We will search for relevant publications using Ovid MEDLINE (table 1) and medRxiv (Table 2).

Table 1. Search strategy for OVID MEDLINE.

1	"coronavirus disease 2019".ti,ab.
2	(COVID19 or "COVID 19" or "COVID 2019").ti,ab.
3	"severe acute respiratory syndrome coronavirus 2".ti,ab.
4	(SARSCOV2 or "SARS COV2" or "SARS COV 2").ti,ab.
5	"2019 nCov".ti,ab.
6	(novel adj3 coronavirus).ti,ab
7	(Wuhan and coronavirus).ti,ab.
8	Coronavirus/ or Coronavirus Infections/
9	(COVID or coronavirus or nCoV).ti,ab.
10	8 or 9
11	limit 10 to yr="2019 - Current"
12	1 or 2 or 3 or 4 or 5 or 6 or 7 or 11
13	"Severity of Illness Index"/
14	"disease sever*".ab,ti.
15	(sever* adj3 scor*).ti,ab.
16	(prognos* adj3 scor*).ti,ab.
17	(predic* adj3 scor*).ti,ab.
18	Early Warning Score/
19	"National Early Warning Score".ti,ab.
20	("sepsis related organ failure assessment" or "sequential organ failure assessment").ti,ab.
21	(SOFA or qSOFA).ti,ab.
22	(CRB or CURB).ti,ab.
23	("Pneumonia severity index" or PSI).ti,ab.
24	("Pandemic medical early warning score" or PMEWS).ti,ab.
25	A-DROP.ti,ab.
26	SMART-COP.ti,ab.
27	(SCAP or "CURXO 80" or CURXO80).ti,ab.
28	"Community Assessment Tool".ti,ab.
29	13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28
30*	12 and 29
31	SARS Virus/ or Severe Acute Respiratory Syndrome/
32	(SARS or "Severe Acute Respiratory Syndrome" or SARSCoV).ti,ab.
33	Middle East Respiratory Syndrome Coronavirus/
34	(MERS or "Middle East Respiratory Syndrome" or MERSCoV).ti,ab.
35	Coronavirus/ or Coronavirus Infections/
36	(coronavirus).ti,ab.
37	31 or 32 or 33 or 34 or 35 or 36 or 37
38	29 and 37
39*	38 not 30
* Results from search row 30 will be screened first to inform ongoing work with coronavirus disease 2019; results from search row 39 will be screened to answer the secondary objective.	

Additional publications will be identified from reference lists of included studies.

Google Scholar and location-specific search engines such as the World Health Organization's regional index medicus projects will be searched using equivalent search terms to identify grey literature, and publications that are not indexed in Ovid Medline.

Data management

Identified studies will be stored in an Endnote database.

Table 2. Search strategy for medRxiv.

Results identified from each search will be included in screening

abstract or title "COVID prognostic score" (match all words)
abstract or title "coronavirus prognostic score" (match all words) and posted between "01 Dec, 2019 and 28 Apr, 2020"
abstract or title "SARS-CoV2 prognostic score" (match all words)
abstract or title "COVID severity score" (match all words)
abstract or title "coronavirus severity score" (match all words) and posted between "01 Dec, 2019 and 28 Apr, 2020"
abstract or title "SARS-CoV2 severity score" (match all words)

Study selection

One reviewer will screen all titles and abstracts against the eligibility criteria. Those that are excluded by the first reviewer will be reviewed by a second reviewer. Studies marked for inclusion by either reviewer will be included in the full-text review.

Full-text studies will be independently reviewed by two reviewers. Disagreements will be resolved by discussion between the reviewers and a third reviewer. Data will be extracted by two reviewers.

Study records

Data will be extracted on to a Microsoft Excel spreadsheet (Appendix). Data extraction will be piloted by two reviewers. Data items to be collected are listed in Table 2. A second set of questions will be completed for studies that developed a novel prognostic score (Table 3). Study quality will be assessed according to the Prediction model Risk Of Bias ASsessment Tool (PROBAST).[26,27] Reviewers' data forms will be collated. Disagreements will be resolved by discussion between the reviewers and a third reviewer.

