CORRESPONDENCE

The Inherent Problems With the Generalizability of the CALL Score: Towards Reliable Clinical Prediction Models for Coronavirus Disease 2019 (COVID-19)

To THE EDITOR—As the coronavirus disease 2019 (COVID-19) pandemic has evolved and caused a large number of infections and deaths, clinical prediction models that either aim to identify subjects at risk in the general population or to predict disease progression and mortality are becoming increasingly important. A recently published clinical prediction model, the CALL score, developed by Ji et al, showed good performance in predicting COVID-19 disease progression [1]. To develop a robust and user-friendly model, the authors performed internal validation, selected only 4 predictors, and developed a simple prediction model. As a result, the developed model was expected to be relatively robust and easy to use for clinicians. Although the model showed a high predictive ability, we have 2 concerns regarding its generalizability.

Our first concern is the variable selection. In that study, the authors defined 4 variables (CALL; comorbidity, age, lymphocyte, and lactate dehydrogenase [LDH]) as candidate predictors based on the *P* value-based criteria (ie, $P < .05$), in the multivariable Cox regression model. Although these predictors were clinically important in the study, it is unknown whether they would be equally important in other clinical settings. In other words, the CALL model may not have included other clinically important variables. For example, many studies suggested that obesity, defined as body mass index ≥ 30 kg/m^2, is an important risk factor for severe illness from COVID-19 [2]; however, the CALL model did not contain obesity as a predictor. Lack of such an important predictor may lead to a lower predictive ability of the model after external validation. One simulation study suggested that 5 of the 6 variables selected via the *P* value-based

criterion ($P < .05$) showed more than 50% biased estimation of β coefficients in the multivariable logistic regression model [3]. To make the prediction model more generalizable, variable selection based on both clinical reasoning and external literature should be considered [3, 4].

Our second concern is regarding missing values. The authors did not show the number of study participants with missing data for the predictors. When missing data exists, the authors should consider the mechanism of missingness (missing completely at random, missing at random, or missing not at random), and they should deal with missing data using an appropriate method, such as multiple imputation [5]. If the authors used complete-case analysis despite the presence of missing data, they would have overestimated the results due to overfitting [6]. The discriminative ability of the CALL score is 0.91, which was too high to predict the outcome using only 4 variables; hence, overfitting due to missing data was suspected. To confirm transparency and generalizability, presenting and dealing with missing values are needed.

In summary, the authors may have shown an overestimated model performance due to inappropriate variable selection and missing data. Many COVID-19 prediction models are known to have a high risk of bias, and such unreliable models may result in more harm than good in clinical decision making [7]. To develop prediction models with high predictive abilities, using robust methodologies, such as the TRIPOD guidelines [5], is essential.

Notes

Takashi Yoshioka,[1,2,] **Satoshi Funada,**[3,4] **and Yan Luo**[4]

[1]Center for Innovative Research for Communities and Clinical Excellence (CiRC2LE), Fukushima Medical University, Fukushima City, Fukushima, Japan, [2]Department of Healthcare Epidemiology, Kyoto University School of Public Health in the Graduate School of Medicine, Kyoto City, Kyoto, Japan, [3]Department of Urology, Kyoto University Graduate School of Medicine, Kyoto City, Kyoto, Japan, and [4]Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto City, Kyoto, Japan

References

1. 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**; 71:1393–9. Available at: https://doi.org/10.1093/cid/ciaa414.
2. Chang TH, Chou CC, Chang LY. Effect of obesity and body mass index on coronavirus disease 2019 severity: A systematic review and meta-analysis. Obes Rev **2020**; 21:e13089. Available at: https://doi.org/10.1111/obr.13089.
3. Steyerberg EW, Uno H, Ioannidis JPA, et al. Poor performance of clinical prediction models: the harm of commonly applied methods. J Clin Epidemiol **2018**; 98:133–43. Available at: https://doi.org/10.1016/j.jclinepi.2017.11.013.
4. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. Fam Med Community Heal **2020**; 8:e000262. Available at: https://doi.org/10.1136/fmch-2019-000262.
5. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med **2015**; 162:55. Available at: https://doi.org/10.7326/M14-0697.
6. Ambler G, Omar RZ, Royston P. A comparison of imputation techniques for handling missing predictor values in a risk model with a binary outcome. Stat Methods Med Res **2007**; 16:277–98. Available at: https://doi.org/10.1177/0962280206074466.
7. Wynants L, Calster BV, Collins GS, et al. Prediction models for diagnosis and prognosis of Covid-19: systematic review and critical appraisal. BMJ **2020**; 369:m1328. Available at: https://doi.org/10.1136/bmj.m1328.