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, Raritan, NJ

Background

- **Negative controls (NC)** are drug exposures / outcome pairs with no known causal relationship that can be used as a bias diagnostic tool in observational studies [1]
- NCs can be used to **calibrate p-values** accounting for random and systematic error [1]
- **Manual identification of NCs is labor intensive** [2] necessitating automated strategies (e.g. the use of Common Evidence Model (CEM)) which reduce the effort in finding NC
- We apply **NC selection strategies to a previously performed study** to understand their impact on the study results

Methods

- The study selected for replication is a **dabigatran versus warfarin exposure with a risk outcome of major gastrointestinal (GI) bleed** by Graham et. al. [3] – empirical calibration was added using NCs
- Performed in IBM MarketScan® **Medicare** Supplemental Database (01/2000-10/2016)
- **Common Evidence Model (CEM)** curates and standardizes publicly available information on exposures and outcomes from product labels, published literature, and spontaneous reports
- **NC Selection Strategies** (in order of complexity):
  - Strategy 1 – All Available Prevalent Outcomes *
  - Strategy 2 – Exclude Outcomes with CEM Evidence (Exact Outcome Terms Only)
  - Strategy 3 – Exclude Outcomes with CEM Evidence (with Associated Related Outcomes)
  - Strategy 4 – Automated Method of Outcome Selection using CEM Evidence [2]
  - Strategy 5 – Automated Method of Outcome Selection using CEM Evidence with Manual Curation ‡

* Outcomes that occur for the first time after exposure to dabigatran or warfarin
‡ Manual curation will be performed by two physicians independently

Results

Table 1 – Number of Negative Control Outcomes Selected for the Graham et al. Study Across Five Selection Strategies

Strategy	Number of Outcomes	Outcomes In Calibration
Strategy 1 – All Available Prevalent Outcomes	23,960 (1,000 sample taken)	451
Strategy 2 – Exclude Outcomes with CEM Evidence (Exact Outcome Terms Only)	722	430
Strategy 3 – Exclude Outcomes with CEM Evidence (with Associated Related Outcomes)	690	402
Strategy 4 – Automated Method of Outcome Selection using CEM	690	402
Strategy 5 – Automated Method of Outcome Selection using CEM Evidence with Manual Curation	113	105

- Outcomes are not used in calibration if they never occurred after exposure to dabigatran or warfarin (**Table 1**)
- Strategy 1 sample had <5% overlap with other strategies
- Strategy 3 and 4 completely overlap; CEM data summary and model selected same outcomes

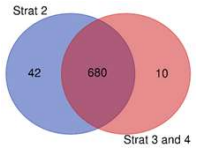

Figure 1 – Comparing Overlap between Number of Outcomes in Strategy 2 and Strategy 3/4

- Strategy 2 and 3/4 have large overlap (**Figure 1**):
  - 42 only in Strategy 2 as Strategy 3/4 eliminated them using related evidence
  - 10 only in Strategy 3/4 due to concept optimization
- In Strategy 5, 179 codes reviewed, physicians agreed 120 times, left with 113 negative controls
  - Ex) "Orthostatic hypotension" (OH) eliminated as bleeding is a well established adverse event for drugs like dabigatran/warfarin and bleeding could lead to OH. CEM did not have evidence to eliminate this outcome for consideration as a NC.

Results Continued

- Comparing higher differenced Strategies (1, 3, 5) while Strategy 1 and Strategy 3 have a <5% overlap in concepts, they seem to be characterizing similar bias (**Figure 2**)

Table 2 – Characteristics of the original Graham study and our replication (before calibration)

Study	Number of Exposures		Number of Events		Risk Ratio (CI) (uncalibrated)	P-value (uncalibrated)
	Dabigatran	Warfarin	Dabigatran	Warfarin		
Graham	67,207	67,207	623	513	1.28 (1.14-1.44)	<0.001
Our Replication	16,734	16,734	210	167	1.15 (0.94-1.41)	0.17

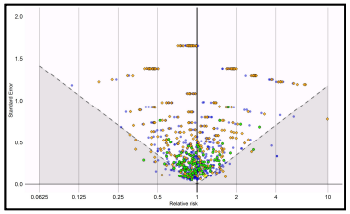

Figure 2 – Plot of Strategy 1, 3, and 5's Negative Controls Blue Circle = Strategy 1, Orange Diamond = Strategy 3, Green Square = Strategy 5

Table 3 – Comparing Calibrated p-values Across Negative Control Strategies

Study	Calibrated p-value (CI)
Strategy 1	0.24 (0.15-0.35)
Strategy 2	0.17 (0.09-0.28)
Strategy 3	0.18 (0.09-0.28)
Strategy 4	0.18 (0.10-0.29)
Strategy 5	0.03 (0.01-0.13)

- Graham study had larger patient population (**Table 2**)
- Automated strategies using CEM (Strategy 2 through 4) perform similarly when calibrating the p-value (**Table 3, Figure 3**)
- Strategy 5 comes to the same interpretation as the original study (increased risk of GI bleed for dabigatran) (**Figure 4**)

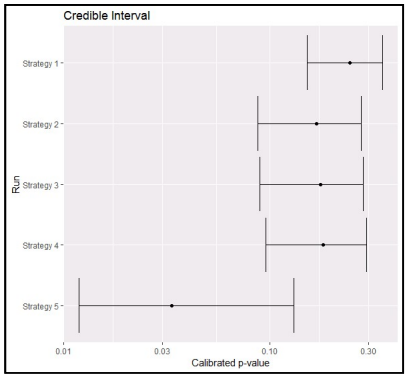

Figure 3 – the Credible Interval of the calibrated p-values

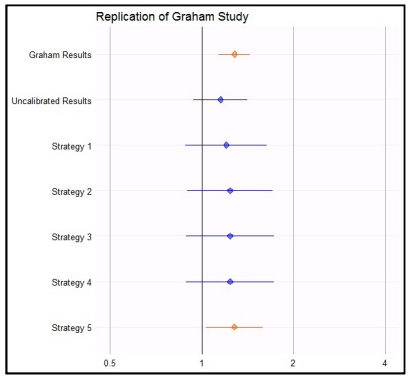

Figure 4 – Risk Ratios with Confidence Intervals for risk of major bleed with dabigatran vs. warfarin. Strategies implement calibrated results

Conclusions

- Automated methods seem to perform similarly
- The more controls you have the more certain you are in the p-value estimate
- There are relationships between exposure and outcomes that CEM cannot detect, manual curation is still recommended

[1] Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p-values. Stat Med. 2014 Jan 30;33(2):209-18. doi: 10.1002/sim.5925. Epub 2013 Jul 30. PubMed PMID: 23900808; PubMed Central PMCID: PMC4285234.
[2] Voss, E.A., et al., Accuracy of an automated knowledge base for identifying drug adverse reactions. J Biomed Inform, 2017. 66: p. 72-81.]
[3] 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.

Contact: evoss3@its.jnj.com