A structured database of Adverse Drug Reaction based on information from the Summary of Product Characteristics

Tomas Bergvall¹
Gunnar Dahlberg¹
Nils Opitz²
Bénédicte Capelli³
Xavier Kurz³
Niklas Norén¹
Jim Slattery³

1. Uppsala Monitoring Centre
2. Bayer Pharmaceuticals
3. European Medicines Agency
Acknowledgements

The research leading to these results was conducted as part of the PROTECT consortium (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium, www.imi-protect.eu) which is a public-private partnership coordinated by the European Medicines Agency.

The PROTECT project has received support from the Innovative Medicine Initiative Joint Undertaking (www.imi.europa.eu) under Grant Agreement n° 115004, resources of which are composed of financial contribution from the European Union's Seventh Framework Program (FP7/2007-2013) and EFPIA companies’ in kind contribution.

The views expressed are those of the authors only.
Background to project

- In addition to beneficial effects, all medicine can cause adverse drug reactions (ADR)
- ADRs should be coded in MedDRA terminology on the Summary of Product Characteristics but frequently non-MedDRA text is used
- This presents an obstacle to incorporation of the data into systems to help identify ADRs in clinical practice or to speed up drug safety assessments
Typically:

• If a doctor sees a patient with a problem that may be caused by a medicine s/he needs to read all the relevant SPCs to decide which medicine is likely to be the culprit.
• If a drug safety expert thinks that a new adverse drug reaction has been observed the SPC must be read to exclude the possibility that it is a known effect coded in an unconventional way.
PROTECT Work Package 3

- SP3: Structured database of SPC section 4.8
  - Structured database of known ADR of centrally-authorised products used in the EU
  - Currently 651 authorised (August 2012)
  - Map to 19550 Preferred Terms (MedDRA 15.0)
  - Plan an ongoing maintenance of the database
Process and secondary aims

- Extract data on adverse reactions from SPCs for CAPs, if possible with gradation according to levels of evidence with which the ADR has been found/detected.
- Validate both against company core data sheets and with MedDRA.
- Pilot a process for extending the coverage to mutual recognition and nationally authorised products.
- Start an incremental process to identify groups of preferred terms in MedDRA that are synonymous in the context of ADR reporting.
- Test the use of the database in the Adverse Drug Reaction signal detection process (EMA);
- Establish processes for maintenance, access to and dissemination of the dataset, and a rational path for extending the product coverage.
Mapping

1. Verbatim term copied from SPC → Exact match to MedDRA?
   - Yes → Stop
   - No → Approximate match?

2. Approximate match?
   - No → Ad hoc search by Research group
   - Yes → Expert check

3. Expert check
   - Disagree → Fail
   - Agree → Add to list of acceptable synonyms

4. Success
Approximate Matching Process

- Development at Uppsala Monitoring Centre
- Porter stemming
- Stop word removal
- Synonym replacement
  - Ad hoc identification of synonyms and spelling variations
- Word order permutation
Development of method

- Trained on a set of 75 European SPCs
  - Identified additional synonyms and Latin/Greek spelling variations
    - ae->e, oe->e, y->i, etc.
- Tested on a set of 169 centrally authorized products (CAPs)
- Matched MedDRA terms extracted from 414 CAPs compared against 1036 potential drug-ADR signals from VigiBase™
Question

- Given that:
  - MedDRA Preferred terms are a well-defined, structured and, in computing terms, fairly small set
  - The development of the acceptable synonyms list steadily reduces the number of terms to be matched
- Is it worth using sophisticated matching processes on a moderate sized and largely resolvable problem?
Results of approximate matching

- Algorithm performance on the 414 SPCs
- Verbatim matches
  - 72% hit rate
- Baseline algorithm
  - 87% hit rate of MedDRA terms
- Additional synonyms and spelling variations
  - 98% hit rate of MedDRA terms
Impact on project

- Ad hoc searching of MedDRA proved to be enormously time consuming! Some redundancy in PTs means that it is not just finding one match that matters but finding the best – and possibly alternative – matches.

- From the EMA point of view the UMC algorithm reduced the coding from an intractable problem to one on which we were able to meet our time-lines for extending the dataset from 2009 to 2011.
Insurance against unpredictable coding

- The list of synonyms is a useful resource but:
- In extending the dataset to older products and nationally authorised we cannot be sure that choice of non-standard terms is stable over time or unrelated to cultural preferences.
- Hence to UMC matching is still proving useful in extending the dataset to other products used in the PROTECT project.
Use of the coding dataset

- Database not an end in itself but a tool for:
  - Expediting signal detection
  - Adjusting statistical SD for established ADRs
  - Research

- Thus the SP itself is only developing a process for recording and maintenance

- Proof of usefulness will come from other SPs and from adoption by other research projects
UMC estimates of impact on current Signal Detection

- Considered 1036 safety signals detected in the Vigibase spontaneous reporting database
- 16% could be immediately removed as ‘known’ using the matching
- Estimated workload reduction of 40 man.hours per year at UMC.
Improvements to signal detection?

- A database of known ADRs may help us to focus on as yet unidentified problems.
Spontaneous reporting databases

• When a patient has an adverse event that may be caused by a medicine a report is generated:
  – List of medicines
  – Adverse event(s)
  – Demographics, times and additional circumstances

• Collections of such reports are SR databases

• Drug-event combinations representing potential ADRs can be ‘trawled’ using disproportionality statistics (eg PRR)
**Proportional Reporting Ratio**

\[
\text{PRR} = \frac{a/(a+b)}{c/(c+d)}
\]

\[
\text{Var}\{\ln \text{PRR}\} = 1/a - 1/(a+b) + 1/c - 1/(c+d)
\]

\[
\text{LL}\{\text{PRR}\} = \frac{\text{PRR}}{\exp(1.96 \times \text{Var}\{\ln \text{PRR}\})}
\]

<table>
<thead>
<tr>
<th></th>
<th>Event E</th>
<th>Not event E</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Drug D</strong></td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td><strong>Not drug D</strong></td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
<tr>
<td><strong>Column total</strong></td>
<td>a+c</td>
<td>b+d</td>
<td>a+b+c+d</td>
</tr>
</tbody>
</table>
Masking

- Some the products included in the denominator of the PRR may be causally associated with event E
- These will tend to reduce the PRR
- It makes sense to remove those drugs with a known causal association from the calculation
- The structured database of ADRs makes this a relatively straightforward procedure
- Total impact still to be evaluated
Improvements to clinical practice?

- Computers are a routine part of clinical practice
- Linking the database with the clinical practice record could immediately alert the doctor when a patient has suffered an event known to be associated with one of their treatments
Conclusions

• Substantial performance gain from approximate mapping of European SPCs to MedDRA (72% - 98%)
• Useful workload reduction in the signal detection process
• Application of matching process to other data ongoing
• Application of database to pharmacovigilance in process of evaluation
The End

- Address for database: