Terminology for the description of the diagnostic studies in the field of EBM

Natalia Grabar\textsuperscript{1}, Ludovic Trinquart\textsuperscript{2}, Isabelle Colombet\textsuperscript{3}

(1) CNRS UMR 8163 STL, Université Lille 3, 3 rue Barreau, 59653 Villeneuve d’Ascq, France
(2) French Cochrane Center, France; AP-HP, Paris, France
(3) Université Paris Descartes, Paris F-75006 France; HEGP AP-HP, 20 rue Leblanc, Paris F-75015 France
Plan of the presentation

- Context
- Material
- Methods
- Results and Discussion
- Conclusion and Perspectives
Systematic Reviews

- Systematic reviews:
  - provide a synthesis of multiple primary research studies

- Cochrane Collaboration effort
- Evidence-Based Medicine (EBM)
- Exploited by health professionals
  - decision-making support
- Several types of Systematic reviews:
  - therapeutic, preventative, diagnostic, prognostic
- Challenging process:
  - identification of the relevant studies
Steps of the Systematic Reviews

- Identification of the relevant studies
  1. exhaustive search strategy in several bibliographical databases
  2. scoping with the poor indexing of the diagnostic studies (Haynes & Wilczynski, 2005)
  3. methodological search filters not recommended (Leeflang et al, 2006)
    - because of the omission of an important number of relevant studies
  4. supervised machine learning not efficient
    - because of the small or non existing learning data

- Manual process

  – Help the process of the selection of the relevant studies
  – Semantic-based approach
    - terminology of the diagnostic studies
Terminology of the diagnostic studies

- Such resource does not exist
- Three kinds of approaches for the creation of terminologies:
  - top-down
    - main top concepts defined and then populated
    - modelization of the EBM area and its link with patient records (Pisanelli et al, 2003)
    - ontology of therapeutic studies (Zaveri et al, 2010)
      128 elements exploited for manual tagging
      interannotator agreement: 0.53-0.82
  - transformation-based approach
    - HTML, XML (Girardo et al, 2002)
    - databases (Krivine et al, 2009; Kamel et al, 2009)
  - bottom-up
    - terms observed within the textual material and organized into classes...
    - use of corpora and of the NLP methods
    - easily accessed and natural data
Material

- Corpora (scientific literature and reports)
  - diagnostic:
    - STARD initiative: 6 articles
    - diagnostic studies: 15 articles, 5 abstracts
    - 105,000 occurrences
  - prognostic: 6 citations (36,000 occ)
  - therapeutic: 7 citations (37,000 occ)
  - observational: 6 citations (40,000 occ)
- MeSH terminology
  - reference data
  - to be enriched?
Methods

Four-step method:

1. Extraction of terms from corpora
2. Alignment of the extracted terms with MeSH
3. Evaluation of the extracted terms
4. Structuring of terms
Methods

- **Extraction of terms**
  - Pre-processing and normalization of the corpora
    - segmentation into words and sentences
    - POS-tagging (cancers/Noun, cancerous/Adj)
    - lemmatization (cancers → cancer)
  - Rule-based term extractor YaTeA
    - syntactic patterns

- **Alignment with MeSH**
  - exact match
  - match of normalized forms
Methods

- Evaluation of the extracted terms
  - independent evaluation by two experts
    - physician, biostatistician with experience in SR
  - consensus in cases of disagreement
  - frequencies from the processed corpora
  - if high frequency in diagnostic corpus
    - high specificity of terms
  - global inter-expert agreement (Cicchetti & Feinstein, 1990):
    - chance-corrected kappa
    - simple raw specific agreement indexes

- Structuring of the extracted terms
  - bottom-up approach
  - categorization into categories, sub-categories, etc.
### Results and Discussion

<table>
<thead>
<tr>
<th>corpus</th>
<th>extracted</th>
<th>MeSH</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>diag</td>
<td>7,448</td>
<td>1,218</td>
<td>16.3</td>
</tr>
<tr>
<td>observ</td>
<td>1,640</td>
<td>722</td>
<td>44.0</td>
</tr>
<tr>
<td>progn</td>
<td>2,383</td>
<td>531</td>
<td>22.3</td>
</tr>
<tr>
<td>therap</td>
<td>1,602</td>
<td>590</td>
<td>36.8</td>
</tr>
</tbody>
</table>

- Terms from the diagnostic corpus are rare in MeSH
- 219 (3%) terms selected by experts
  - 26 (13%) in MeSH
    - E (n=11), G (n=2), N (n=11)
  - 193 provided by YaTeA
  - 36 terms added by experts
    - variations of the extracted terms
    - npv, ppv, dor, cut point ...
- 255 terms in total
Results and Discussion

- Important rejection rate (97%):
  - tagging errors of the NLP tools
  - terms corresponding to journals, authors...
  - too general terms (*public health, confidence interval*)
  - non diagnostic terms (*clinical trials*)

  \[ \Rightarrow \] Specificity of the task and shortcomings of NLP tools

- Inter-expert agreement (Cicchetti & Feinstein, 1990):
  - inter-expert agreement kappa: 0.106
    - selection rate low and heterogeneous
  - average positive agreement (selection): 0.14
  - average negative agreement (rejection): 0.84

- Possible source for the enrichment of MeSH
Five levels of terms

MeSH terms: under Test characteristics
Conclusion and Perspectives

- Experience on building of a terminology of diagnostic studies
- Exploitation of automatic NLP methods
- Creation of the semantic resource
- Small part of terms is already recorded in MeSH
  - Possible enrichment of MeSH
  - Better indexing of diagnostic studies
- Structuring of the extracted terms
- Exploitation for the selection of relevant citations