

**TITLE: There are no shortcuts 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@esm.ox.ac.uk](mailto:gary.collins@esm.ox.ac.uk)

**Jie Ma, MSc**

Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Windmill Road, Oxford OX3 7LD, United Kingdom.

**Paula Dhiman, PhD**

Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Windmill Road, Oxford OX3 7LD, United Kingdom.

Word count: 900

The authors report no conflict of interest.

Data sharing not applicable to this article as no datasets were generated or analysed during the current study

**Letter Re: Dong Y, Zhou H, Li M, Zhang Z, Guo W, Yu T, Hui Y, Wang Q, Zhao L, Luo S, Fan H, Hu D. A novel simple scoring model for predicting severity of patients with SARS-CoV-2 infection. *Transbound Emerg Dis* 2020 [DOI: 10.1111/tbed.13651]**

A recent living systematic review has identified 145 covid-19 prediction models published up until May 2020, to support clinical decision-making during the global covid-19 pandemic (Wynants et al., 2020). Despite this surge in developing prediction models, the systematic review concluded that all these models are at high risk of bias citing concerns regarding poor data quality, flaws in the statistical analysis and incomplete or poor reporting. As a consequence none are recommended for use (Wynants et al., 2020). We therefore read with interest the recent paper by Dong and colleagues describing the development of a prediction model to identify covid-19 patients with severe disease (Dong et al., 2020). Unfortunately, as with existing covid-19 prediction models (Wynants et al., 2020), we observed a number of concerns in the study by Dong and colleagues with regards to the study design, statistical analysis and reporting, which we believe warrant highlighting to readers.

In terms of design, the sample size in a prediction model study depends on the proportion of individuals experiencing the event to be predicted (in Dong's study, this is individuals with

severe disease) and the number of predictors considered for inclusion in the model. To minimise overfitting, using sample size formulae for deriving prediction models (Riley et al., 2019, 2020), based on information reported in the Dong study (30 predictors, outcome proportion of 0.361), the minimum sample size in the most optimistic scenario (e.g., that the model gives the highest R-squared) would be 430 individuals (156 events). To precisely estimate the intercept, requires 355 individuals (128 events). The study by Dong and colleagues included 147 individuals in total, where 53 had the outcome of severe disease – this is noticeably lower than the aforementioned sample size in even the most optimistic scenario. The prediction model by Dong is therefore, at considerable risk of overfitting – such that the model will appear to work well in the data used to develop the model, but has poor accuracy when applied in new data (or new individuals).

Due to the potential for overfitting during the development of a prediction model, it is widely recommended that all studies include some meaningful internal validation (using cross-validation or bootstrapping), to estimate the performance of the model and correct for any optimism/overestimation (Moons et al., 2015). The study by Dong, carried out no such internal validation, and merely evaluated the predictive accuracy of the model using the same data that was used to develop the model, reporting an area under the receiver operating characteristic curve of 0.843. Whilst this may appear to be high, and thus indicating good discrimination, this value will almost certainly be optimistic and overestimated. The lack of internal validation, and no external validation (i.e., evaluating the model in a separate dataset), is a major limitation.

Other concerns include the data quality, namely the presence and handling of missing data. Missing values are largely unavoidable, and the study by Dong included 30 predictors - in the absence of any mention of missing data, one can only assume that individuals with missing were excluded from the analysis – such an approach not only reduces sample size, which as we have already discussed raises the risk of overfitting, but also makes a strong and unlikely assumption on the reason for missingness. The authors are strongly recommended to report their missing data and consider approaches to handle missing data such as multiple imputation (Vergouwe et al., 2010). Dong et al also assumed all continuous measurements were linearly associated with the outcome, this is not only unrealistic, but the consequence of making such an assumption will be a loss in predictive accuracy (Collins et al., 2016). The authors also used univariate analyses to screen predictors to include in the multivariable modelling, which has long been discredited as important predictors can be erroneously omitted (which only become significant after adjustment for other predictors) or unimportant predictors spuriously included (Sun et al., 1996).

