

Evaluating automated selection of outcome negative controls for estimating the total error distribution of self-controlled cohort study designs

James P. Gilbert, Rachel E. Teneralli, Erica A. Voss, Christopher A. Knoll, David M. Kern, M Soledad Cepeda, Patrick B. Ryan

Background

The use of negative controls for empirical calibration of p-values and confidence intervals (CIs) of effect estimates is an important tool to improve the quality, reproducibility, and robustness of population effect estimation studies (1,2). Negative controls are outcomes or exposures, related to an effect estimate of interest, that are considered to lack a causal relationship to one another, indicating that any observed effects are a product of systematic error and not due to a causal relationship (3). However, sets of negative controls are challenging to create requiring significant manual curation, a rate limiting factor in high-throughput studies conducted previously (4–6). In this work, we evaluated a fully automated method of generating outcome controls using the, previously presented, Common Evidence Model (CEM) (7) applied to a series of self-controlled cohort (SCC) designs (8) across all outcomes and exposures to 4 US healthcare datasets as part of the REWARD framework (4–6).

Methods

We applied the data in CEM to build sets of exposure controls using evidence from spontaneous reports, drug labels and literature to build fully automated null distributions. We evaluated the use of outcome negative controls against 4 Observational Health Data Sciences and Informatics (OHDSI) gold standard benchmark drug exposures (diclofenac, ciprofloxacin, metformin and sertraline) with 25 manually curated negative controls (8). Using a new-user cohort design and self-controlled cohort study we generated effect estimates for all outcome cohorts, as defined as SNOMED conditions with at least two diagnosis codes recorded. We applied this to 4 US medical insurance claims databases conforming to the Observational Medical Outcomes Partnership (OMOP) Common Data Model. We measured the concordance between error distributions by comparing the absolute systematic error, which summarises the mean and variance of fitted null distributions. Using the manually calibrated effect estimates for all outcomes exposed to the reference drugs as a gold standard positive set, we compared all outcomes with significant effects ($p \leq 0.05$ and a CI that does not include 1.0) to measure the false discovery rate (FDR), sensitivity and specificity of effect estimates calibrated with automated controls.

Results

The manual and automated negative controls show similarity with the resulting error distributions across all 4 databases. See Figure 1, left for diclofenac and right for expected absolute systematic error of all distributions. Comparing the automated null distributions against a baseline set of uncalibrated p-values and CIs, across all 4 databases and exposures, we observed a mean FDR of 6.76% (standard deviation [SD] $\pm 8.32\%$), sensitivity of 0.57 (SD ± 0.013) and specificity of 0.99 (SD ± 0.106) for auto null distributions compared to mean FDR of 60.56% (SD $\pm 15.02\%$) sensitivity of 0.90 (SD ± 0.107) and specificity of 0.71 (SD ± 0.106) for uncalibrated effect estimates.

Figure 2 shows the full error distribution for all Rx Norm exposures captured within the dataset across all diseases, as defined by standard SNOMED concept codes. We note that the extremes of this distribution indicate that bias may be large for some exposures.

Conclusion

The method of automated negative control selection appears to have strong concordance with the gold standard outcome controls sets, indicating their applicability. The low FDR, when compared with uncalibrated effect estimates, indicates that the method has good concordance with a manual approach to filtering negative controls and helps filter out a significant amount of noise that is largely due to residual bias, a major problem when applying uncalibrated effect estimates at scale.

However, the low number of significant results indicates that the sensitivity of the automated approach may be improved. In addition to improved mappings between drug and condition concepts, methods such as ranking negative controls based on co-occurrence of exposure and outcomes across data domains may help to further increase the sensitivity of automated approaches such as the method outlined here

The analysis total error distribution also indicates that this approach can be applied to produce more reliable effect estimates at scale, resulting in fewer false positives. Crucially, it appears that this process allows for reasonable estimates of systematic bias without the painstaking process of curating negative controls by hand, paving the way for high throughput population level effect estimate studies utilising observational data.

