Accuracy of an Automated Knowledgebase for Identifying Adverse Drug Reactions

Erica Voss
(Martijn Schuemie presenting)
Background

For method **evaluation** and **calibration** we need controls:

- **Positive controls** – drugs-outcome pairs where the drug is known to cause the outcome
- **Negative controls** – drug-outcome pairs where we’re pretty sure there’s no causal relationship
Background

In the past, creating positive and negative controls was hard work

Defining a Reference Set to Support Methodological Research in Drug Safety

Patrick B. Ryan · Martijn J. Schuemie · Emily Welebob · Jon Duke · Sarah Valentine · Abraham G. Hartzema


Preciosa M. Coloma · Paul Avillach · Francesco Salvo · Martijn J. Schuemie · Carmen Ferrajoli · Antoine Pariente · Annie Fourrier-Réglat · Mariam Molokhia · Vaishali Patadia · Johan van der Lei · Miriam Sturkenboom · Gianluca Trifirò
Objective

To build a machine learning classifier using LAERTES to automatically identify positive and negative controls
## Predictors

<table>
<thead>
<tr>
<th>Data Source Type</th>
<th>Data Source &amp; Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Literature</strong></td>
<td>Medline MeSH Clinical Trials</td>
</tr>
<tr>
<td></td>
<td>Medline MeSH Case Reports</td>
</tr>
<tr>
<td></td>
<td>Medline MeSH Other</td>
</tr>
<tr>
<td></td>
<td>Medline SemMedDB Clinical Trials</td>
</tr>
<tr>
<td></td>
<td>Medline SemMedDB Case Reports</td>
</tr>
<tr>
<td></td>
<td>Medline SemMedDB Other</td>
</tr>
<tr>
<td><strong>Product Labels</strong></td>
<td>European Product Label Adverse Drug Reactions</td>
</tr>
<tr>
<td></td>
<td>Structured Product Label Adverse Drug Reactions from SPLICER</td>
</tr>
<tr>
<td><strong>Spontaneous Reports</strong></td>
<td>FAERS Report Count</td>
</tr>
<tr>
<td></td>
<td>FDA Adverse Event Reporting System (FAERS) Proportional Reporting Ratio (PRR)</td>
</tr>
</tbody>
</table>
Model

Regularized logistic regression

Result: single score reflecting probability that drug causes outcome (given the available information)

Negative control: $p < x$
Positive control: $p > y$
The LAERTES Universe

Need to have enough evidence on the drug and the outcome to have some confidence that

Lack of evidence of an effect =
Evidence of lack of an effect
The LAERTES Universe

Drugs (ingredients) and outcomes must have at least

• 1 FAERS record, and
• 1 Medline ADR record, and
• 1 product label

Outcomes use hierarchy: evidence of child counts as evidence for parent (e.g. acute MI is counted as MI)
The LAERTES Universe

- 992 distinct drugs (ingredients)
- 3,488 outcomes
- $992 \times 3,488 = 3.5\text{ mln}$ drug-outcome pairs where we can predict
Evaluation

Use previously created reference sets for training + evaluation (using cross-validation):
- OMOP reference set
- EU-ADR reference set

External set for evaluation only (train on OMOP and EU-ADR sets):
- AZCERT
## Results – OMOP & EU-ADR sets

<table>
<thead>
<tr>
<th>Column(s) in Model</th>
<th>OMOP AUC</th>
<th>EU-ADR AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medline Clinical Trial</td>
<td>0.74 (0.69-0.79)</td>
<td>0.73 (0.63-0.83)</td>
</tr>
<tr>
<td>Medline Case Reports</td>
<td>0.85 (0.81-0.89)</td>
<td>0.88 (0.81-0.96)</td>
</tr>
<tr>
<td>Medline Other</td>
<td>0.85 (0.80-0.89)</td>
<td>0.87 (0.79-0.95)</td>
</tr>
<tr>
<td>Medline SemMedDB Clinical Trial</td>
<td>0.58 (0.55-0.61)</td>
<td>0.57 (0.51-0.63)</td>
</tr>
<tr>
<td>Medline SemMedDB Case Reports</td>
<td>0.58 (0.55-0.61)</td>
<td>0.59 (0.52-0.65)</td>
</tr>
<tr>
<td>EU Product Labels</td>
<td>0.57 (0.54-0.60)</td>
<td>0.53 (0.49-0.57)</td>
</tr>
<tr>
<td>US Product Labels</td>
<td>0.87 (0.84-0.91)</td>
<td>0.80 (0.71-0.89)</td>
</tr>
<tr>
<td>FAERS</td>
<td>0.73 (0.67-0.78)</td>
<td>0.70 (0.57-0.82)</td>
</tr>
<tr>
<td>FAERS PRR</td>
<td>0.64 (0.58-0.70)</td>
<td>0.75 (0.63-0.86)</td>
</tr>
<tr>
<td>All Predictors</td>
<td>0.94 (0.91-0.97)</td>
<td>0.92 (0.86-0.98)</td>
</tr>
</tbody>
</table>

Using leave-one-out crossvalidation
Results – AZCERT set

55 drugs in universe and AZCERT ‘certain’ category are considered positive

Assuming all 865 drugs in universe and not in AZCERT are negative:
AUC = 0.92 (0.89-0.95)

Assuming worst case: 1% lowest predicted are positive:
AUC = 0.79

Assuming best case: 5% highest predicted are positive:
AUC = 0.94
Conclusions

• Able to automatically ‘predict’ positive and negative controls with high accuracy

• Challenge: outcomes can be at all levels in the hierarchy (e.g. lots of evidence for ‘Condition’, and all drugs seem to cause conditions)
Next steps

• Apply the model to find controls
• Continue fitting the model as data in LAERTES is refreshed
• Include additional predictors?
• Try other types of classifiers?
Null distribution

The null distribution represents the expected relative risk under the null hypothesis that there is no effect (RR = 1).

It represents the uncertainty due to having a limited sample size.
Null distribution

The colored area indicates the probability that the observed RR would be observed if the true RR = 1

This is what we call the p-value

Drug X
Null distribution

clopidogrel – GI bleed:
\[ p < .001 \]
Null distribution

clopidogrel – GI bleed
Negative controls

We used exactly the same method, parameter settings and database to estimate the RR of ketoconazole – GI bleed.
Negative controls

55% of negative controls have p < .05 (Expected: 5%)
Negative controls

We can use the negative controls to estimate the real null distribution.
Under this empirical null distribution, we cannot reject the null hypothesis for clopidogrel: $p > 0.05$