

Flaws in the development and validation of a covid-19 prediction model

Gary S. Collins, PhD

Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Windmill Road, Oxford OX3 7LD, United Kingdom. Email: gary.collins@csm.ox.ac.uk

Richard D. Riley, PhD

Centre for Prognosis Research, School of Medicine, Keele University, Staffordshire, ST5 5BG, United Kingdom.

Maarten van Smeden, PhD

Julius Center for Health Science and Primary Care, University Medical Center Utrecht, University of Utrecht, Utrecht, the Netherlands.

Corresponding author:

Gary S. Collins, Ph.D

e-mail: gary.collins@scm.ox.ac.uk

To the Editor –

The covid-19 pandemic has seen the development of a number of clinical prediction models to support assessing disease severity or aiding prognosis. A recent systematic review has identified 145 models, concluding all to be at high risk of bias citing concerns with data quality, statistical analysis and reporting leading to the conclusion that none are recommended for use [1]. We therefore read with interest the recent paper by Dong and colleagues describing the development of a prediction model for assessing survival in covid-19 patients [2]. Unfortunately, we observed a number of concerns in the study which we believe deserve highlighting to readers.

In their paper, the authors randomly split their data into training and validation sets – a practice widely known to be statistically inefficient [3] – it reduces the sample size for developing the model (increasing the risk of overfitting), whilst the validation cohort is often too small to evaluate performance. The authors then develop a model on the training cohort containing ~75 deaths (60% of 121 deaths) and examining over 30 candidate predictors – this will almost certainly lead to an overfit model [4]. The resulting model was then evaluated in the validation cohort which contained only ~46 deaths (40% of 121 deaths), which is much lower than the recommended 100 deaths needed for validation [5]. The preferred approach, making the most of the data, would be to develop the model on all available data, and use methods like bootstrapping or cross-validation to both adjust the performance measures (e.g., c-index) for optimism (which are almost certainly too high) and shrink the model coefficients (which will likely be overestimated [6]). Other major analysis concerns include categorization of continuous predictors (which results in loss of information [7]), no mention of missing data, use of lasso followed by ‘multivariate’(sic) Cox regression to screen predictors for inclusion, incorrect (i.e., does not reflect the actual model building process) and confusing implementation of cross-validation on the validation data, weak assessment of model calibration by binning observations, and assessment of both the AUC and the c-index – both assess discrimination (however, only the latter is correct for survival outcomes), and no assessment of clinical utility.

Our second point relates to model presentation. The authors presented their model as a nomogram presumably to aid clinical uptake [8]. However, for other investigators to externally validate this model in their own data to see, it is important that the model underpinning the nomogram is fully reported, namely all the regression coefficients and importantly the baseline survival at the timepoints of interest (i.e., at 14 and 21 days) – this latter information is missing, impeding independent external validation of the model.

The TRIPOD Statement (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) for reporting (covid) prediction models should be consulted (www.tripod-statement.org) so that important information is presented when authors describe their study [9]. Accompanying the TRIPOD Statement is an Explanation and Elaboration paper that describes the rationale for reporting, as well as discussing various methodological aspects [10].

The authors report no conflicts of interest.

REFERENCES

1. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* **2020**; 369:m1328.
2. Dong Y-M, Sun J, Li Y-X, et al. Development and Validation of a Nomogram for Assessing Survival in Patients with COVID-19 Pneumonia. *Clinical Infectious Diseases* **2020**; :ciaa963.
3. Steyerberg EW, Harrell Jr FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* **2001**; 54:774–781.
4. 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* **2018**; :962280218784726.
5. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* **2016**; 35:214–226.
6. Riley RD, Van Der Windt DA, Croft P, Moons KG. *Prognosis Research in Healthcare: Concepts, Methods and Impact*. Oxford: Oxford University Press, 2019.
7. Collins GS, Ogundimu EO, Cook JA, Le Manach Y, Altman D. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Stat Med* **2016**; 35:4124–4135.
8. Bonnett LJ, Snell KIE, Collins GS, Riley RD. Guide to presenting clinical prediction models for use in clinical settings. *BMJ* **2019**; 365:1737.
9. Collins GS, Reitsma JB, Altman D, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis: The TRIPOD statement. *Ann Intern Med* **2015**; 162:55–63.
10. 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–W73.