PROTECT

imi efpia

Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium

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

PROTECT

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

PROTECT

Motivations

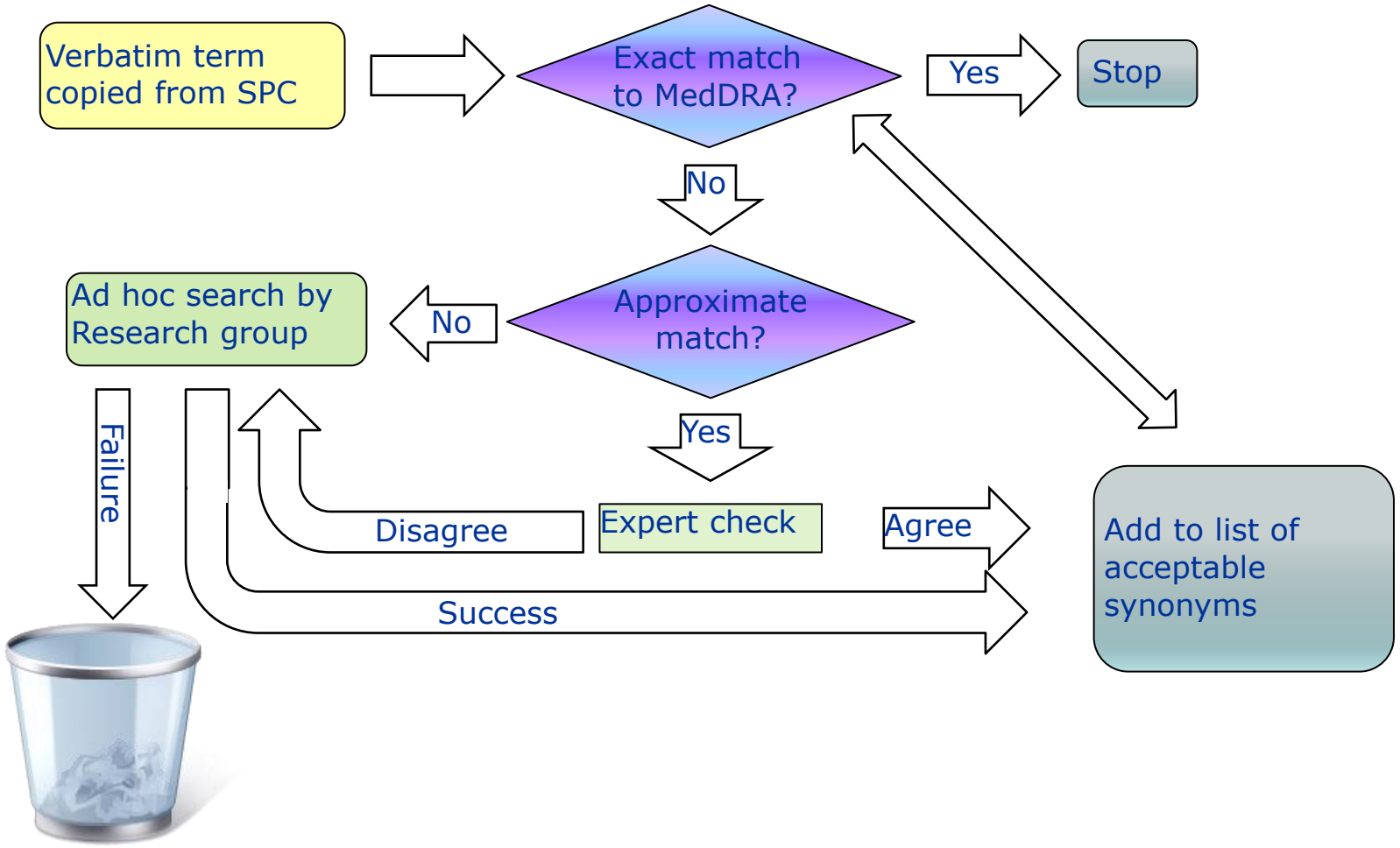
- **Clinical:** Determine quickly if an adverse event might be associated with a medicine
- **Pharmacovigilance:** Automatic checks of whether a new signal is already known

PROTECT

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
Failure
Disagree
Expert check
Agree
Success
Add to list of acceptable synonyms

PROTECT

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 from adoption by other research projects

Address for database

- http://www.imi-protect.eu/methodsRep.shtml

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:

Terminology level	Total	Before	After
MedDRA PT	25	18	7
MedDRA HLT	23	17	6
SMQ narrow	19	12	7
CustomGroup	23	17	6

the UPPSALA MONITORING CENTRE

Interpretation?

- May indicate that important medical distinctions exist between PTs
- May be due to multiple comparisons – more chances to signal with several categories.

