Use of ontologies in signal detection
SPC section 4.8 coding subpackage

• Map each textual description of an adverse drug reaction in section 4.8 of the SPC for centrally authorised products to formal MedDRA terms
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 automated systems to help identify ADRs in clinical practice or to speed up drug safety assessments.
Motivations

- **Clinical:** Determine quickly if an adverse event might be associated with a medicine

- **Pharmacovigilance:** Automatic checks of whether a new signal is already known
Study 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 against company core data sheets and MedDRA
- Pilot extension to non CAPs.
- 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 ADR signal detection process;
- Establish processes for maintenance, access to and dissemination of the dataset, and a rational path for extending the product coverage.
Mapping

Verbatim term copied from SPC

Exact match to MedDRA?

Yes → Stop

No →

Approximate match?

No → Ad hoc search by Research group

Yes →

Expert check

Disagree → Failure

Agree → Add to list of acceptable synonyms

Success
Results of the UMC matching

- Algorithm performance on the 414 SPCs
- Verbatim matches
  - 72% hit rate of MedDRA terms
- Matching with algorithm
  - 98% hit rate of MedDRA terms
Database at February 2015

• SP3: Structured database of SPC section 4.8
  – Structured database of known ADR of centrally-authorised products used in the EU
  – Currently 653 products
  – 430 substances
  – 45298 ADRS
  – 2660 Preferred Terms
  – Ongoing maintenance of the database but only to June 2012 at present
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

• Proof of usefulness will come from adoption by other research projects
Address for database

MedDRA grouping subprojects

1. Determine whether grouping medical terms using the currently established hierarchies can speed up signal detection. Lead: UMC

2. Determine if creating new groupings using a currently established ontology can speed up signal detection. Lead: INSERM
Retrospective study

- 13 medical concepts with medium to high probability of being drug-related (Trifirò et al, Drug Safety 2009)
- 43 EMA labelling changes (Alvarez et al, Drug Safety 2010)
- Scope of study
  - Sets of individual Preferred Terms
  - High-Level Terms
  - Narrow SMQs
  - Groupings of manually selected Preferred Terms (custom groups)
PTs highlight early!

- Individual MedDRA Preferred Terms trump other groupings in terms of timeliness
- Results for 43 EMA labelling changes:

<table>
<thead>
<tr>
<th>Terminology level</th>
<th>Total</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>MedDRA PT</td>
<td>25</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>MedDRA HLT</td>
<td>23</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>SMQ, narrow</td>
<td>19</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>CustomGroup</td>
<td>23</td>
<td>17</td>
<td>6</td>
</tr>
</tbody>
</table>
Interpretation?

- May indicate that important medical distinctions exist between PTs
- May be due to multiple comparisons – more chances to signal with several categories.
An attempt to expedite signal detection by grouping related adverse reaction terms. Richard Hill, Johan Hopstadius, Magnus Lerch, and G. Niklas Noren

Ontology  *noun*

- the branch of metaphysics dealing with the nature of being.
What ICT people say

- In the context of computer and information sciences, an ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application. In the context of database systems, ontology can be viewed as a level of abstraction of data models, analogous to hierarchical and relational models, but intended for modeling knowledge about individuals, their attributes, and their relationships to other individuals. Ontologies are typically specified in languages that allow abstraction away from data structures and implementation strategies; in practice, the languages of ontologies are closer in expressive power to first-order logic than languages used to model databases. For this reason, ontologies are said to be at the "semantic" level, whereas database schema are models of data at the "logical" or "physical" level. Due to their independence from lower level data models, ontologies are used for integrating heterogeneous databases, enabling interoperability among disparate systems, and specifying interfaces to independent, knowledge-based services. In the technology stack of the Semantic Web standards [1], ontologies are called out as an explicit layer. There are now standard languages and a variety of commercial and open source tools for creating and working with ontologies. Tom Gruber
Ontologies for the simple minded

