Why are Machine Learning-based prediction models still unpopular in clinical practice?

Constanza L Andaur Navarro[1,2], Johanna Damen[1,2], Toshihiko Takada[1,2], Paula Dhiman[3], Gary S Collins[3], Jie Ma[3], Ram Bajpai[4], Richard D. Riley[4], Lotty Hooft[1,2], Karel G Moons[1,2]

[1]*Julius Center for Health Sciences and Primary Care, UMC Utrecht, Utrecht University, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands*
[2]*Cochrane Netherlands, UMC Utrecht, Utrecht University, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands*
[3]*Center for Statistics in Medicine, University of Oxford, Windmill Road, OX1 2JD Oxford, United Kingdom*
[4]*School of Primary, Community and Social Care, Keele University, David Weatherall Building, ST5 5BG Keele, United Kingdom*

Background. Studies addressing diagnostic and prognostic prediction models are abundant in many clinical domains. At the same time, many systematic reviews showed that the quality of reporting of prediction model studies is suboptimal.[1] Due to the increasing availability of larger, routinely collected and complex data, and the rising application of Artificial Intelligence (AI) and Machine Learning techniques (ML) for clinical predictions, the number of prediction models is expected to increase even further. These AI/ML-based prediction model studies are often labeled as a "black box" and not much is know yet about the quality of reporting. Aims. The aim of this systematic review is to evaluate the reporting and methodological conduct of prediction model studies that applied AI/ML techniques for model development or validation. Methods. Our protocol was registered in PROSPERO (CRD42019161764). A search was performed in January 2020 to identify primary studies developing and/or validating prediction models using any AI/ML methodology across all medical fields. Studies were included if predicted patient-related outcomes, used any study design and were published in 2018-2019. We assessed (1) the quality of reporting by measuring the adherence to Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis guideline (TRIPOD) and (2) the risk of bias in prediction model development or validation using the Prediction model Risk of Bias Assessment Tool (PROBAST). Results. Initial results from the review will be presented, stratified by medical field and prevalent AI/ML methods. Conclusion. Emerging issues will be discussed, as well as the necessity for specific reporting (TRIPOD-AI/ML) and risk of bias (PROBAST AI/ML) assessment for AI/ML-based prediction model studies.

Keywords
Systematic Review, Machine Learning, Prediction.

References
[1] Heus P, Damen JAAG, Pajouheshnia R, et al. Poor reporting of multivariable prediction model studies: Towards a targeted implementation strategy of the TRIPOD statement. BMC Med. 2018;16(1):1-12.