Prognostic scores for coronavirus disease 2019:

A draft analysis plan using data collected by the ISARIC Collaborators

# Background

There is international interest in whether prognostic scores can be used to improve the management of patients with coronavirus disease 2019 (COVID-19). Questions of interest include whether one score can be used across settings, or whether multiple scores will be needed. There are also questions regarding the changing utility of scores over the course of the pandemic. The International Severe Acute Respiratory and emerging Infection Consortium (ISARIC) Collaborators are collecting clinical data in more than thirty countries on patients admitted to hospital with COVID-19. This large dataset may allow very precise estimates of the mean effectiveness of scores, but it may also conceal true heterogeneity that is clinically relevant for clinicians intending to apply a score.

## Prognostic scores

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,1-3 and patients who can safely be discharged from hospital to complete their treatment at home.4 Scores may be used at different points during a patient’s journey. Common timepoints include at presentation in the community, on arrival at a hospital, or on transfer to an intensive care unit (ICU). Early Warning Scores are intended for repeated use throughout a patient’s stay in hospital to identify early physiological signs of deterioration.5 For the purpose of this analysis, we are interested in scores that can be calculated on admission to hospital.

# Objectives

1. To assess the utility of prognostic scores for COVID-19, comparing between different time periods and between settings.
2. To produce R code to allow investigators to assess scores within their local settings.

# Methods

## Participants

We will include all adults in the dataset. Patients whose data were entered into the ISARIC dataset were suspected to have COVID-19 at the time of admission. In clinical practice, any scores will be applied in this population. We will repeat the analysis limited to cases with laboratory-confirmed SARS-CoV-2 infection as a sensitivity analysis.

 The proposed analysis does not allow for censoring of patients who do not yet have an outcome. We therefore need to exclude participants who are still in hospital. To reduce bias from this, we will also exclude patients admitted in the two weeks immediately preceding the data extraction. This

also allows for potential delays in entering admission data on to the database, which would otherwise lead to an underestimate of the availability of data to calculate the scores.

## Outcomes

Many scores are developed with an outcome of mortality.6-8 Other outcomes include ICU admission,1,9 need for intensive respiratory or vasopressor support,10 or a composite of these. For emerging infections, the utility of various outcomes can vary over time. For example, when a pathogen is newly detected, patients may be admitted to intensive care units for the purpose of infection prevention (for example, use of negative-pressure side rooms). Conversely, in a pandemic situation, resources constraints may prevent ICU admission for patients who would otherwise have been admitted. An outcome of mortality is vulnerable to the same potential biases in reverse, as ineligibility for ICU admission may become a risk factor for mortality. For an outcome on in-hospital mortality, it is also important to consider the appropriate classification of patients who are discharged to a palliative care facility or to their own home with end-of-life medication. Coding these individuals as ‘survivors’ may inappropriately make certain comorbidities appear protective against mortality.

 In ISARIC collaborator meetings, the main two outcomes to be assessed were agreed as mortality, and a composite of mortality and need for invasive mechanical ventilation. This second outcome is intended to reduce some of the risks described above. Patients who are admitted to ICU for reasons other than ventilatory support will not be classified as having the outcome. Those who are unable to receive mechanical ventilation due to resource constraints are, sadly, likely to be identified in the mortality category.

## Candidate scores

Scores identified by the ongoing systematic review of prognostic scores for COVID-19 will be included.11 We will also include commonly used scores for other respiratory infections and acute illness, and any candidate scores proposed by investigators.

## Assessment of usability

For each component of a score, we will identify whether the variable is available for each participant. Where a score allows for a score to be calculated from a variety of variables (for example, Pneumonia Severity Index includes points for either arterial oxygen tension <8 kPa or peripheral oxygen saturation <90%), this will be considered non-missing if either variable is available. Similarly, for scores that include confusion, this will be considered non-missing either if the presence or absence of confusion is reported, or if Glasgow Coma Scale is reported.

## Calculation of scores

For each participant who has all variables for a score non-missing, the score will be calculated. Secondarily, the scores will be calculated for all participants, with missing data imputed as the absence of the risk factor. For scores that require a physiological or biochemical value to be included, we will use the midpoint of the usual physiological range (adjusted for age and sex if appropriate). No attempt will be made to impute missing age data.

## Calibration

For scores developed for COVID-19, we will plot the observed outcomes in the ISARIC dataset against the predicted outcomes from the study in which the score was developed. For all scores, we will plot the proportions with an outcome against the score.

## Discrimination

We will plot receiver operating characteristic (ROC) curves for each score. The area under the curve (equivalent to the concordance [*c*] statistic) will be calculated for each. Positive- and negative-predictive values will be calculated for various score thresholds.

## Units of analysis

To assess a prognostic score reliably requires a minimum of 100, and ideally 200, participants who experience the outcome event.12 For countries reaching the 200 events threshold, we will conduct the analysis for individual countries. Other countries will be grouped according to World Bank Income classifications and continents. Heterogeneity between areas will be assessed

## Variation over time

For populations with sufficient numbers of patients, we will divide the period of recruitment and compare the performance of scores at different time points. Initially the time will be divided into three equal periods. We will also investigate dividing time according to dates of events in participating countries, such as peak numbers of cases.

## Statistical tests

Differences in discrimination of scores within a population will be assessed using a non-parametric test for paired data.13 Heterogeneity between populations will be assessed using a non-parametric test for unpaired data.14

# Outputs

This analysis is intended to produce:

1. Data for a peer-reviewed manuscript to be submitted jointly by the ISARIC investigators, in line with the ISARIC publication policy.
2. R code that participating sites can use to analyse their own data.

# Participatory approach

All contributors to the ISARIC database are invited to participate in this analysis through review and input on the statistical analysis plan and resulting publication. The outputs of this work will be disseminated as widely as possible to inform patient care and public health policy, this will include submission for publication in an international, peer-reviewed journal. ISARIC aims to include the names of all those who contribute data in the cited authorship of this publication, subject to the submission of contact details and confirmation of acceptance of the final manuscript within the required timelines.

# 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. 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

3. 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

4. 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.

5. 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

6. 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

7. 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

8. 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

9. 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

10. 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

11. Pritchard M, Sigfrid L, Thomson B, et al. Prognostic scores developed or validated for coronavirus disease 2019: A rapid systematic review. 2020. <https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020183746>.

12. 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**(2): 214-26. doi: 10.1002/sim.6787

13. Venkatraman ES, ;, Begg CB. A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment. *Biometrika* 1996; **83**(4): 835-48.

14. Venkatraman ES. A permutation test to compare receiver operating characteristic curves. *Biometrics* 2000; **56**(4): 1134-8. doi: 10.1111/j.0006-341x.2000.01134.x