• An attempt to list the objects required to answer a problem and formally describe their properties and the relationships between them in a way that allows a computer to handle them.
MedDRA terms are associated with semantic definitions

- Semantic definitions manually revised for more than 2000 PTs

<table>
<thead>
<tr>
<th>Semantic definition</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>hasFindingSite</td>
<td></td>
</tr>
<tr>
<td>hasAssociatedMorphology</td>
<td></td>
</tr>
<tr>
<td>associatedWith</td>
<td></td>
</tr>
<tr>
<td>hasClinicalCourse</td>
<td></td>
</tr>
<tr>
<td>interprets</td>
<td></td>
</tr>
<tr>
<td>hasDefinitionalManifestation</td>
<td></td>
</tr>
</tbody>
</table>

10028596 Myocardial infarction

<table>
<thead>
<tr>
<th>hasFindingSite</th>
<th>Myocardium structure [body structure]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coronary artery structure [body structure]</td>
</tr>
<tr>
<td></td>
<td>Healed infarct [morphologic abnormality]</td>
</tr>
<tr>
<td></td>
<td>Acute coronary syndrome [disorder]</td>
</tr>
<tr>
<td>hasClinicalCourse</td>
<td>Sudden onset AND/OR short duration [qualifier value]</td>
</tr>
<tr>
<td></td>
<td>Angiography [procedure] &amp; Abnormal [qualifier value]</td>
</tr>
<tr>
<td></td>
<td>Electrocardiogram, rhythm [procedure] &amp; Abnormal [qualifier value]</td>
</tr>
<tr>
<td></td>
<td>Blood test [procedure] &amp; Abnormal [qualifier value]</td>
</tr>
<tr>
<td>interprets</td>
<td>Pain [finding]</td>
</tr>
<tr>
<td></td>
<td>Chest pain [finding]</td>
</tr>
<tr>
<td></td>
<td>Myocardial infarction [disorder]</td>
</tr>
<tr>
<td></td>
<td>Dyspnea [finding]</td>
</tr>
</tbody>
</table>
Rationale

- MedDRA organised primarily by medical categories
- Terms that may be related within a particular case presentation are not easily identified.
- Interchangeable terms are not necessarily clear
- An ontology that identified the properties of terms might help to extract sensible groupings not manifest in the current hierarchy.
Study objectives

- Test similarity between groupings created using ontology and MedDRA SMQs.
- Compare signal detection with new groupings to signal detection with SMQs
Example: Upper GI bleeding + occult blood + blood in stool

\[
\text{hasFindingSite some 'Upper gastrointestinal tract structure'} \\
\text{AND hasAssociatedMorphology some 'Hemorrhage'} \\
\text{OR} \\
\text{interprets some 'Occult blood screening'} \\
\text{AND hasInterpretation some 'Positive'} \\
\text{OR} \\
\text{interprets some 'Evaluation of stool specimen'} \\
\text{AND hasAssociatedMorphology some 'Hemorrhage'}
\]
Searching for upper GI bleeding with the query tool

**SIMPLE QUERY**

Finding Site
- Upper digestive tract structure (body structure)
- Hemorrhage (morphologic abnormality)

**RESULTS**

38 MedDRA terms matching your criteria
- Anastomotic ulcer haemorrhage
- Anastomotic ulcer perforation
- Angina bullosa haemorrhagica
- Duodenal ulcer haemorrhage
- Duodenal ulcer perforation
- Duodenal varices
- Duodenitis haemorrhagic
- Erosive duodenitis
- Gastric haemorrhage
- Gastric ulcer haemorrhage
- Gastric ulcer haemorrhage, obstructive
- Gastric ulcer perforation

**MATCH DEFINITION**

Specify the list of properties for our ADRs search.

Criterion to include:

Criterion to exclude:

**PREDEFINED SEARCHES:**

- Disorder related to an organ (e.g. hepatitis)
- Disorder related to a causative agent (e.g. viral hepatitis)
- Disorder related to blood component (e.g. pancytopenia)
- Disorder without organ localization (e.g. depression)
- Laboratory finding (e.g. blood creatinine increased)
- Investigation finding (e.g. electrocardiogram abnormal)
- Operation (e.g. duodenal operation)

**FINDING SITE**

Body site affected by a condition

Manual search: Type text here...

Browse SNOMED-CT:
- Anatomical or acquired body structure (body structure)
Comparison with target SMQ

2 non retrieved terms:
- Duodenal operation
- Ulcer haemorrhage

4 additional retrieved terms:
- Aorto-oesophageal fistula
- Erosive duodenitis
- Gastric antral vascular ectasia
- Portal hypertensive gastropathy

\[
Recall = \frac{\text{relevant terms} \cap \text{retrieved terms}}{\text{relevant terms}} = 92.6\%
\]

\[
Precision = \frac{\text{relevant terms} \cap \text{retrieved terms}}{\text{retrieved terms}} = 86.2\%
\]
## Results for all 13 topics

<table>
<thead>
<tr>
<th>Safety topic</th>
<th>MedDRA gold standards</th>
<th>sem.match</th>
<th>RESULTS</th>
<th>F-measure</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bullous eruptions</td>
<td>HLT Bullous conditions</td>
<td>++</td>
<td>Recall</td>
<td>81.3%</td>
<td>47.3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Precision</td>
<td>33.3%</td>
<td></td>
</tr>
<tr>
<td>2. Acute renal failure</td>
<td>SMQ Acute renal failure</td>
<td>++</td>
<td></td>
<td>57.9%</td>
<td>54.3%</td>
</tr>
<tr>
<td></td>
<td>HLT Renal failure and impairment</td>
<td>+</td>
<td></td>
<td>47.8%</td>
<td>51.2%</td>
</tr>
<tr>
<td>3. Anaphylactic shock</td>
<td>SMQ Anaphylactic/anaphylactoid shock</td>
<td>++</td>
<td></td>
<td>34.5%</td>
<td>36.4%</td>
</tr>
<tr>
<td></td>
<td>HLT Anaphylactic responses</td>
<td>+</td>
<td></td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>4. Rhabdomyolysis</td>
<td>SMQ Rhabdomyolysis/myopathy</td>
<td>+</td>
<td></td>
<td>46.7%</td>
<td>47.2%</td>
</tr>
<tr>
<td>5. Aplastic anaemia / pancytopenia</td>
<td>SMQ Cytopenia and haematopoietic...</td>
<td>++</td>
<td></td>
<td>55%</td>
<td>55%</td>
</tr>
<tr>
<td></td>
<td>HLT Marrow depression and hypoplastic...</td>
<td>+</td>
<td></td>
<td>98.3%</td>
<td>80%</td>
</tr>
<tr>
<td>6. Neutropenia</td>
<td>SMQ Leukopenia SELECT</td>
<td>++</td>
<td></td>
<td>100%</td>
<td>77.8%</td>
</tr>
<tr>
<td></td>
<td>HLT Neutropenias</td>
<td>++</td>
<td></td>
<td>100%</td>
<td>74.3%</td>
</tr>
<tr>
<td>7. Cardiac valve fibrosis</td>
<td>HLTGT Cardiac valve disorders SELECT</td>
<td>++</td>
<td></td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>8. Extrapyramidal disorders</td>
<td>SMQ Extrapyramidal syndrome</td>
<td>+</td>
<td></td>
<td>64.9%</td>
<td>55.6%</td>
</tr>
<tr>
<td></td>
<td>HLT Dystiniasis and movement disorders</td>
<td>+</td>
<td></td>
<td>100%</td>
<td>83.3%</td>
</tr>
<tr>
<td>9. Confusional state</td>
<td>SMQ Infectious encephalopathy</td>
<td>+</td>
<td></td>
<td>16.0%</td>
<td>21.1%</td>
</tr>
<tr>
<td></td>
<td>HLT Confusion and disorientation</td>
<td>+</td>
<td></td>
<td>100%</td>
<td>60%</td>
</tr>
<tr>
<td>10. Thrombocytopenia</td>
<td>SMQ Thrombocytopenia</td>
<td>++</td>
<td></td>
<td>100%</td>
<td>60.5%</td>
</tr>
<tr>
<td></td>
<td>HLT Thrombocytopenias</td>
<td>++</td>
<td></td>
<td>100%</td>
<td>53.7%</td>
</tr>
<tr>
<td>11. Upper gastrointestinal bleeding</td>
<td>SMQ Gastrointestinal haemorrhage SELECT</td>
<td>++</td>
<td></td>
<td>92.6%</td>
<td>89.3%</td>
</tr>
<tr>
<td>12. Peripheral neuropathy</td>
<td>SMQ Peripheral neuropathy</td>
<td>++</td>
<td></td>
<td>66.7%</td>
<td>40.4%</td>
</tr>
<tr>
<td></td>
<td>HLTGT Peripheral neuropathies</td>
<td>++</td>
<td></td>
<td>98.9%</td>
<td>85.2%</td>
</tr>
<tr>
<td>13. Maculo-papular erythematos eruptions</td>
<td>HLT Rashes, eruptions and exanthem NEC</td>
<td>+</td>
<td></td>
<td>100%</td>
<td>33.9%</td>
</tr>
</tbody>
</table>

