Name:	Anthony G. Sena
Affiliation:	Janssen Research and Development
Email:	asena5@its.jnj.com
Presentation type:	Collaborator Demonstration

Automated selection of Negative Control Exposure-Outcome pairs for use in Observational Studies: A capabilities demonstration

Anthony G. Sena[1,2], Richard D. Boyce, PhD[3,2], Patrick B. Ryan, PhD[1,2,4], Martijn J. Schuemie, PhD[1,2], Erica A. Voss, MPH[1,2]
[1]Janssen Research and Development, Raritan, NJ, [2]Observational Health Data Sciences and Informatics (OHDSI), New York, NY, [3]University of Pittsburgh, Pittsburgh, PA, [4]Columbia University, New York, NY

Abstract

Negative control exposure-outcome pairs should be more commonly employed in observational studies to reduce the potential for systematic error[1,2]. When employed, negative controls help to identify and calibrate for confounding as well as other sources of error. The process for identifying negative controls (either exposure controls or outcome controls) for use in observational studies requires time-consuming, manual curation of information from various knowledge sources. Here we demonstrate a means for automating the selection of candidate negative controls by leveraging an open-source standardized evidence base known as LAERTES (Large-scale Adverse Effects Related to Treatment Evidence Standardization) made available through the Observational Health Data Sciences and Informatics (OHDSI) collaborative.

Introduction

The LAERTES evidence base integrates several sources of evidence for investigating the association between drugs and health outcomes of interests (HOIs) into a single data source. Evidence is sourced from spontaneous reports, scientific literature, and both American and European product labeling. A statistical model was created using the evidence in LAERTES for discriminating between known positive drug-HOI causal relationships and drugs known to be unassociated with an outcome.

ATLAS is an open-source platform made available through OHDSI that provides a single user-experience for generating evidence. ATLAS has been integrated with LAERTES to provide a workflow for automatically identifying positive and negative controls.

Methods

The LAERTES evidence base contains information from spontaneous reports, scientific literature, and product labeling. Spontaneous reporting evidence is culled from the FDA Adverse Event Reporting System (FAERS) and includes counts of reports and proportional reporting ratio (PRR) scores[3,4]. Evidence from the scientific literature was generated through two methods: one leveraging Medical Subject Headings (MeSH) tags following the process described by Avillach et al.[5], and another (SemMedDB) that uses relationships semantically tagged within Medline abstracts natural language processing[6]. The results of these two methods are additionally stratified by Medline publication types: clinical trials, case report, and all other abstracts (i.e., of type Meta-Analysis, Comparative Study, Multicenter Study, or Journal Article). Finally, American product labels are parsed by a method developed by Duke et al.[8], and ADRs mentioned in European labels are provided by the PROTECT project[7].

The evidence in LAERTES was narrowed down to a "universe" of drugs and HOIs that met a specific, minimum criterion of evidence. The minimum criteria is defined as: the condition or ingredient had at least one FAERS evidence item, one Medline evidence item and one evidence item from a product label. This step was taken because a condition or drug that did not meet the minimum criteria might indicate a lack of clinical experience.

Logistic regression was used to build a multivariable prediction model on the LAERTES data that could discriminate between positive and negative controls. Regularization with a Laplace prior on the regression coefficients was used to allow the model to perform parameter selection. A union of the OMOP Reference Set [9] and the Exploring and Understanding Adverse Drug Reactions (EU-ADR) Reference Set [10-12] was used to train our classifier. The prediction accuracy was estimated using leave-pair-out cross-validation.

An automated process was developed to expose the prediction model along with the evidence "universe" to return positive and negative controls. The input to the algorithm is a list of OMOP concept ids that define either a drug ingredient or an HOI. When given a drug ingredient, the algorithm will return the full "universe" of HOIs with counts for each evidence source and the predicative values. It works in a similar fashion when given an HOI, the algorithm will return the same information for all drug ingredients in the "universe".

Results

Figures 1 provides an example of input to the negative controls algorithm while Figure 2 shows the output including the evidence counts, by source and the prediction values.

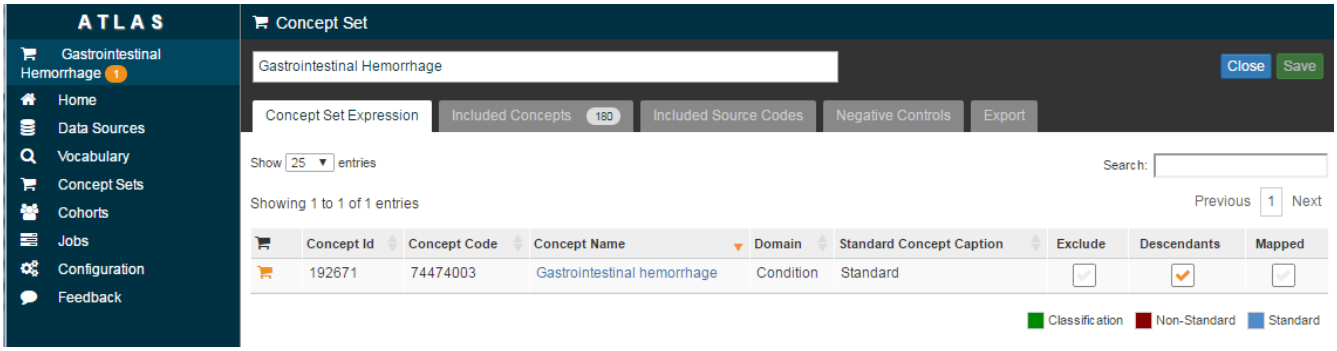

Figure 1. Input to the negative control algorithm (HOI: Gastrointestinal hemorrhage)

