Adjusting for Healthcare Utilization Improves the Performance of Self-Controlled Case Series Studies using Electronic Health Records

Undina Gisladottir[1], Nicholas Tatonetti[1,2,3]

Background

Adverse drug events (ADEs) can cause severe morbidities, hospitalization, and in some cases even death[1]. Randomized control trials are the gold standard for determining drug safety and efficacy[2]. However, the strict exclusion criteria for such trials can lead to unforeseen adverse drug events (ADEs) when released to the general public. The use of real-world data, such as electronic health records (EHR) data, allows for the active surveillance of the safety of drugs on the market in the population they are used clinically[3]. A self-controlled case series study using observational data, such as EHR, is an attractive approach to identifying ADEs, as it controls for time-invariant confounders such as sex and race/ethnicity[4]. However, ascertainment bias in the EHR leads to inherent differences between the 'risk' and 'baseline' periods, which results in a greater number of false positives. To counter this, some groups use negative controls to adjust the relative risk but this requires expert knowledge, is time-consuming, and fails to directly address the ascertainment bias that may disproportionately affect the study drugs compared to the negative controls.

Methods

In this study, we used the electronic health records from the New York-Presbyterian (NYP) Columbia University Irving Medical Center (CUIMC) data warehouse for patients whose first interaction with the hospital was after January 1, 2004. We used a reference set of adverse drug events (ADEs) that focuses on four outcomes: acute kidney injury (AKI), acute liver injury (ALI), acute myocardial infarction (AMI), and gastrointestinal bleeds (GI) and contains 399 drug-event pairs annotated as a positive or negative control[5]. We calculated an interval-specific utilizations score to adjust for bias between 'risk' and 'baseline' periods using the patient's number of visits, diagnosis, drug prescriptions etc. We used a logistic regression with an elastic net penalty to create the utilization scores. Specifically, we modelled risk vs. baseline period using the features listed above that capture the patient's healthcare utilization. To evaluate this score we estimated the drug effect on known drug-event pairs from a reference adverse drug event (ADE) set using EHR data and claims data to replicate our findings. We compared the performance of the conditional poisson regression model that included the interval-specific utilization score as a covariate to one that did not. We evaluated the performance of the models using four metrics: AUROC, coverage, bias, and mean squared error using 1,000 bootstrap iterations to produce 95% confidence intervals.

Results

After excluding drug-event pairs that had fewer than 10 patients, we were able to apply our prediction task to 224 drug-event pairs using the New York-Presbyterian (NYP) Columbia University Irving Medical Center (CUIMC) cohort. When comparing the performance of the traditional SCCS model to one that included the utilization score as a covariate, our method reduces bias and increases coverage while maintaining the same performance (according to AUROC and MSE).