How a model is presented is important so that others can use or evaluate it in their own setting. The authors have presented both their logistic regression model and a simple score that was based on the values of the regression coefficients from the logistic regression model. However, the presented logistic regression model does not include the intercept, rendering it useless to predict individualized risk in new individuals. The full regression model, including the intercept, is important for individualized predictions (Collins et al., 2015). Further, no

information was presented on how the simple score was derived from the logistic regression model. Such a simplification also limits its usefulness for individualized prediction, as the simplified model yields no predicted risks – just a unitless numerical integer which is then dichotomized to classify an individual into low or high risk (though without defining what high or low risk). If a user-friendly format to aid model uptake is required then a principled approach should be adopted (Bonnett et al., 2019).

When writing up findings from a (covid) prediction model study, authors should adhere to the TRIPOD Statement ([www.tripod-statement.org](http://www.tripod-statement.org)), so that key information is reported when describing their prediction model study. Full and transparent reporting enables readers to have the minimal information required to judge the quality of the study (Collins et al., 2015). The accompanying TRIPOD Explanation and Elaboration paper also discusses various methodological considerations, highlighting appropriate and inappropriate methods in the design and analysis of a prediction model study (Moons et al., 2015).

## ETHICS STATEMENT

The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to. No ethical approval was required as this is a Letter to the Editor with no original research data.'

## REFERENCES

- Bonnett, L.J., K.I.E. Snell, G.S. Collins, and R.D. Riley, 2019: Guide to presenting clinical prediction models for use in clinical settings. *BMJ* **365**, 1737, DOI: 10.1136/bmj.1737.
- Collins, G.S., E.O. Ogundimu, J.A. Cook, Y. Le Manach, and D. Altman, 2016: Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Stat Med* **35**, 4124–4135, DOI: 10.1002/sim.6986.
- Collins, G.S., J.B. Reitsma, D. Altman, and K.G. Moons, 2015: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis: The TRIPOD statement. *Ann Intern Med* **162**, 55–63.
- Dong, Y., H. Zhou, M. Li, Z. Zhang, W. Guo, T. Yu, Y. Gui, Q. Wang, L. Zhao, S. Luo, H. Fan, and D. Hu, 2020: A novel simple scoring model for predicting severity of patients with SARS-CoV-2 infection. *Transbound Emerg Distbed*.13651, DOI: 10.1111/tbed.13651.
- Moons, K.G.M., D.G. Altman, J.B. Reitsma, J.P.A. Ioannidis, P. Macaskill, E.W. Steyerberg, A.J. Vickers, D.F. Ransohoff, and G.S. Collins, 2015: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* **162**, W1–W73.
- Riley, R.D., J. Ensor, K.I.E. Snell, F.E. Harrell, G.P. Martin, J.B. Reitsma, K.G.M. Moons, G. Collins, and M. van Smeden, 2020: Calculating the sample size required for developing a clinical prediction model. *BMJ* **368**, m441, DOI: 10.1136/bmj.m441.
- Riley, R.D., K.I. Snell, J. Ensor, D.L. Burke, F.E. Harrell Jr, K.G. Moons, and G.S. Collins, 2019: Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Statistics in Medicine* **38**, 1276–1296, DOI: 10.1002/sim.7992.
- Sun, G.W., T.L. Shook, and G.L. Kay, 1996: Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol* **49**, 907–916.

- Vergouwe, Y., P. Royston, K.G.M. Moons, and D.G. Altman, 2010: Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol* **63**, 205–214.
- Wynants, L., B. Van Calster, G.S. Collins, R.D. Riley, G. Heinze, E. Schuit, M.M.J. Bonten, D.L. Dahly, J.A.A. Damen, T.P.A. Debray, V.M.T. de Jong, M. De Vos, P. Dhiman, M.C. Haller, M.O. Harhay, L. Henckaerts, P. Heus, N. Kreuzberger, A. Lohmann, K. Luijken, J. Ma, G.P. Martin, C.L. Andaur Navarro, J.B. Reitsma, J.C. Sergeant, C. Shi, N. Skoetz, L.J.M. Smits, K.I.E. Snell, M. Sperrin, R. Spijker, E.W. Steyerberg, T. Takada, I. Tzoulaki, S.M.J. van Kuijk, F.S. van Royen, J.Y. Verbakel, C. Wallisch, J. Wilkinson, R. Wolff, L. Hooft, K.G.M. Moons, and M. van Smeden, 2020: Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* **369**, m1328, DOI: 10.1136/bmj.m1328.