Reference

- An attempt to expedite signal detection by grouping related adverse reaction terms. Richard Hill, Johan Hopstadius, Magnus Lerch, and G. Niklas Noren

  - http://www.imi-protect.eu/documents/HilletalGroupingrelatedmedicaltermsfinalMIE2012.pdf

PROTECT

Ontology *noun*

- the branch of metaphysics dealing with the nature of being.

What ICT people say

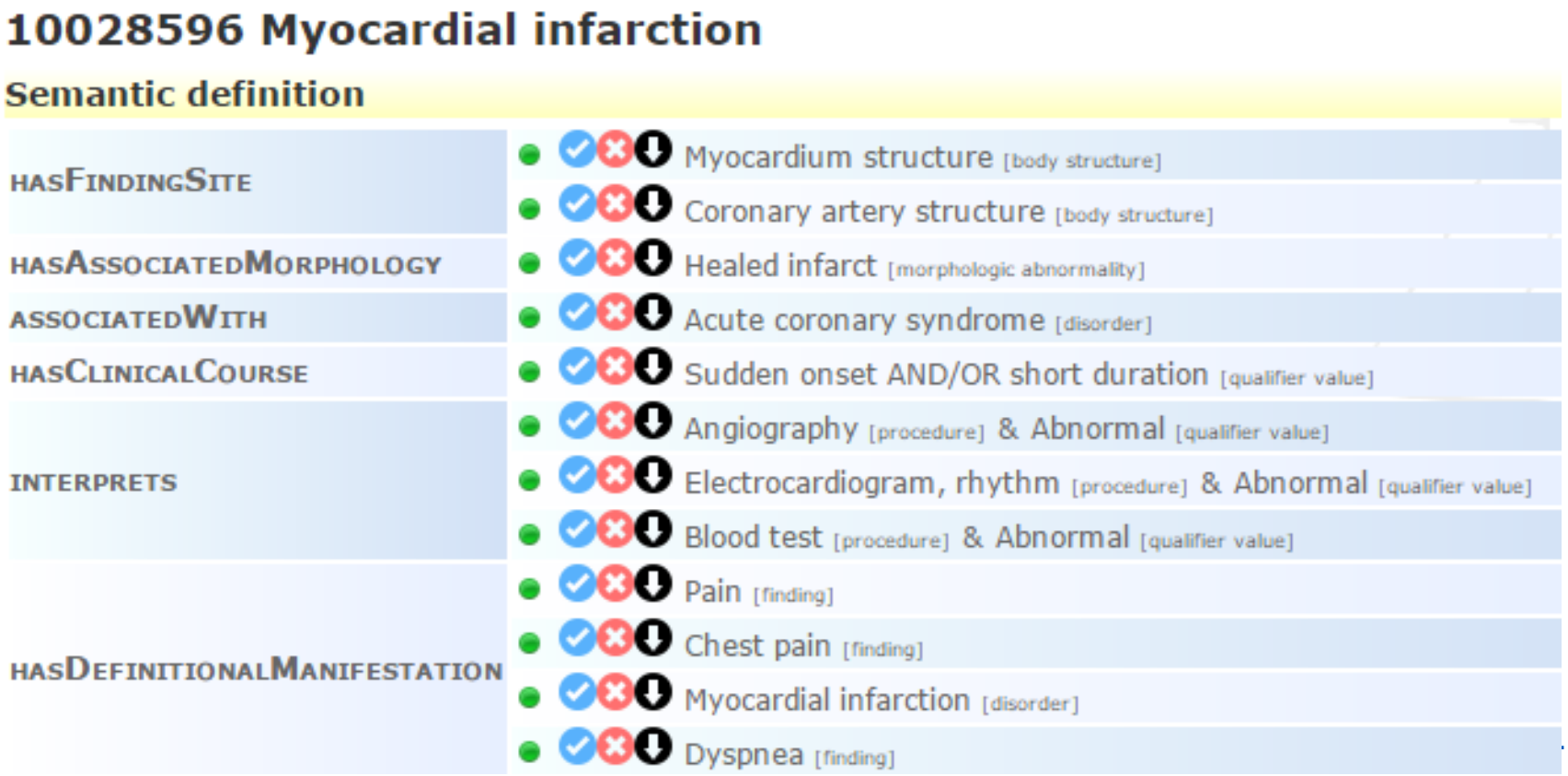
- In the context of computer and information sciences, an ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class 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



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

PROTECT

Example: Upper GI bleeding + occult blood + blood in stool

hasFindingSite some 'Upper gastrointestinal tract structure'

AND hasAssociatedMorphology some 'Hemorrhage'

OR

interprets some 'Occult blood screening'

AND hasInterpretation some 'Positive'

OR

interprets some 'Evaluation of stool specimen'

