

Prediction under hypothetical interventions: evaluation of counterfactual performance using longitudinal observational data

[Nan van Geloven \(LUMC\)](#)

Predictions under hypothetical interventions are estimates of what a person's risk of an outcome would be if they were to follow a particular treatment strategy, given their individual characteristics. Such predictions can give important input to medical decision making. However, evaluating predictive performance of interventional predictions is challenging. Standard ways of evaluating predictive performance do not apply when using observational data, because prediction under interventions involves obtaining predictions of the outcome under conditions that are different to those that are observed for a subset of individuals in the validation dataset. This work describes methods for evaluating counterfactual predictive performance of predictions under interventions for time-to-event outcomes. This means we aim to assess how well predictions would match the validation data if all individuals had followed the treatment strategy under which predictions are made. We focus on counterfactual performance evaluation using longitudinal observational data, and under treatment strategies that involve sustaining a particular treatment regime over time. We introduce an estimation approach using artificial censoring and inverse probability weighting which involves creating a validation dataset that mimics the treatment strategy under which predictions are made. We extend measures of calibration, discrimination (c-index and cumulative/dynamic AUC) and overall prediction error (Brier score) to allow assessment of counterfactual performance. The methods are evaluated using a simulation study, including scenarios in which the methods should detect poor performance. Applying our methods in the context of liver transplantation shows that our procedure allows quantification of the performance of predictions supporting crucial decisions on organ allocation.