Marginal structural models using calibrated weights with SuperLearner: application to longitudinal diabetes cohort

Sumeet Kalia; Oili Saarela; Michael Escobar; Tahen O'Neill; Christopher Meaney; Jesscia Gronsbelt; Babak Allarzadeh; Ervin Sejdic; Rahim Moineddin; Conrad Pow; Frank Sullivan; Michelle Greiver

Introduction

• We were interested in estimating “diabetes care provision” in next calendar year using a composite measure of chronic disease prevention and screening elements (Nabet et al., 2007).
• We constructed the hypothetical treatment cohorts using the referent modality (i.e. treatment naïve cohort) and indexed modality (i.e. treatment drop-in cohort).
• We described the “treatment naïve” cohort as the absence of treatment regimen while the “treatment drop-in” cohort as the initiation of treatment post-baseline (Lin et al., 2021).
• For example, we considered a hypothetical treatment contrast in which the patients are not prescribed glucose-lowering medications during the study period and we may use this cohort to describe the referent regimen in relation to treatment regimens imposed using the treatment drop-in cohorts.
• We characterized the individual-level treatment regimes with respect to the clinical profile of each patient using the conditional average treatment effect using the time-varying outcome-predictors (i.e. effect modifiers) including annual laboratory regulation, vaccination, lifestyle information, billing ICD9 codes and OHIP billing codes.

Methods

• We defined a composite binary endpoint using the sum of eight elements of diabetes provision within a calendar year: (i) primary care visit, (ii) blood pressure, (iii) weight, (iv) hemoglobin A1c, (v) lipid, (vi) ACR, (vii) eGFR and (viii) statin medication.
• We formulated the hypothetical framework using the potential outcomes (i.e. counterfactual outcomes) in the context of causal inference. Since this study relied on the use of longitudinal repeated-measures outcomes, the causal framework was built using the sequential variants of identifiability assumptions, including exchangeability, positivity, and consistency.
• We generated an ensemble of hypothetical predictions using the SuperLearner based on the following base learners: (i) least absolute shrinkage and selection operator, (ii) ridge regression, (iii) elastic net, (iv) random forest, (v) gradient boosting machines, (vi) neural network.
• Each statistical learning algorithm was fitted using the sum of eight elements of diabetes provision within a calendar year:
• We noticed an adequate prevalence of diabetes provision was consistent across the entire study period (for three consecutive years) among patients who did not receive a prescription for Metformin, Sulfonylurea and SGLT-2.
• Any combination of prescriptions related to glucose-lowering medications led to improved prevalence of adequate diabetes provision in next calendar year.
• The adequate prevalence of diabetes provision was consistently lower (for three consecutive years) among patients who did not receive a prescription for Metformin, Sulfonylurea and SGLT-2.

Results

Figure 3: Risk difference with 95% confidence interval using the SuperLearner prediction in test sample

Interpretation

• We noticed an improvement in diabetes provision with respect to increase in age groups (with the exception for 80+ years).
• Male patients tend to receive improved diabetes care with higher prevalence than female patients for each calendar year.
• The adequate prevalence of diabetes provision was consistently lower (for three consecutive years) among patients who did not receive a prescription for Metformin, Sulfonylurea and SGLT-2.

Material

• We demonstrated the application of hypothetical dynamic treatment regimes using the National Diabetes Action Canada Repository in which we applied a collection of mainstream statistical learning algorithms.
• The NDR collates electronic health records (EHRs) on diabetes patients across multiple practice-based research networks (PBRNs) located in Alberta, Manitoba, Quebec, Ontario, and Newfoundland.
• As of 2020/02, the NDR collects information on 148,767 diabetes patients distributed across 1,342 primary care providers with 145,958 age and sex matched controls (i.e. non-diabetics) for comparative research. The EHR in NDR contains patient-level demographics, medical diagnosis, procedures, medications, immunization, laboratory test results, vital signs and risk factors.
• The research ethics board (REB) application for this project was approved by North York General Hospital (REB #20-0012). Analyses were performed using the R software (v.4.1.0) in Secure Analytic Virtual Environment at the Centre for Advanced Computing Located at Queen's University.

References