Accuracy of an Automated Knowledgebase for Identifying Adverse Drug Reactions

Erica A. Voss, MPH[1,2], Richard D. Boyce, PhD[3,2], Patrick B. Ryan, PhD[1,2,4], Peter R. Rijnbeek, PhD2[5,2], Martijn Schuemie, PhD[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, [5]Erasmus University Medical Center, Rotterdam, Netherlands

Abstract

An adverse drug reaction (ADR) is a response to a drug which is noxious and unintended at a dose that is normally used in humans (1). Large-scale Adverse Effects Related to Treatment Evidence Standardization (LAERTES) provides a standardized structure to enable analyses across relevant ADR evidence. This work reviews using LAERTES to build a classifier that groups pairs of drug and health outcomes of interest (HOI) either exhibiting properties of an ADR or not and performs a quantitative assessment of this classifier's predictive accuracy. The model addresses the immediate need of automating the selection of positive and negative controls and holds promise of trying to automate what human annotators have manually had to do to produce reference sets.

Introduction

Pharmacovigilance, the understanding and monitoring of a drug's safety profile, is of critical importance to both patients and healthcare providers. Drug safety evidence is produced across many disparate sources including randomized clinical trials (RCTs), drug product labels, spontaneous reports, and literatures. The process of safety assessment by pharmacovigilance practitioners is time-consuming and resource-intensive, as evidence is often compiled from varying, often unstructured, and generally incomplete, systems which yield conflicting information that must be interpreted and reconciled with appropriate clinical and analytical expertise. Large-scale Adverse Effects Related to Treatment Evidence Standardization (LAERTES) provides a standardized structure with a standardized vocabulary to enable analyses across relevant adverse drug reaction (ADR) evidence. The research goal was to build a classifier that leverages the LAERTES data and tries to identify drugs that cause certain outcomes (positive controls) and drugs that lack evidence for causing certain negative outcomes (negative controls). We performed a quantitative assessment of the predictive accuracy of the evidence base for discriminating between positive and negative controls. If it is determined this automated classifier could produce reference sets this would address the immediate need of automating the selection of positive and negative controls and holds promise of trying to automate what human annotators have manually had to do to produce reference sets.

Materials and Methods

Each piece of evidence from LAERTES was identified by a drug and health outcomes of interest (HOI) pair. A piece of evidence was also quantified by a statistic value, such as a count, or a summary associated measure. As evidence sources were loaded into LAERTES, any reference to a drug or condition was associated with standardized vocabularies (RxNorm for drugs and SNOMED for conditions). A subset of drugs and conditions was chosen to be the "universe" and only they were used for analysis. The term "universe" will be used within this work to refer to the drug ingredients and conditions that have enough evidence; having enough evidence meant that a single ingredient or single condition had at least one piece of evidence appear in each of the three main LAERTES evidence source types (literature, product labeling, and spontaneous reporting).

The goal of the analysis was to test automated discrimination between positive and negative controls using a model built on LAERTES data. Two manually generated reference sets, the Observational Medical Outcomes Partnership (OMOP) Reference Set (2) and the Exploring and Understanding Adverse Drug Reactions (EU-ADR) Reference Set (3), were used to train a regularized logistic regression model in discerning between positive and negative controls and then its predictive accuracy was estimated (using cross-validation). Only the most prevalent features that are present in LAERTES were used because some evidence was considered too infrequently reported to be useful. To test the generalizability of the combined model, a third reference set was used, the Arizona Center for Education and Research on Therapeutics (AZCERT) dataset (4).

Results

Table 1 provides the measured performance of models based on individual evidence types as well as the full model with all evidence types (however due to regularization not all evidence may play a role in the model). Looking at model fit to the OMOP reference set using individual pieces of evidence, we see that the US product labels were the most predictive (Area Under the Curve [AUC]=0.81

[95% CI: 0.77-0.85]) with the FAERS counts second (AUC=0.74 [95% CI: 0.68-0.80]). For the EU-ADR reference set, the Medline case reports were most predictive (AUC=0.88 [95% CI 0.81-0.95) followed by US product labels (AUC=0.77 [95% CI: 0.68-0.87]).

Figure 1 shows histograms of the predicted probability as well as the AUCs of the model for the positive and negative controls in all three reference sets. The plots suggest that the predicted probabilities produced by the algorithm were generally useful for segregating positive and negative controls. With AZCERT, the model was good at separating the ingredients that were not positive controls from those that were. However due to imbalance between the two classes Figure 1 looks to suggest that the classifier had a hard time with the positive controls, however review of the predictive probabilities shows there is separation. The AUCs for all models run shows they are fairly predictive and there is good separation between the positive and negative controls.

Table 1: AUC (Area Under the Receiver Operating Characteristics Curve) and 95% confidence interval for individual predictors and a regularized regression model using all predictors, using leave-pair-out cross-validation.

Column(s) in Model	OMOP AUC	EU-ADR
Medline Clinical Trial	0.57 (0.54-0.59)	0.64 (0.56-0.71)
Medline Case Reports	0.68 (0.63-0.72)	0.88 (0.81-0.95)
Medline Other	0.52 (0.50-0.54)	0.55 (0.50-0.60)
Medline SemMedDB Clinical Trial	0.58 (0.55-0.61)	0.57 (0.51-0.62)
Medline SemMedDB Case Reports	0.58 (0.55-0.61)	0.58 (0.51-0.65)
EU Product Labels	0.57 (0.54-0.60)	0.53 (0.49-0.56)
US Product Labels	0.81 (0.77-0.85)	0.77 (0.68-0.87)
FAERS *	0.74 (0.68-0.80)	0.68 (0.56-0.80)
FAERS PRR **	0.56 (0.49-0.63)	0.59 (0.47-0.72)
All Predictors	0.85 (0.81-0.90)	0.93 (0.87-0.99)

OMOP: Observational Medical Outcomes Partnership, EU-ADR: Exploring and Understanding Adverse Drug Reactions, AUC: area under the curve, LBCI: lower bound 95% confidence interval, UPCI: upper bound 95% confidence interval, FAERS: FDA Adverse Event Reporting System, PRR: proportional reporting ratio

* natural logs were taken to scale predictor

** geometric mean was used to scale predictor

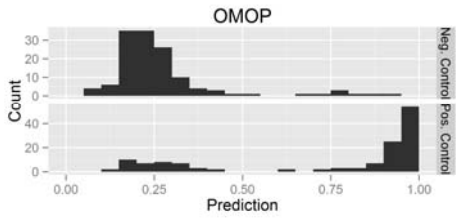
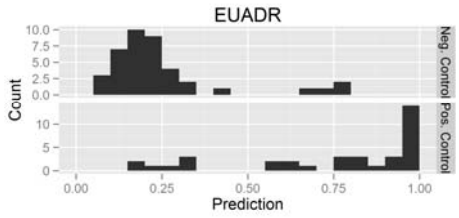
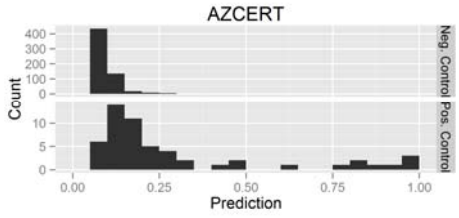
Histogram Comparing Predictive Probability Separation	AUC
	0.85 (0.81-0.90)
	0.93 (0.87-0.99)
	0.87 (0.63-0.93)*

Figure 1. Histograms of predicted probabilities with AUCs for positive and negative controls in the various reference sets, using the model trained on both OMOP and EU-ADR reference set.

*interval was calculated by assuming a 5% of the negative controls were misclassified.

Conclusions

The goal of this paper was to explore the use of LAERTES for automating the selection of positive and negative controls. We demonstrated that using LAERTES, evidence was predictive of the reference sets, particularly when all the predictors with enough data were utilized. The model classifier also performed well on AZCERT but due to in-balance of positive and negative controls it was harder to see the separation between the groups. LAERTES may have potential in identifying drug outcome pairs that warrant further evaluation, however this would need to be studied in the future. LAERTES holds promise of being a powerful tool in selecting positive controls in a faster manner than traditional manual efforts would allow.

References

1. Safety of Medicines: A guide to detecting and reporting adverse drug reactions [PDF]. Geneva: World Health Organization; 2002 [03/24/2012]. Available from: http://whqlibdoc.who.int/hq/2002/WHO_EDM_QSM_2002.2.pdf.

2. Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. Defining a reference set to support methodological research in drug safety. Drug safety. 2013 Oct;36 Suppl 1:S33-47. PubMed PMID: 24166222. Epub 2013/11/06. eng.

3. Coloma PM, Avillach P, Salvo F, Schuemie MJ, Ferrajolo C, Pariente A, et al. A reference standard for evaluation of methods for drug safety signal detection using electronic healthcare record databases. Drug safety. 2013 Jan;36(1):13-23. PubMed PMID: 23315292. Epub 2013/01/15. eng.

4. Woosley RL, Romero K. Assessing cardiovascular drug safety for clinical decision-making. Nat Rev Cardiol. 2013 06//print;10(6):330-7.