Analysis plan

We will tabulate results, identifying scores that have been investigated for patients with COVID-19 and providing details of measures of effectiveness of these scores. We will also report the risk of bias, assessed according to PROBAST.[26,27] Results will be presented according to the setting, population and time-point of disease when they were used, and the outcomes they were used to predict. We anticipate too few results to assess for publication bias. We will only attempt to synthesize results quantitatively if multiple studies have used the same score in the same setting with the same outcome measure.

Table 3. Data to be collected (based on guidance by Moons and colleagues)[22]

Section	Question	Details / options
Study ID	Citation	
Source of data		e.g. single-hospital cohort, multi-centre cohort
Score used	Was the score used to guide clinical decisions?	Yes or no
Disease	Disease being investigated	COVID-19/SARS/MERS
Setting	Setting of study	e.g. primary care, hospital, ICU*
	Country	List all countries contributing data
	Study dates	
Participants	Eligibility	Study's eligibility criteria
	Diagnosis	Clinical or laboratory-confirmed COVID-19
	Number	Number of participants included
	Numbers by outcome	e.g. number recovered, number died, number ongoing care
	Ages	Median or mean, and IQR or SD
	Ethnicities	
	Sex ratio	
	Details of comorbidities	
Treatments	Details of treatments if differentially given based on prognosis score	
Score	Development or validation of a score	
	If validation, additional details	e.g. internal, temporal, bootstrapping, external
	Score name	
	Components of score	
	When in patient journey was score calculated	
	If score calculated retrospectively, was this blind to outcome?	
	If development, handling of predictors	
Outcome	Type of outcome	Single or composite
	Definition of outcomes recorded	e.g. Death / ICU/HDU admission / prolonged hospital stay
	How outcomes were assessed	
	Was assessor blind to the prognostic score?	
	Duration of follow-up	Median (IQR) or mean (SD)
Missing data	Number of participants with any missing data	
	Number of participants missing each predictor or outcome	
	Handling of missing data	
Model development†	Modelling method	e.g. logistic, survival, machine learning
	Modelling assumptions satisfied	
	Method for selection of predictors	

	into multivariable modelling	
	What parameters were considered	e.g. host-related, virus-related, physiological parameters, blood parameters, radiology
	Method for selection of predictors during multivariable modelling	
	Criteria used for selection of predictors	e.g. P-value, Akaike Information Criterion
	How were the predictors converted into a score?	e.g. 1 point per parameter, use of nomogram
Relationship between score and outcome	Measure used by study authors	e.g. AUC, OR, PPV*
	Value	Value of the Measure
Interpretation	Interpretation of results	
	Comparison with other studies / discussion of generalizability	
Notes or observations		Free-text

* Multiple rows may be used if results are presented in multiple settings, for multiple scores, or with multiple measures of the relationship.
† Only for use in studies developing a novel prognostic score

AUC, area under the (receiver operating characteristic) curve; COVID-19, coronavirus disease 2019; ICU, intensive care unit; IQR, inter-quartile range; MERS, Middle East respiratory syndrome; OR, odds ratio; PPV, positive-predictive value; SARS, severe acute respiratory syndrome; SD, standard deviation

References

1.   Mandell LA, Wunderink RG, Anzueto A, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis* 2007; **44 Suppl 2**: S27-72. doi: 10.1086/511159

2.   Alam N, Hobbelink EL, van Tienhoven AJ, van de Ven PM, Jansma EP, Nanayakkara PW. The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review. *Resuscitation* 2014; **85**(5): 587-94. doi: 10.1016/j.resuscitation.2014.01.013

3.   Haegdorens F, Monsieurs KG, De Meester K, Van Bogaert P. An intervention including the national early warning score improves patient monitoring practice and reduces mortality: A cluster randomized controlled trial. 2019; **75**(9): 1996-2005. doi: 10.1111/jan.14034

4.   Singanayagam A, Chalmers JD, Hill AT. Severity assessment in community-acquired pneumonia: a review. *QJM* 2009; **102**(6): 379-88. doi: 10.1093/qjmed/hcp027

5.   National Institute for Health and Care Excellence. Pneumonia: Diagnosis and management of community- and hospital-acquired pneumonia in adults [Clinical guideline 191]. London, UK: National Institute for Health and Care Excellence; 2014.