AND hasAssociatedMorphology some 'Hemorrhage'

Searching for upper GI bleeding with the query tool

SIMPLE QUERY

Finding Site

- [x] Upper digestive tract structure (body structure)

Associated Morphology

- [x] Hemorrhage (morphologic abnormality)

> Search <

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

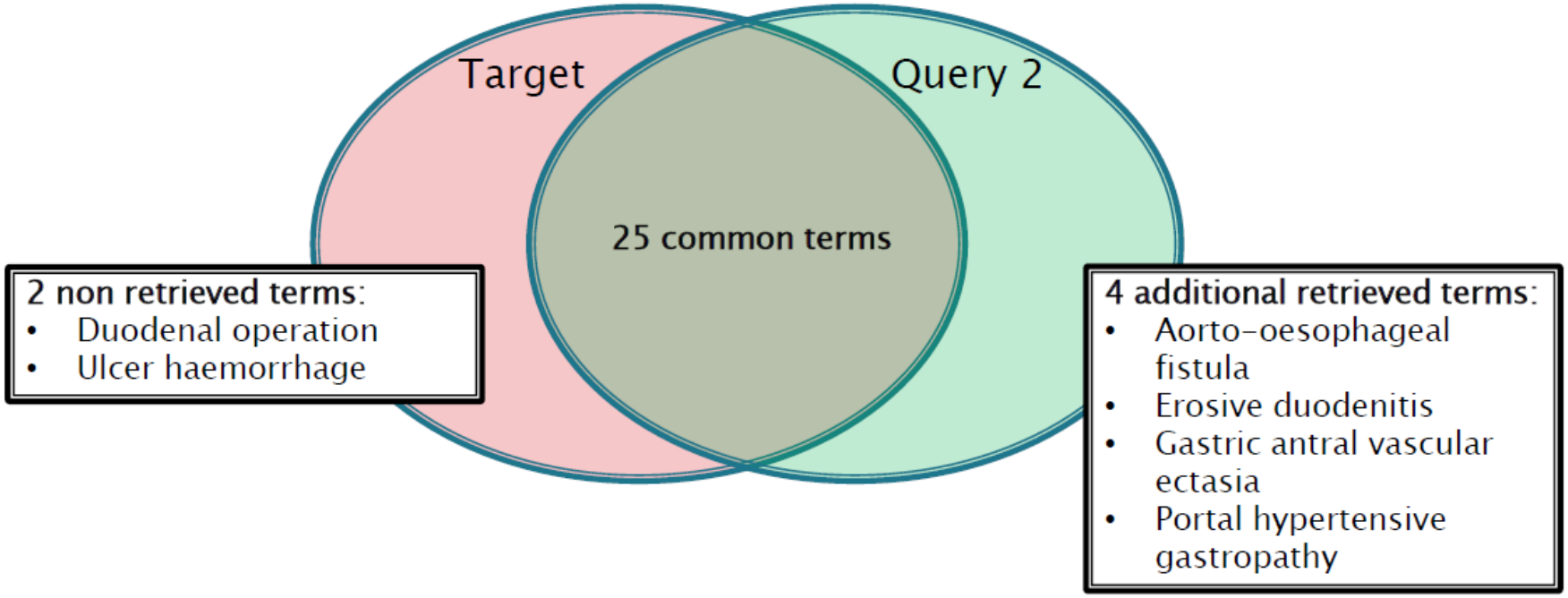
Body site affected by a condition

Manual search Type text here...

Browse Snomed-CT:

- [ ] Anatomical or acquired body structure (body structure)

Comparison with target SMQ



$$Recall = \frac{relevant\ terms \cap retrieved\ terms}{relevant\ terms} = 92.6\%$$

$$Precision = \frac{relevant\ terms \cap retrieved\ terms}{retrieved\ terms} = 86.2\%$$

