Identifying Causal Effects of Medication Exposures from Electronic Health Record Data with i2b2

Victor M. Castro, MS[1], Michael C. Ollendieck[1], MD, Shawn N. Murphy, MD, PhD[2]
[1]Partners Healthcare, Somerville, MA; [2]Massachusetts General Hospital, Boston, MA

Abstract

We describe a method for inferring causal drug effects (both adverse and beneficial) from large observational real-world EHR datasets using a case-control design applied with coarsened exact matching. Using various parameters, we demonstrate the ability to detect true causal associations with osteoporosis by looking across 900+ medications. These tools are implemented within the i2b2 framework and made available as R packages.

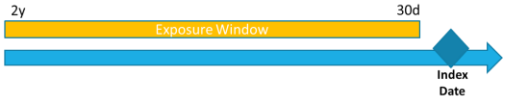
Introduction

Drug safety and efficacy surveillance is a major focus of drug regulatory authorities in many countries. While spontaneous adverse event reporting systems provide a means to directly capture drug side effects, they suffer from reporting bias and do not capture non-acute events[1]. Electronic health record (EHR) data offers a real-world setting to identify unanticipated side effects in large populations. However, a major limitation of observational EHR data is confounding by unmeasured observations. In this work, we describe a method for inferring causal drug effects (both adverse and beneficial) from large observational EHR datasets using a case-control design applied with coarsened exact matching (cem). We apply the method to detect drugs with causal links to osteoporosis by looking across 900+ medications. Using matching and adjusting for utilization measures we can mitigate confounding effects.

Methods

Case-control Design

The case-control design has been used in retrospective observational studies to compare patients with a given condition to patients without the condition by looking at their medical history preceding the onset of the condition. In this design, we identify the index event in patients and define an exposure period prior to the event to determine if they were exposed to a drug of interest during that time. Figure 1 below provides a conceptual overview of the design.



The index date for cases is defined as the first event of the condition in their medical record. Control pool patients are defined as patients without the condition of interest and with a visit on the same year of the case index date. Each case is assigned a set of eligible controls that can be matched against based on the presence of a visit on the same year.

Matching: We utilized coarsened exact matching, implemented in the cem package in R, to match controls to cases on patients age at index date, gender, race and year of index event[2]. CEM provides exact matching on categorical variables (gender and race in this case) and coarsens continuous variables to find suitable matches.

Exposure window: The exposure window is defined in days prior to the index event. The method allows various settings with a default of 2 years to 30 days prior to the index event. Patients without any visits in the exposure windows are excluded from the analysis. In addition, the count of visits (log-transformed) within the exposure window is specified as a covariate in effect estimates.

Effect Estimates: Estimate of drug effects are computed by 3 methods: 1) unadjusted risk ration (RR); 2) logistic regression adjusted for count of visits in the exposure window and year of index date and 3) conditional logistic regression with further adjustment for matching strata.

Experimental Setting

Population/ Data: We retrieved all EHR data of 69,121 patients consented in the Partners Healthcare Biobank made available in an i2b2 datamart[3]. The Biobank population is a subset of the overall Partners population that will ultimately be linked with genomic and other data sources that may further be integrated in the current methods.

Outcome of interest: We selected osteoporosis as the outcome of interest defined as an first recorded ICD-10 code of M80* and M81*.