6.   Myles PR, Nguyen-Van-Tam JS, Lim WS, et al. Comparison of CATs, CURB-65 and PMEWS as triage tools in pandemic influenza admissions to UK hospitals: case control analysis using retrospective data. *PloS One* 2012; **7**(4): e34428. doi: 10.1371/journal.pone.0034428

7.   Semple MG, Myles PR, Nicholson KG, et al. An evaluation of community assessment tools (CATs) in predicting use of clinical interventions and severe outcomes during the A(H1N1)pdm09 pandemic. *PloS One* 2013; **8**(9): e75384. doi: 10.1371/journal.pone.0075384

8.   Wu JT, Leung K, Bushman M, et al. Estimating clinical severity of COVID-19 from the

transmission dynamics in Wuhan, China. *Nat Med* 2020. doi: 10.1038/s41591-020-0822-7

9.	Christian MD, Hawryluck L, Wax RS, et al. Development of a triage protocol for critical care during an influenza pandemic. 2006; **175**(11): 1377-81. doi: 10.1503/cmaj.060911 %J Canadian Medical Association Journal

10.	Christian MD, Sprung CL, King MA, et al. Triage: Care of the Critically Ill and Injured During Pandemics and Disasters: CHEST Consensus Statement. *Chest* 2014; **146**(4): e61S-e74S. doi: 10.1378/chest.14-0736

11.	Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020; **395**(10223): 507-13. doi: 10.1016/s0140-6736(20)30211-7

12.	World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report 94. Geneva: World Health Organization, 2020.

13.	Wynants L, Van Calster B, Bonten MMJ, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ (Clinical research ed)* 2020; **369**: m1328. doi: 10.1136/bmj.m1328

14.	Lim WS, van der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax* 2003; **58**(5): 377-82. doi: 10.1136/thorax.58.5.377

15.	Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med* 1997; **336**(4): 243-50. doi: 10.1056/nejm199701233360402

16.	Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985; **13**(10): 818-29.

17.	Intensive Care National Audit & Research Centre. ICNARC report on COVID-19 in critical care 24 April 2020 London: ICNARC, 2020.

18.	España PP, Capelastegui A, Gorordo I, et al. Development and validation of a clinical prediction rule for severe community-acquired pneumonia. *Am J Respir Crit Care Med* 2006; **174**(11): 1249-56. doi: 10.1164/rccm.200602-177OC

19.	Charles PG, Wolfe R, Whitby M, et al. SMART-COP: a tool for predicting the need for intensive respiratory or vasopressor support in community-acquired pneumonia. *Clin Infect Dis* 2008; **47**(3): 375-84. doi: 10.1086/589754

20.	Committee for the Japanese Respiratory Society Guidelines for the Management of Respiratory Infections. Severity classification of pneumonia. *Respirology* 2006; **11**(S3): S84-S5. doi: 10.1111/j.1440-1843.2006.00937_5.x

21.	Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis: For the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; **315**(8): 762-74. doi: 10.1001/jama.2016.0288

22.	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 medicine* 2014; **11**(10): e1001744. doi: 10.1371/journal.pmed.1001744

23.	Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009; **338**: b605. doi: 10.1136/bmj.b605

24.	Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *Journal of Clinical Epidemiology* 2005; **58**(5): 475-83. doi: 10.1016/j.jclinepi.2004.06.017

25.	Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Statistics in medicine* 2016; **35**(2): 214-26. doi: 10.1002/sim.6787

26.	Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019; **170**(1): 51-8. doi: 10.7326/m18-1376

27.   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**(1): W1-w33. doi: 10.7326/m18-1377