Can we combine propensity score modelling and patient-level prediction to make counter-factual predictions?

Jenna M. Reps and Chris Knoll
Janssen Research and Development, Titusville, NJ, United States

Background

The OHDSI community have developed advanced frameworks for causal inference and patient-level prediction [1]. However, the causal inference is population based and the patient-level prediction cannot be used to aid treatment choice as it does not account for confounding or other forms of bias commonly seen in observational data. This means it is currently not possible to use the current OHDSI tools and large observational healthcare data to personalize medical treatment choices.

If large clinical trial data were available, where patients were randomly assigned treatment, it would be possible to developed prediction models that can personalize treatment. This is due to the randomization minimizing potential bias, given the data are large enough. Using randomized data, prediction models could be developed that include the treatment choice as a predictor. Before a patient is assigned treatment, the model could be applied under each hypothetical treatment option and the treatment that has the lowest probability for a set of negative outcomes or highest probability for a set of positive outcomes would be given to the patient.

Propensity score models are often used in population level causal inference to mimic randomization [2]. In this feasibility study we investigate using propensity score modelling to create a data set that is similar to a randomized clinical trial. We find a group of hypertension patients who had fairly equal propensity for two different initial treatments and restrict the model development population to this population while developing models using treatment as a predictor. We develop a model for a known adverse drug reaction to one of the treatments.

Methods

We use Optum's de-identified Clinformatics® Data Mart Database (Optum Clinformatics®) mapped to the OMOP CDM. Optum Clinformatics® is a large US insurance claims database.

We identified new users of lisinopril or amlodipine with >365 days prior observation, aged 18 or older with an exposure on or after 2020. We restricted to patients with a history of hypertension. The CohortMethod R package was used to create a propensity model to predict the probability of being given lisinopril. The prediction target population was created by trimming to a preference score between 0.3-0.7 and then applying 1-1 matching.

We then developed patient-level prediction models, using the R PatientLevelPrediction package [3], for the task: which patients in the target population will have liver injury during treatment exposure. Liver injury is listed as a known side effect of taking lisinopril.

We developed a gradient boosting machine model with treatment as a predictor as well as age/sex, conditions grouped using the hierarchy and drug ingredients. The model development used 75% of the data for model training and 25% of the data for model testing. 3-fold cross validation was applied in the training data to learn the optimal hyper-parameters. A final model was fit using the optimal hyper-parameters and all the training data. The model performance was accessed internally by comparing the predicted risk and the true risk in the testing data. Model performance was evaluated using the area under the receiver operating characteristic curve (AUROC), calibration-in-the-large and brier score.

We then calculated the predicted risk for all test patients if they had been given lisinopril and the predicted risk for all test patients if they had been given amlodipine. We divided the mean lisinopril risk by the mean amlodipine risk to estimate the population risk ratio. This was compared with the relative risk obtained from a propensity matched cohort method analysis.

Results & Discussion

We identified 978,893 new users of lisinopril or amlodipine with inclusion criteria. After trimming 398,943 remained (195,009 lisinopril and 203,934 amlodipine). After 1:1 matching 320,932 remained in our final target population (160,466 each). The propensity plots for the full, trimmed and trimmed than matched as shown in Figure 1.

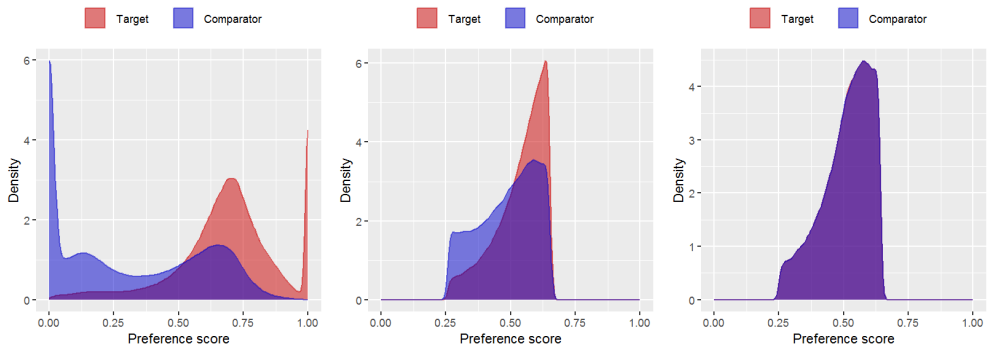

Figure 1 - Propensity score plots for the full population, trimmed population and trimmed then matched populations.

10,272 patients were excluded due to having prior liver injury, leaving 310,660 patients. 4380 patients (1.4%) experienced liver injury while exposed to the drugs.

The outcome model obtained an AUROC of 0.77 (0.76- 0.78) in the test data and a calibration-in-the-large predicted risk of 0.01402 and observed risk of 0.01410. Figure 2 shows the discrimination and calibration plots.

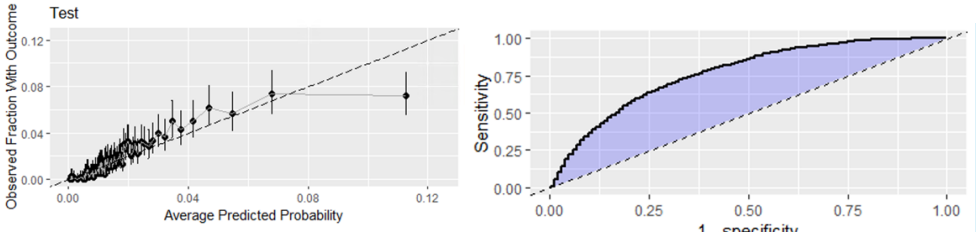

Figure 2 - Calibration and ROC plots for the test data.

The mean predicted risk when lisinopril was set to true and amlodipine set to false for all the patients (test patients) was 0.01548191 (0.01540209) and the mean predicted risk lisinopril was set to false and amlodipine set to true for all the patients was 0.01267826 (0.0126168). This gives a population risk ratio of 1.22 (1.22 in test set patients only).

The odds ratio estimate from cohort method after trimming and matching was 1.39 (1.31 - 1.48). When restricting to the same patients used in the prediction test set, the estimate was 1.16 (0.91 - 1.48).

The results show that the outcome model has a high discrimination and calibration and in addition, the population level estimate of the effect estimated using the prediction model is close to the effect estimated using cohort method. This shows the methodology may be suitable for counterfactual prediction.

It is of note, that the methodology is only applicable when a patient could have been given an alternative treatment. If a patient has a high propensity for one treatment, then there is likely a clinical reason for them to have that treatment and therefore personalizing their treatment is not suitable. To implement this methodology in a real-world setting, first the propensity model must be applied to the patient and if the patient has a preference score between 0.3-0.7, then the outcome model can be applied setting each treatment to true to predict their risks under each treatment.

Conclusions

In this paper we investigate a simple and novel approach to develop counter-factual prediction models that could be used to aid medical treatment choice. Our preliminary results suggest that the methodology may work, and future research should expand this feasibility study to more drugs and outcomes. In addition, it may be possible to pair this methodology with risk-benefit analyses to help guide patients towards their preferred medical treatment.

References

1. Schuemie M, Reps J, Black A, Defalco F, Evans L, Fridgeirsson E, Gilbert JP, Knoll C, Lavallee M, Rao GA, Rijnbeek P, Sadowski K, Sena A, Swerdel J, Williams RD, Suchard M. *Health-Analytics Data to Evidence Suite (HADES): Open-Source Software for Observational Research*. Stud Health Technol Inform. 2024 Jan 25;310:966-970. doi: 10.3233/SHTI231108.
2. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011 May;46(3):399-424. doi: 10.1080/00273171.2011.568786. Epub 2011 Jun 8. PMID: 21818162; PMCID: PMC3144483.
3. Jenna M Reps, Martijn J Schuemie, Marc A Suchard, Patrick B Ryan, Peter R Rijnbeek, Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data, *Journal of the American Medical Informatics Association*, Volume 25, Issue 8, August 2018, Pages 969–975, https://doi.org/10.1093/jamia/ocy032