Statistical signal detection for spontaneous reports

Suzie Seabroke

PROTECT Symposium February 19-20 2015
Contents

1. Which disproportionality method to use?
2. Subgroups and stratification
3. Unmasking
4. Drug-drug interactions
5. Duplicate detection
Study Objectives

1. to evaluate the performance of different signal detection algorithms

2. to investigate the impact of stratified and subgroup analyses in routine first pass signal detection

Within several spontaneous databases of varying size and characteristics
Partners and Databases

• **Regulatory Authority**
  - European Medicines Agency
  - MHRA (UK)

• **Pharma Industry**
  - AstraZeneca
  - Bayer*
  - GlaxoSmithKline
  - Roche*

• **Research**
  - Uppsala Monitoring Centre

* Participated in study #1 only on method comparison
Signal detection performance measured using two/three performance indicators:

These are calculated on the entire dataset and also as they evolve over time:

1) **Sensitivity (the true positive rate)**
   - Proportion of true ADRs that are correctly detected

2) **Precision (positive predictive value)**
   - Proportion of detected signals that correspond to a true ADR

3) **Time to detection for the true positives (median or average)**
   - How much time is gained by earlier signalling?
• 220 drugs were selected to represent a variety of therapeutic areas and patient populations

• True ADRs were those listed in the SPC section 4.8 or company core data sheets
## Study 1 – Which method to use?

<table>
<thead>
<tr>
<th>Disp. measure</th>
<th>Implementation</th>
</tr>
</thead>
</table>
| PRR           | PRR_{025} ≥ 1 & n ≥ 3  
RRR_{025} ≥ 1 & n ≥ 5  
PRR ≥ 3 & $\chi^2$ ≥ 4 & n ≥ 3  
PRR ≥ 2 & $\chi^2$ ≥ 4 & n ≥ 3  
PRR ≥ 2 & p ≤ 0.05 & n ≥ 3 |
| ROR           | ROR_{025} ≥ 1 & n ≥ 3  
ROR_{025} ≥ 1 & n ≥ 5  
ROR_{025} > 1 with shrinkage  
ROR_{025} > 2 & n ≥ 5 |
| IC            | IC_{025} > 0 |
| EBGM          | EB05 ≥ 1.8 & n ≥ 3 & EBGM ≥ 2.5  
EB05 ≥ 1.8 or positive trend flag  
EB05 > 2.0 or positive trend flag |
| Urn           | RR > 1 & unexpectedeness > 1 / 0.05  
RR > 1 & unexpectedeness > 500 / 0.05 |
Study 1 - Precision and sensitivity for all measures across databases

![Graph showing precision and sensitivity for various methods across databases. The graph includes data points for different datasets and owners.](image)

**METHODS**
- EB05 (2.0, trend)
- EB05 (1.8, trend)
- EB05 (1.8, 3, 2.5)
- PRR025 (1.0, 3)
- PRR025 (1.0, 5)
- PRR (2.0, 3, 4.0)
- PRR (2.0, 3, 3.84)
- PRR (3.0, 3, 4.0)
- PRR025 (1.0, 3)
- PRR025 (1.0, 5)
- PRR025 (1.0, SHR)
- PRR025 (2.0, 5)
- URN1
- URN500
- IC

**DATASET OWNER**
- UMC
- EMA
- MHRA
- AZ
- Bayer
- GSK
- Roche
Study 1 - Performance of measures after database standardisation
Study 1 - Mean precision and sensitivity over databases

![Graph showing mean adjusted precision over databases for different studies](image)

- EB05 (2.0, trend)
- EB05 (1.8, trend)
- EB05 (1.8, 3, 2.5)
- PRR025 (1.0, 3)
- PRR025 (1.0, 5)
- PRR (2.0, 3, 4.0)
- PRR (2.0, 3, 3.84)
- PRR (3.0, 3, 4.0)
- ROR025 (1.0, 5)
- ROR025 (1.0, SHR)
- ROR025 (2.0, 5)
- URN1
- URN500
- IC

Mean adjusted precision over databases
Study 1 – Envelope of precision and sensitivity achievable with PRR

Lower confidence bound threshold varies from 0 to 5 in steps of 0.1

Precision

- EB05 (2.0, trend)
- EB05 (1.8, trend)
- EB05 (1.8, 2.5)
- PRR025 (1.0, 3)
- PRR025 (1.0, 5)
- PRR (2.0, 3, 4.0)
- PRR (2.0, 3, 3.84)
- PRR (3.0, 3, 4.0)
- ROR025 (1.0, 3)
- ROR025 (1.0, 5)
- ROR025 (2.0, 5)
- URN1
- URN500
- IC
Study 1 - Change in precision over time
Study 1 - Conclusions

- All disproportionality methods can achieve similar overall performance by choice of algorithm
- Choice of algorithm can provide very different levels of performance
- Relative performance of an algorithm in one database can be predicted from research in others
- Precision seems to decrease over time on the market

### Study 2 – Subgroups & Stratification

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Strata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0-23months, 2-11, 12-17, 18-35, 36-64, 65-74, 75+ years, unknown</td>
</tr>
<tr>
<td>Gender</td>
<td>Male, female, unknown</td>
</tr>
<tr>
<td>Time period</td>
<td>5-yearly</td>
</tr>
<tr>
<td>Vaccines/Drugs</td>
<td>Vaccines, non-vaccines</td>
</tr>
<tr>
<td>Event seriousness</td>
<td>Serious, non-serious</td>
</tr>
<tr>
<td>Reporter qualification</td>
<td>Consumer only, healthcare professional only, mixed</td>
</tr>
<tr>
<td>Report source</td>
<td>Spontaneous only</td>
</tr>
<tr>
<td>Country of origin</td>
<td>Individual country of origin</td>
</tr>
<tr>
<td>Region of origin</td>
<td>North America, Europe, Asia, Japan, Rest of the World</td>
</tr>
</tbody>
</table>
Study 2 - Methods

- Stratified analyses conducted using Mantel-Haenszel approach to obtain a single adjusted value

- Subgroup analyses calculated disproportionality statistics within individual strata separately

- Stratified/subgroup results compared to crude unadjusted results

- Disproportionality statistics:
  - $ROR_{0.25} \geq 1 \& n \geq 3$
  - $IC_{0.25} > 0$
  - $EBGM \geq 2.5, EB05 \geq 1.8$ and $n \geq 3$
Study 2 - Precision and sensitivity for stratified & subgroup analyses (ROR)

- **Precision**
- **Sensitivity**

**Graphs:**
- UMC
- EMA
- AZ
- GSK
- MHRA

**Legend:**
- Age
- Gender
- Time period
- Vaccine
- Seriousness
- Reporter
- Spontaneous
- Country of origin
- Continent of origin
Study 2 - Precision and sensitivity for stratified & subgroup analyses (Bayesian)
Study 2 - Precision and sensitivity for stratified, subgroup & random strata
Study 2 - Conclusions

- Subgroup analyses consistently performed better than stratified analyses

- Subgroup analyses are beneficial in large, international databases. Smaller databases may need to consider a likely tradeoff between sensitivity and precision

- Choice of variables for subgroup analyses will likely vary between different datasets

Influence of Masking on Disproportionality

- Developed masking ratio to quantify masking effect of given product
- Assessed extent and impact of masking in Eudravigilance and Pfizer spontaneous database

- Prevalence of important masking quite rare (0.003% DECs)
- Important masking mainly concerns rarely reported events


Drug-Drug Interaction Detection

- Objective: Compare sensitivity & specificity of 4 different measures to detect drug-drug interactions

- Reference set:
  - established DDIs & D-E pairs with no known association
  - emerging DDIs from Stockley’s interaction alerts 2007-2009 & D-E pairs not included in same reference

- WHO Vigibase used for analysis
Conclusion: Statistical interaction measures with additive baseline models should be preferred over multiplicative models for detecting drug-drug interactions in spontaneous data.

Duplicate Detection

- Objective: compare probabilistic record matching algorithm (VigiMatch) with rule-based approaches

- MHRA, DHMA & AEMPS participated

- Initial evaluation: suspected VigiMatch duplicates 2000-2010 were assessed by respective national centre

- Second evaluation: direct comparison between VigiMatch & MHRA rule-based algorithm
Duplicate Detection

- Initial evaluation showed VigiMatch to return few false positives in all 3 national centres

- Direct comparison:

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>Rule-based screening</th>
<th>vigiMatch™</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unrelated</strong></td>
<td>56</td>
<td>0</td>
</tr>
<tr>
<td>Not duplicates, but otherwise related</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>Likely duplicates, not yet confirmed</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Confirmed duplicates</strong></td>
<td><strong>30</strong></td>
<td><strong>87</strong></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Duplicate Detection

Conclusion: Probabilistic record matching should be considered as an alternative to rule-based methods for duplicate detection in spontaneous data