

Assessing Negative Control Exposure-Outcome Pair Selection Strategies on a Replication Study

Erica A. Voss, MPH ^{1,2,3}, Martijn J. Schuemie PhD ^{2,3}, Johan van der Lei MD, PhD ^{1,2},
Peter R. Rijnbeek, PhD ^{1,2}

1 - Erasmus University Medical Center, Rotterdam, Netherlands,

2 - OHDSI Collaborators, Observational Health Data Sciences and Informatics (OHDSI), New York, NY,

3 - Janssen Research and Development, LLC, Raritan, NJ

Is this the first time you have submitted your work to be displayed at any OHDSI Symposium?

Yes _____ No X (I have submitted posters in the past, but not of this work)

Abstract

Negative controls, exposure-outcome pairs where no causal relationship is believed to exist, can be used to evaluate whether causal effect estimation methods are unbiased, and subsequently perform empirical calibration. Manual curation of negative controls is difficult. The knowledge base called the Common Evidence Model helps with the selection of negative controls by enabling several selection strategies using its curated evidence. The strategies range in difficulty to implement. This work assesses the different strategies for the selection of negative controls by applying them to a typical effect estimation study. We observed little impact on estimated residual bias and empirical calibration using the different strategies, meaning the simpler of the strategies may be used. This knowledge yields time savings as the simpler of the strategies takes less time to implement and reduces complexity of selecting negative controls as it is an easier process.

Introduction

A notable contribution the Observational Health Data Sciences and Informatics (OHDSI) collaboration has made to observational research is the empirical calibration of p-values [1]. Empirical calibration uses negative controls, which are exposures-outcome pairs thought to have no known causal relationship. In empirical calibration the study design used to answer the question of interest is also used to estimate effects for a set of negative controls. These estimates give “an indication of what can be expected when the null hypothesis [(usually the hypothesis of no effect)] is true, and we use them to estimate an empirical null distribution” [1]. OHDSI now recommends that “observational studies always include negative controls to derive an empirical null distribution and use these to compute calibrated p-values” [1].

Knowing which exposure-outcome pairs make good negative controls is not easy. The initial method for finding negative controls was to manually generate a list using evidence from drug product labels, the book *Drug-Induced Diseases: Prevention, Detection and Management* [2], and medical literature found in PubMed. It is prohibitive to curate evidence across these sources when they exist in separate locations and use different terminologies.

In response OHDSI developed a knowledge base called the Common Evidence Model (CEM) [3]. Building on prior work [4, 5], CEM curates and standardizes publicly available information on exposures and outcomes from product labels, published literature, and spontaneous reports. The standardization is both in structure and terminologies. Once evidence is brought together in CEM, processes are needed to determine which exposure-outcome pairs make proper negative controls. It is currently unknown what the impact of different selection strategies are on the empirical calibration. This research will apply the negative control selection to a previously performed study with the goal of understanding if the different processes change the interpretation of the study.

Methods

Five negative control selection strategies are evaluated. The strategies are listed in order of implementation complexity starting with the simplest.

Strategy 1 – All Relevant Exposure-Outcome Pairs

Strategy 2 – Exclude Pairs with CEM Evidence (Exact Outcome Terms Only) [3]

Strategy 3 – Exclude Pairs with CEM Evidence (with Associated Related Outcomes) [3]

Strategy 4 – Automated Method of Pair Selection using CEM Evidence [4]

Strategy 5 – Automated Method of Pair Selection using CEM Evidence [4] with Manual Curation

Strategy 1 leverages observational data to find all possible outcomes that occur for the first time after the exposure of interest. This provides a list of possible exposure-outcome pairs and no additional filtering takes place. Since this strategy yields many pairs, we take a random sample of 1,000 in our evaluation. Strategy 2 has five steps for filtering pairs: (a) limiting to prevalent exposure-outcome pairs whose terms exist in CEM, (b) filtering to preferred terms (i.e. excluding terms that are too broad, already suggesting an exposure-outcome relationship, and associated to pregnancy), (c) excluding outcomes that are indicated or contraindicated for an exposure, (d) eliminating pairs when evidence in CEM maps exactly to the outcome term, and (e) of the outcome terms left those that are in the same hierarchy only keeping the top level term. Strategy 3 follows the same approach as Strategy 2, however this strategy leverages the Observational Medical Outcomes Partnership (OMOP) Vocabulary to associate hierarchically related CEM evidence to the outcome term [6]. Strategy 4 has five steps for filtering pairs: (a) limiting to prevalent exposure-outcome pairs whose terms exist in CEM, (b) filtering to preferred terms, (c) excluding outcomes that are indicated or contraindicated for an exposure, (d) using CEM evidence to build a classification model that is then used to select exposure outcome pairs not in an adverse event relationship [6], and (e) of the outcome terms left those that are in the same hierarchy only keeping the top level term. Finally, Strategy 5 is manual curation of the list prepared in Strategy 4. This was performed independently by two physicians, and when there was initial disagreement the physicians reached a consensus.

To evaluate the impact of the different negative control selection strategies, we replicate a study by Graham et al. performed in a Medicare population [7]. Specifically, Graham et al. assessed dabigatran versus warfarin exposure with outcome risk of major gastrointestinal (GI) bleed. This was a new user cohort design with propensity score adjustment to handle potential confounding between exposure and outcome. A hazard ratio of 1.28 with a 95% confidence interval of 1.14-1.44 was reported by Graham et al. [7]. Schuemie et al. replicated this new user cohort design study and found an uncalibrated similar hazard ratio. Even with calibration, the significance of the result did not change [8]. Similarly, to Schuemie et al. we replicate Graham’s study, and use the five strategies to select the negative controls to use for empirical calibration. This work will help us understand the impact of different negative control selection strategies on empirical calibration.

Results

Table 1 shows the number of negative controls selected by the different strategies on all possible exposure-outcome pairs.

Table 1 – Number of Negative Controls Pairs Selected for the Graham et al. Study Across Five Selection Strategies	
Strategy	Number of Pairs
Strategy 1 – All Relevant Exposure-Outcome Pairs	23,960 (1,000 sample taken)
Strategy 2 – Exclude Pairs with CEM Evidence (Exact Outcome Terms Only)	722
Strategy 3 – Exclude Pairs with CEM Evidence (with Associated Related Outcomes)	690
Strategy 4 – Automated Method of Pair Selection using CEM	690
Strategy 5 – Automated Method of Pair Selection using CEM Evidence with Manual Curation	100

Strategy 2 through Strategy 4 yield similar calibration results when applied to the Graham study. Strategy 5 is still undergoing physician review and may update our current conclusions.

Conclusion

In the replication of the Graham et al. study we found that our main selection strategies of interest (Strategy 2-4) performed similarly. It has yet to be seen what manual curation does. Since the strategies increase in complexity to implement this suggests the simpler methods can be used. This work only reviewed replication of one study and further evaluation is needed to know if this finding holds true across studies. However, if this holds true across studies this knowledge yields time savings as the simpler of the strategies takes less time to implement and reduces complexity of selecting negative controls as it is an easier process.

References

1. Schuemie, M.J., et al., *Interpreting observational studies: why empirical calibration is needed to correct p-values*. Stat Med, 2014. **33**(2): p. 209-18.
2. Tisdale, J.E. and D.A. Miller, *Drug-induced diseases: prevention, detection, and management*. 2010: ASHP.
3. *Common Evidence Model (CEM)*. [web page] 2018.12.13]; Available from: <https://github.com/OHDSI/CommonEvidenceModel>.
4. Voss, E.A., et al., *Accuracy of an automated knowledge base for identifying drug adverse reactions*. J Biomed Inform, 2017. **66**: p. 72-81.
5. *Large-scale adverse effects related to treatment evidence standardization (LAERTES): an open scalable system for linking pharmacovigilance evidence sources with clinical data*. J Biomed Semantics, 2017. **8**(1): p. 11.
6. *OMOP Common Data Model (CDM)*. [web page]; Available from: <https://github.com/OHDSI/CommonDataModel>.
7. Graham, D.J., et al., *Cardiovascular, bleeding, and mortality risks in elderly Medicare patients treated with dabigatran or warfarin for nonvalvular atrial fibrillation*. Circulation, 2015. **131**(2): p. 157-64.
8. Schuemie, M.J., et al., *Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data*. Proc Natl Acad Sci U S A, 2018. **115**(11): p. 2571-2577.