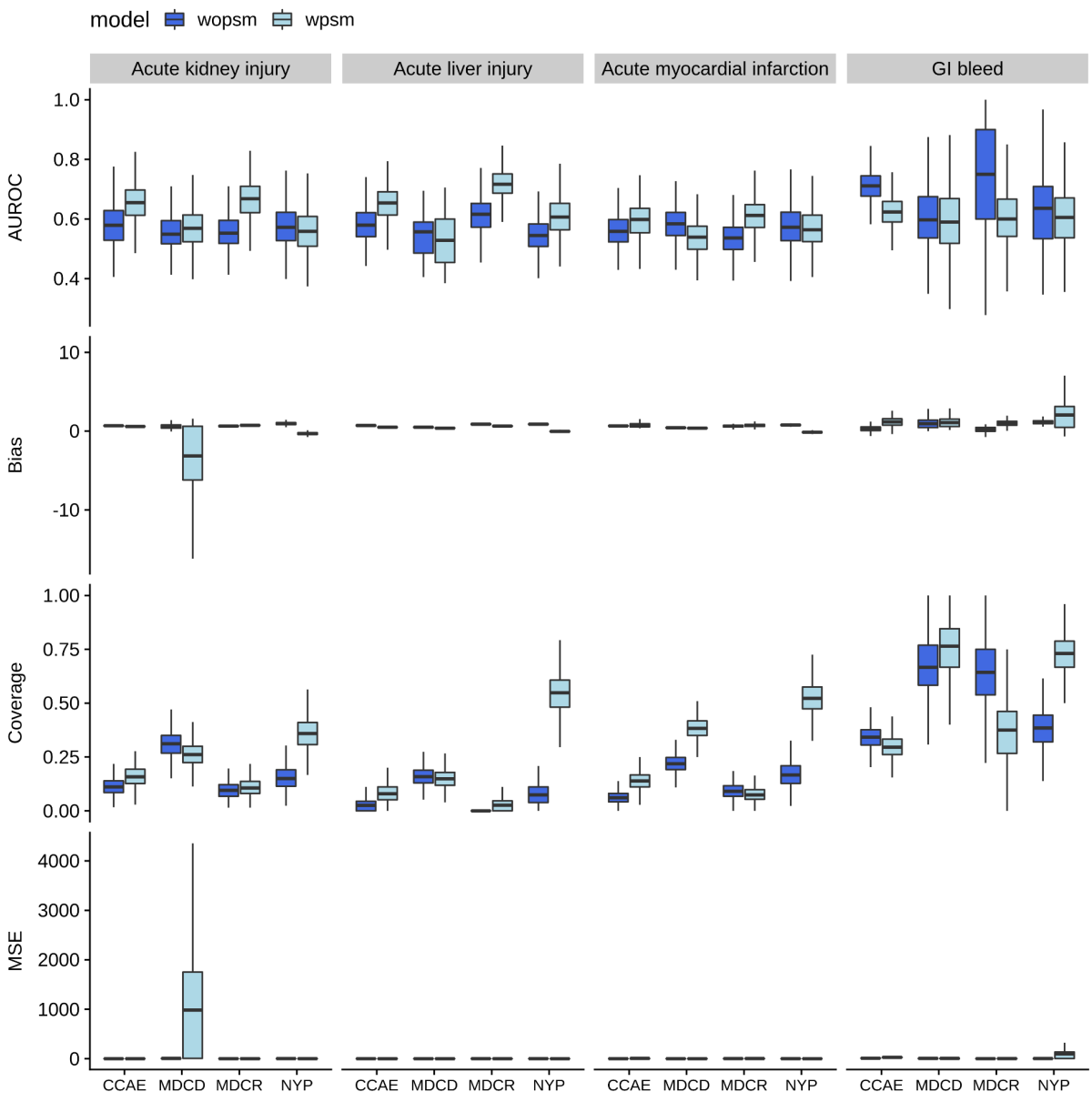

Figure 1. Results using NYP/CUIMC data. The results of the prediction task of positive and negative controls from an ADE reference set, that includes four events of interest acute kidney injury (AKI), acute liver injury (ALI), acute myocardial infarction (AMI), and gastrointestinal bleed (GI), using MarketScan data (commercial (CCAE), medicaid (MDCD), and medicare (MDCR)) and NYP CUIMC EHR data (NYP). To evaluate the performance of the models we used all drug-event pairs to calculate the AUROC, but only negative controls to measure the bias, coverage, and mean squared error (MSE). We see similar AUROC and MSE but an increase in coverage for all adverse events and a decrease in bias when comparing models that included the utilization score as a covariate (light blue) to those that did not (dark blue). This performance improvement is most apparent for the electronic health records (NYP).

Conclusion

In this study, we presented a novel method of including an interval-specific utilization score, which

captures the patient's healthcare utilization, as a covariate in the SCCS model to adjust for bias between risk and baseline periods. Our results show that adjusting for healthcare utilization may reduce bias in SCCS studies when using observational data, enabling more reliable drug safety estimates.

References/Citations (see: www.icmje.org/index.html)

1. Sahilu T, Getachew M, Melaku T, Sheleme T. Adverse drug events and contributing factors among hospitalized adult patients at jimma medical center, Southwest Ethiopia: A Prospective Observational Study. Curr Ther Res Clin Exp. 2020 Oct 29;93:100611.
2. Kabisch M, Ruckes C, Seibert-Grafe M, Blettner M. Randomized controlled trials. Dtsch Ärztebl Int. 2011 Sep;108(39):663–8.
3. Huang YL, Moon J, Segal JB. A Comparison of active adverse event surveillance systems worldwide. Drug Saf. 2014 Aug 1;37(8):581–96.
4. Suchard MA, Zorych I, Simpson SE, Schuemie MJ, Ryan PB, Madigan D. Empirical performance of the self-controlled case series design: lessons for developing a risk identification and analysis system. Drug Saf. 2013 Oct;36(S1):83–93.
5. Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. defining a reference set to support methodological research in drug safety. Drug Saf. 2013 Oct;36(S1):33–47.