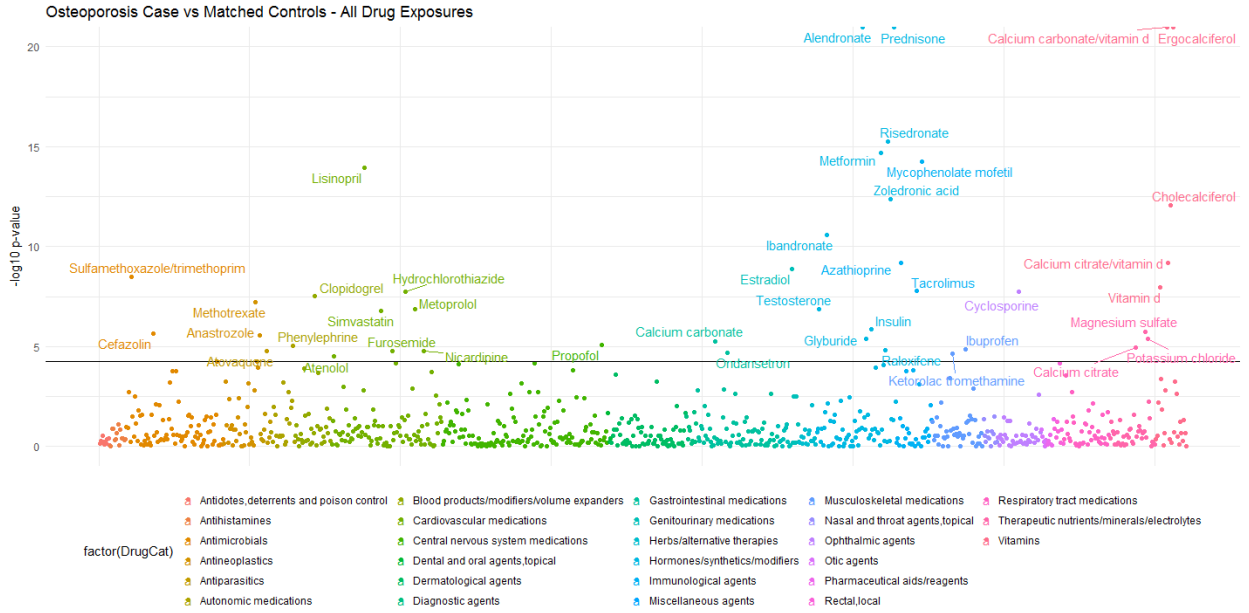
Drugs of interest: We looked across all RxNorm drug ingredients and combinations. Drugs prescribed to fewer than 100 patients were excluded from the analysis. A total of 901 medications were evaluated.

Results

We identified 7,119 patients with an index osteoporosis in the Partners Biobank. These patients were matched to 14,570 controls with no mention of osteoporosis in their medical record. Figure 3 is a plot of each of 901 drugs and their association to osteoporosis onset after exposure based on conditional regression. Drugs such as prednisone with known side effects of osteoporosis are identified. Significant drugs with reduced odds ratio include metformin. Adjusting for the count of visits significantly reduces false drug-outcome associations.

Medications indicated for osteoporosis are also identified even though they occur prior to the onset of the outcome by 30 days. This indicates some level of signal bias still exists potentially caused missing data in the EHR or inaccurate definitions of the outcome definition.

The methods have been implemented as 2 R packages: Ri2b2matchcontrols and Ri2b2casecontrol and will be made



publicly available after further testing.

Figure 1. Drugs associated with osteoporosis identified with the Ri2b2casecontrol method. The y-axis represents -log10 p-value from a conditional logistic regression analysis. Each point corresponds to a medication. Colors represents class of medication.

Conclusion

A case-control approach applied to large EHR datasets can identify true causal effects of drug exposures with the aim of monitoring the safety of medications and identifying candidates for drug repurposing. The Ri2b2casecontrol tools implements the methods in an i2b2 framework. Future efforts will be aimed at minimizing bias due to missing data and inaccurate outcome definitions. We also aim to test the tools in larger datasets and additional outcomes.

References

1. Madigan D, Stang PE, Berlin JA, Schuemie M,…Ryan PB. A systematic statistical approach to evaluating evidence from observational studies. Annual Review of Statistics and Its Application. 2014 Jan 3;1:11-39.
2. Gainer VS, Cagan A, Castro VM, Duey S,…Murphy SN. The Biobank Portal for Partners personalized medicine: a query tool for working with consented biobank samples, genotypes, and phenotypes using i2b2. Journal of personalized medicine. 2016 Feb 26;6(1):11.
3. Iacus SM, King G, Porro G. CEM: Coarsened exact matching software. Journal of statistical Software. 2009;30(9):1-27.