### Mean values

<table>
<thead>
<tr>
<th>(+) and (++) groupings</th>
<th>HLT/HLGT</th>
<th>TOT</th>
<th>Mean F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td>92.8%</td>
<td>78.8%</td>
<td></td>
</tr>
<tr>
<td>Precision</td>
<td>60.3%</td>
<td>55.9%</td>
<td></td>
</tr>
<tr>
<td>F-measure</td>
<td>54.0%</td>
<td>62.3%</td>
<td></td>
</tr>
</tbody>
</table>

Recall, precision and F-measure rates of automatic OnetoADR groupings. **SELECT** (third column) indicates that a manual selection of PTs has been made in the MedDRA grouping taken as gold standard. **(++)** MedDRA reference groupings having a perfect semantic match with the safety topic. **(+)** MedDRA reference groupings having an imperfect semantic match with the safety topic (broader or narrower topic).
Comparison of statistical signal detection

- EB05 calculated using automated grouping and target SMQ
- Spontaneous reports from FDA public access data
Very similar results

\[ y = 1.0082x + 0.0048 \]
\[ R^2 = 0.9931 \]

\( R^2 \): coefficient of determination (correlation coefficient squared), \( R^2 \) ranges from 0 to 1, where 1 indicates that the regression line perfectly fits the data.
Questions

• How much of the clinical knowledge about a MedDRA can be reflected in the ontology?
  – Many subtle distinctions and ambiguities of language to be considered
  – The more complex the logical structure the slower the computing

• Are SMQs a good Gold Standard in view of the results showing that PT give better signal detection?

• Might better SD be obtained with simpler rather than more sophisticated groupings? Note the existence of some terms with practically indistinguishable meanings.
References


- Automatic generation of MedDRA terms groupings using an ontology G Declerck, C Bousquet, MC Jaulent MIE 2012, 73-77
Overall messages

- Preferred Terms are probably the best of the existing levels of MedDRA for signal detection.
- Automated groupings look promising for locating terms for SMQ development.
- Difficult to make any recommendation concerning signal detection with new groupings at present.
What was not covered

- Synonym identification in MedDRA
- Grouping of products
  - How to treat multi-constituent products
  - How to treat classes of products
    - With exception that vaccines were a subgroup in another PROTECT work package