Accuracy of an automated knowledge base for identifying drug adverse reactions

E.A. Voss a, b, c, R.D. Boyce d, c, P.B. Ryan a, e, c, J. van der Lei b, c, P.R. Rijnbeek b, c, M.J. Schuemie a, c

a Janssen Research & Development, LLC
b Erasmus University Medical Medical Center
c Observational Health Data Sciences and Informatics (OHDSI)
d University of Pittsburgh
e Columbia University

Summary

- **[LAERTES]** a method for automatically aggregating disparate sources of drug adverse event information together into a single repository,

- **[MAIN RESEARCH]** developing a predictive model to classify drug-adverse event relationships,

- **[REAL WORLD APPLICATION]** applying those predictions to a real world problem of identifying negative controls for statistical method calibration.
Why is your research important?

• There is a general need for development of reference sets of drug/condition pairs and if they are a adverse drug event pair or not

• Real World Example:
Methods

Published Literature
Product Labels
Spontaneous Reports

OMOP Vocabulary

LAERTES

Adverse Drug Reaction?
- Rx1-Dx1 ✓
- Rx1-Dx2 ✗
- Rx2-Dx1 ✗
- Rx3-Dx2 ✓
- ...

R
Methods

• **Logistic regression** was used to build multivariate models 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.

• Parameters given to model:
  • Medline Clinical Trial
  • Medline Case Report
  • Medline Other
  • SemMedDB Clinical Trial
  • SemMedDB Case Report
  • EU Product Labels
  • US Product Labels
  • FAERS
  • FAERS PRR
Methods

• OMOP/EU-ADR Reference Sets used for training/testing of pilot method. AZCERT used for additional validation.

• First built a model and evaluated performance on individual reference sets EU-ADR and OMOP

• Second built a model on the combination of EU-ADR and OMOP and evaluated performance on a third reference set AZCERT
Results

• LAERTES had nearly 8 millions rows of evidence
  – 3797 distinct ingredients
  – 9403 distinct conditions

• The model performed well on predicting the reference sets [AUC (95% CI)] :
  – OMOP  – 0.93 (0.86-0.97)
  – EU-ADR  – 0.92 (0.86-0.97)
  – AZCERT  – 0.92 (0.89-0.95)

Fluvoxamine later added to the list
Conclusion

• LAERTES data was predictive of the reference sets

• Method provides a scalable alternative to the time/resource-intensive manual curation process of reference sets of positive/negative used in drug safety research