ASSESSING ADHERENCE TO EXISTING MODEL REPORTING GUIDELINES BY COMMONLY USED CLINICAL PREDICTION MODELS FROM AN ELECTRONIC HEALTH RECORD VENDOR

Jonathan H. Lu, MS[1]; Alison Callahan, PhD[1]*; Birju S. Patel, MD, MPH[1]*; Keith E. Morse, MD, MBA[2,3]*; Dev Dash, MD[1]; Nigam H. Shah, MBBS, PhD[1]

[1] Center for Biomedical Informatics Research, Stanford University School of Medicine, Stanford, CA
[2] Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA
[3] Department of Clinical Informatics, Lucile Packard Children's Hospital, Palo Alto, CA, USA
* These authors contributed equally.

Model reporting guidelines have been proposed to ensure clinical prediction models are reliable, fair, and useful, but the adherence of currently deployed models to these guidelines has not been studied. Our objective was to assess information requested by model reporting guidelines and 2) whether the documentation for commonly used machine learning models developed by an electronic health record vendor provides the information requested. We queried PubMed using: "machine learning model card" and "reporting machine learning" in November 2020, reviewed references to find additional publications, and excluded publications without specific model reporting recommendations. We merged similar items requested for reporting into representative "atoms." Four independent reviewers and one adjudicator assessed how often model documentation for the most commonly used models developed by Epic Systems reported the atoms.

Combining recommendations from 15 model reporting guidelines, we identified 220 unique requested items. We reviewed documentation of 12 commonly deployed Epic models and assessed completion rates of applicable items. The median completion rate was 39%. While the most commonly requested items were highly reported, information on reliability (including external validation, uncertainty measures, and strategy for handling missing data), transparency (model coefficients) and fairnes ((summary statistics and subgroup analyses, including for age, race/ethnicity, or sex) was missing from at least half of documentation.

In conclusion, there are many recommendations about what should be reported about clinical predictive models. Existing model documentation from the one model vendor examined in this study provided less than half of applicable atoms, and entire reporting guidelines have low adherence rates. Half or less of reviewed documentation reported information related to reliability, transparency and fairness of models. There is a need for better operationalization of reporting recommendations for clinical predictive models.