Generalizability of COVID-19 Clinical Prediction Models

Shubhada Hooli

Department of Pediatrics, Section of Pediatric Emergency Medicine, Baylor College of Medicine, Houston, Texas, United States of America

Carina King

Department of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:

Shubhada Hooli, MD, MPH

Assistant Professor of Pediatrics

Department of Pediatrics, Section of Pediatric Emergency Medicine

Baylor College of Medicine

6621 Fannin Street, Suite A2210.00

Houston, TX 77030-2299

hooli@bcm.edu

m 727-505-9525

Dear Editor,

In “Epidemiological and Clinical Predictors of COVID-19” by Sun and Koh et al. compare exposure, demographic, clinical and diagnostic test characteristics between COVID-19 PCR confirmed cases and PCR negative cases evaluated at the designated screening and referral hospital in Singapore.[1] The authors then present four COVID-19 case prediction models. We question the reproducibility of their results.

Multivariable logistic regression models can be overfitted to their derivation sample when the predictor to outcome of interest ratio is greater than 1:10. Overfitting a logistic regression model can lead to spuriously high area under the receiver operating curve implying good model discrimination. However, this limits the generalizability applied to another population. Each of the models violate this principal (Table 1), although we note that model 4, with the poorest performance was close to meeting this criteria.

Nearly every nation has limited testing resources in the face of this rapidly progressing pandemic. Case identification tools could play a crucial role in containment and mitigation strategies. This is why it is extremely important that models are designed with a focus on generalizability of findings, that includes clearly defined predictors. When describing the clinical characteristics of patients included in the models the authors do not provide sufficient detail for others to replicate and externally validate their tool, with descriptors such as “elevated body temperature” and “elevated respiratory rate”.

No other disease process, in recent memory, has captured the world's attention like COVID-19. Appropriately, scientists are racing to better understand and mitigate this global pandemic; a pre-print review uploaded on the 27th March already identified 19 COVID-19 prediction models and also raised concerns about the quality of these tools.[2] During such dire circumstances, more than ever we must be vigilant to uphold our own standards.

Neither author has any conflicts to disclose.

References:

1. Sun Y, Koh V, Marimuthu K, Ng OT, Young B, Vasoo S, et al. Epidemiological and Clinical Predictors of COVID-19. Clin Infect Dis. **2020**. doi:10.1093/cid/ciaa322

2. Wynants L, Calster BV, Bonten MM, Collins GS, Debray TP, Vos MD, et al. Systematic review and critical appraisal of prediction models for diagnosis and prognosis of COVID-19 infection. medRxiv [Preprint]. March 27, 2020 [cited April 1, 2020]. Available from: doi:10.1101/2020.03.24.20041020

Table 1: Characteristics of COVID-19 Case Prediction Models					
Model	Total (N)	Controls (N)	Cases (N)	Predictors (N)	AUC
1	292	243	49	16	0.91
2	292	243	49	11	0.88
3	292	243	49	13	0.88
4	788	734	54	6	0.65