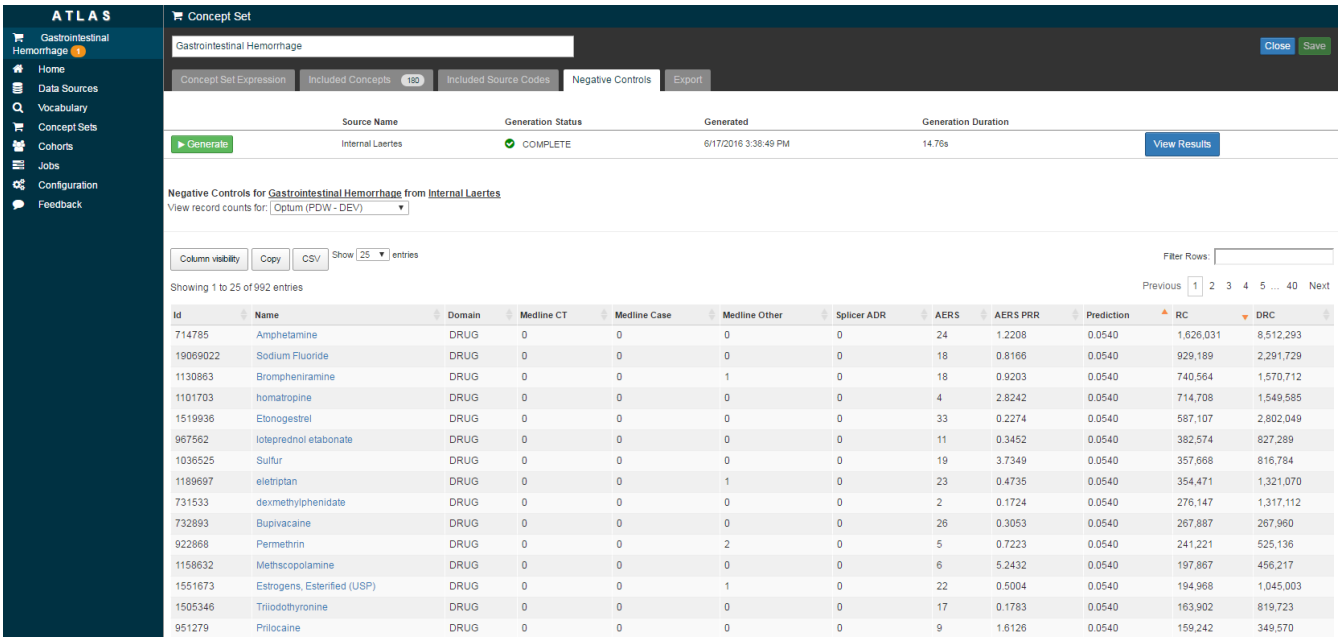

Figure 2. Output of the negative control algorithm: drug ingredients in the "universe", their evidence counts and predictive value.

Conclusion

Automating a list of candidate negative controls using LAERTES and ATLAS lowers the amount of manual effort required of researchers while implementing a study.

Acknowledgements

RDB received funding from the US National Institute on Aging (K01AG044433) and the National Library of Medicine (R01LM011838). The opinions expressed here are those of the authors and not necessarily those of the companies or institutions who employ them.

References

1. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 2010;21: 383–388.
2. 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; 33: 209–218.
3. (20) van Puijenbroek, E.P., et al., A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Pharmacoepidemiol Drug Saf, 2002. 11(1): p. 3-10.
4. (21) Banda, J.M., et al., A curated and standardized adverse drug event resource to accelerate drug safety research. Scientific Data, 2016. 3: p. 160026.
5. (22) Avillach, P., et al., Design and validation of an automated method to detect known adverse drug reactions in MEDLINE: a contribution from the EU-ADR project. J Am Med Inform Assoc, 2013. 20(3): p. 446-52.
6. (23) Kilicoglu, H., et al., Constructing a semantic predication gold standard from the biomedical literature. BMC Bioinformatics, 2011. 12: p. 486.
7. (24) Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (PROTECT): Adverse Drug Reactions Database. [web page] 2015.05.07]; Available from: http://www.imi-protect.eu/adverseDrugReactions.shtml.
8. (8) Duke, J., J. Friedlin, and X. Li, Consistency in the safety labeling of bioequivalent medications. Pharmacoepidemiol Drug Saf, 2013. 22(3): p. 294-301.
9. (25) Ryan, P.B., et al., Defining a reference set to support methodological research in drug safety. Drug Saf, 2013. 36 Suppl 1: p. S33-47.
10. (26) Coloma, P.M., et al., A reference standard for evaluation of methods for drug safety signal detection using electronic healthcare record databases. Drug Saf, 2013. 36(1): p. 13-23.
11. (27) Schuemie, M.J., et al., Detecting adverse drug reactions following long-term exposure in longitudinal observational data: The exposure-adjusted self-controlled case series. Stat Methods Med Res, 2014.
12. (28) Schuemie, M.J., et al., Replication of the OMOP experiment in Europe: evaluating methods for risk identification in electronic health record databases. Drug Saf, 2013. 36 Suppl 1: p. S159-69.