Results for all 13 topics

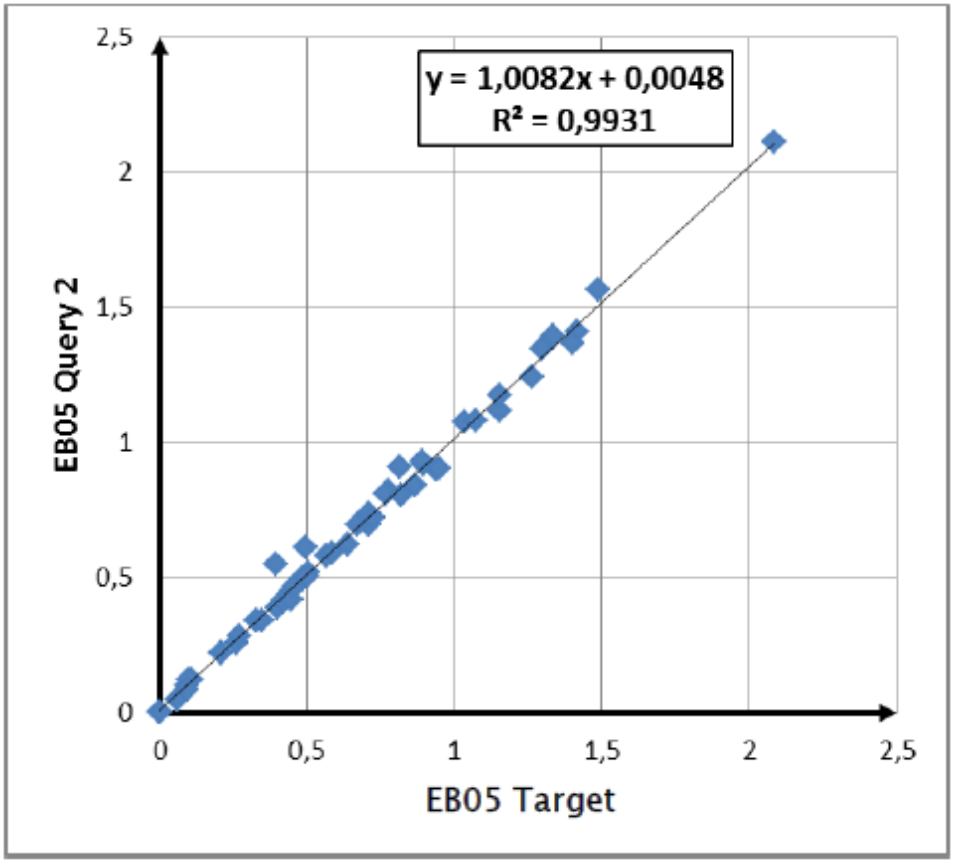
Safety topic	MedDRA gold standards		sem. match	RESULTS Recall	Precision	F-measure
1. Bullous eruptions	HLT	Bullous conditions	++	81,3%	33,3%	47,3%
2. Acute renal failure	SMQ	Acute renal failure	++	57,9%	51,2%	54,3%
	HLT	Renal failure and impairment	+	47,8%	55%	51,2%
3. Anaphylactic shock	SMQ	Anaphylactic/anaphylactoid shock	+	34,5%	38,5%	36,4%
	HLT	Anaphylactic responses	+	100%	100%	100%
4. Rhabdomyolysis	SMQ	Rhabdomyolysis/myopathy	+	46,7%	47,7%	47,2%
5. Aplastic anaemia / pancytopenia	SMQ	Cytopenia and haematopoietic...	+	55%	55%	55%
	HLT	Marrow depression and hypoplastic...	++	93,3%	70%	80%
6. Neutropenia	SMQ	Leukopenia **SELECT**	++	100%	63,6%	77,8%
	HLT	Neutropenias	++	100%	59,1%	74,3%
7. Cardiac valve fibrosis	HLGT	Cardiac valve disorders **SELECT**	++	100%	100%	100%
8. Extrapyramidal disorders	SMQ	Extrapyramidal syndrome	+	64,9%	48,5%	55,6%
	HLT	Dyskinesias and movement disorders	+	100%	71,4%	83,3%
9. Confusional state	SMQ	Noninfectious encephalopathy	+	16,0%	48,1%	24,1%
	HLT	Confusion and disorientation	++	100%	42,9%	60%
10. Thrombocytopenia	SMQ	Thrombocytopenia	++	100%	43,3%	60,5%
	HLT	Thrombocytopenias	++	100%	36,7%	53,7%
11. Upper gastrointestinal bleeding	SMQ	Gastrointestinal haemorrhage **SELECT**	++	92,6%	86,2%	89,3%
12. Peripheral neuropathy	SMQ	Peripheral neuropathy	++	66,7%	29%	40,4%
	HLGT	Peripheral neuropathies	++	98,9%	74,8%	85,2%
13. Maculo-papular erythematous eruptions	HLT	Rashes, eruptions and exanthems NEC	+	100%	20,4%	33,9%
MEAN VALUES	**(+) and (++) groupings**		**SMQ**	63,4%	51,1%	54,0%
			HLT/HLGT	92,8%	60,3%	69,9%
			TOT	**78,8%**	**55,9%**	**62,3%**
	only (++) groupings		**SMQ**	83,4%	54,7%	64,5%
			HLT/HLGT	96,2%	59,5%	71,5%
			TOT	**90,9%**	**57,5%**	**68,6%**

Recall, precision and F-measure rates of automatic OntoADR 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



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

- Evaluation of automated term groupings for detecting upper gastrointestinal bleeding signals with drugs J Souvignet, G Declerck, MC Jaulent, C Bousquet Drug Saf 35 (12), 1195-6

- 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