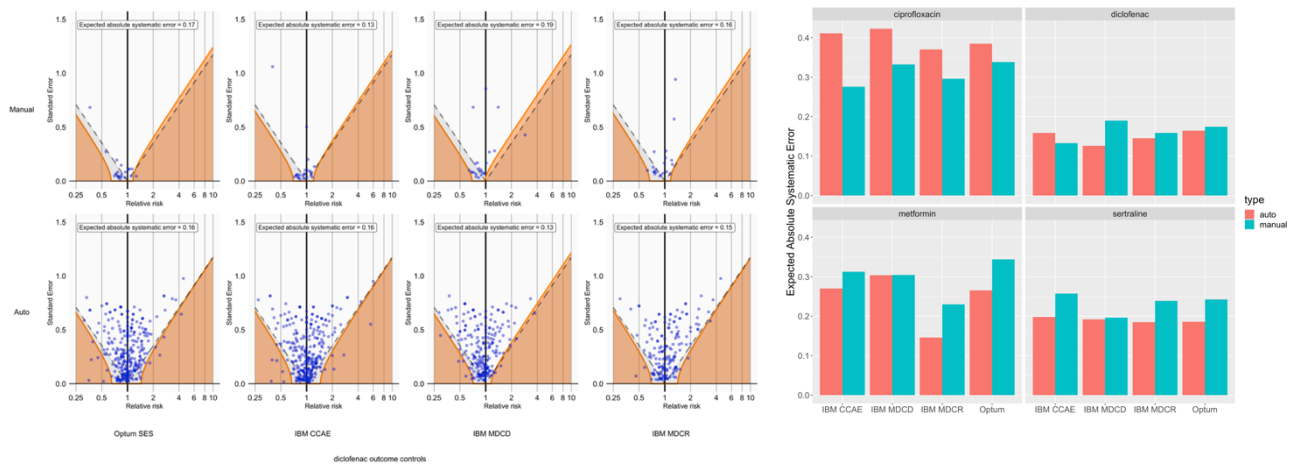


Figure 1. (left) Calibration plots of the exposure diclofenac with many negative control outcomes across 4 US claims databases using the SCC study design for patients exposed to diclofenac. We compare automatic controls (bottom) to the gold standard set (top). Estimates below the dashed line (gray area) have $p < 0.05$ using traditional p-value calculation. Estimates in the orange areas have $p < 0.05$ using the calibrated p-value calculation and the red band indicates the 95% credible interval of the boundary of the orange area. Blue dots indicate negative controls. The shown absolute systematic error summarises the computed null distributions in a single statistic that captures the mean and variance of the distributions. **(right)** Observed expected systematic error for controls across all 4 exposures in automated and manually selected outcome control sets.

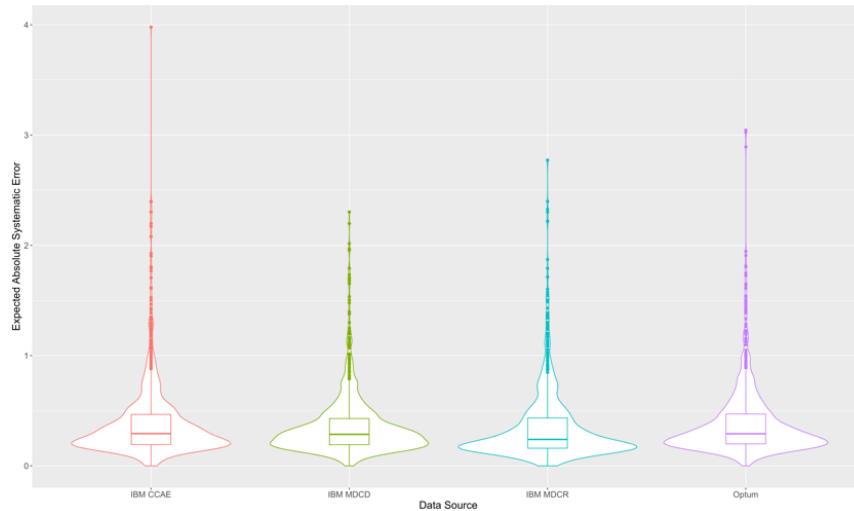


Figure 2. Violin plots of the distribution of expected absolute systematic error across null distributions generated from outcome controls for all exposures in each data source. Though 2 standard deviations of error are below 0.5 for all the datasets, in some cases systematic error appears to be extreme.

Exposure Concept	OMOP Concept Id	N. automated controls
RxNorm – Sertraline	739138	3522
RxNorm – Diclofenac	1124300	2921
RxNorm – Metformin	1503297	1836
RxNorm – ciprofloxacin	1797513	2474

Table 1. Exposure concepts and the number of outcome concepts used as controls in the study. Control concepts were selected from the common evidence model. Each concept was treated as an incident outcome, only patients with 2 diagnosis codes within 365 days included.

References

- Schuemie MJ, Hripcsak G, Ryan PB, Madigan D, Suchard MA. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proc Natl Acad Sci U S A*. 2018 Mar 13;115(11):2571–7.
- Schuemie MJ, Ryan PB, Hripcsak G, Madigan D, Suchard MA. Improving reproducibility by using high-throughput observational studies with empirical calibration. *Philos Trans R Soc A Math Phys Eng Sci*. 2018;376(2128).
- Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative Controls: A tool for detecting confounding and bias in observational studies. *Epidemiology* [Internet]. 2010 May [cited 2020 May 14];21(3):383–8. Available from: <http://journals.lww.com/00001648-201005000-00017>
- Teneralli RE, Kern DM, Cepeda MS, Gilbert JP, Drevets WC. Exploring Real-World Evidence to Uncover Unknown Drug Benefits and Support the Discovery of New Treatment Targets for Depressive and Bipolar Disorders Exploring Real-World Evidence to Uncover Unknown Drug Benefits and Support the Discovery of New Treatment Targets for. *J Affect Disord* [Internet]. 2021 [cited 2021 May 12]; Available from: <https://doi.org/10.1016/j.jad.2021.04.096>
- Kern DM, Cepeda MS, Lovestone S, Seabrook GR. Aiding the discovery of new treatments for dementia by uncovering unknown benefits of existing medications. *Alzheimer’s Dement Transl Res Clin Interv*. 2019 Jan 1;5:862–70.
- Cepeda MS, Kern DM, Seabrook GR, Lovestone S. Comprehensive Real-World Assessment of Marketed Medications to Guide Parkinson’s Drug Discovery. *Clin Drug Investig*. 2019 Nov 1;39(11):1067–75.
- Voss EA, Boyce RD, Ryan PB, van der Lei J, Rijnbeek PR, Schuemie MJ. Accuracy of an automated knowledge base for identifying drug adverse reactions. *J Biomed Inform*. 2017 Feb 1;66:72–81.
- J. Schuemie M, Soledad Cepede M, A. Suchard M, Yang J, Tian Alejandro Schuler Y, B. Ryan P, et al. How Confident Are We About Observational Findings in Health Care: A Benchmark Study. *Harvard Data Sci Rev*. 2020 Jan 31;2(1).