Quality assessment of automatically extracted data from GPs’ EPR

ACHIL research laboratory
Etienne De Clercq (UCL-IRSS), S. Moreels, V. Van Casteren, N. Bossuyt (ISP), G. Goderis, S. Bartholoméeusen (KUL)

(Etienne.DeClercq@UCLouvain.be)

MIE 2012 Congress
Pisa 26/08/12 – 29/08/12
ResoPrim Research Network

Architecture

Collecting

Analyzing

Disseminating

GP EPR

Scientific Institutions

- Scientific Institute of Public Health
- ACHIL satellites

National report
Publications
Etc.

Intermediate Level

Benchmarks

TSP: Trust Service Provider

ACHIL, funded by the National Institute for Health and Disability Insurance
Clinical automatic extracted data

- New coded and active **diagnosis** (ICPC2, ICD10, Belgian Thesaurus)
  (hypertension, diabetes type 2, cardiovascular past event)
- New coded and active **drug prescription** (ATC code)
  (anti-diabetic drugs, anti-hypertension drugs, aspirin, statin)
- New referral
- **Parameters** (2 most recent values extracted): height, weight, smoking status, syst. & diast. Blood pressure, total & LDL cholesterol
ResoPrim documented care

Documented care = Care * Quality of HRIS

Research Question:

Which are the properties of the HRIS?
Properties of the HRIS

PPV

“proportion of patients with a “gold standard” positive value of those with positive AE data”

* e.g.: proportion of drug codes (extracted from EPR) confirmed by the gold std?

Sensitivity

“proportion of patients with positive AE data of those with a “gold standard” positive value”

* e.g. How many patients with drug prescription (gold std) identified by a drug code (extracted from EPR)?
Building our “gold standard”

GPs’ consultation

EPR

Ref. DB

Research DB

Questionnaire

Source validation

ACHIL, funded by the National Institute for Health and Disability Insurance
Pilot Site

ResoPrim phase-2 pilot, (Summer 2007, +/- 7 weeks) data gathered from:

- 43 GP practices
- 4 software systems
- 10,307 patients
- 13,372 contacts
Questionnaire: missing & incoherent values

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>Min – Max (question)</th>
<th>Min – Max (practice)</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing</td>
<td>4.2%</td>
<td>0.5% - 5.7%</td>
<td>0% - 18.4%</td>
</tr>
<tr>
<td>(Diab.: 30%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>incoherence</td>
<td>5.3%</td>
<td>1.5% - 14%</td>
<td>0% - 14.3%</td>
</tr>
</tbody>
</table>

→ 90% seems an acceptable value for Sensitivity and PPV
### Automatic extracted diagnoses

<table>
<thead>
<tr>
<th>Diagnoses</th>
<th>Sensitivity</th>
<th>PPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension</td>
<td>45.5%</td>
<td>82.6%</td>
</tr>
<tr>
<td>Diabetes</td>
<td>47.9%</td>
<td>71.7%</td>
</tr>
<tr>
<td>Past Cardio-Vasc. Event</td>
<td>29.9%</td>
<td>57.5%</td>
</tr>
<tr>
<td>Family PCVE</td>
<td>1.2%</td>
<td>(50.0%)</td>
</tr>
</tbody>
</table>
### Automatic extracted drugs

#### Sensitivity and Positive Predictive Value (PPV) of automatic extraction vs questionnaire

<table>
<thead>
<tr>
<th>Drugs</th>
<th>Sensitivity</th>
<th>PPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-HT drugs</td>
<td>73.6%</td>
<td>97.8%</td>
</tr>
<tr>
<td>Anti-diab. Drugs</td>
<td>68.9%</td>
<td>96.3%</td>
</tr>
<tr>
<td>Aspirin</td>
<td>44.3%</td>
<td>91.9%</td>
</tr>
<tr>
<td>Statin</td>
<td>52.9%</td>
<td>90.4%</td>
</tr>
</tbody>
</table>
## Automatic extracted parameters

Sensitivity and Positive Predictive Value (PPV) of automatic extraction (most recent value) vs questionnaire

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Missing (AE)</th>
<th>Sensitivity (AE)</th>
<th>PPV (AE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypercholesterolemia</td>
<td>19.5%</td>
<td>59.1%</td>
<td>42.9%</td>
</tr>
<tr>
<td>(Smoking)</td>
<td>43.8%</td>
<td>(34.1%)</td>
<td>(82.2%)</td>
</tr>
<tr>
<td>Blood pressure &lt;140/90</td>
<td>6.1%</td>
<td>65.9%</td>
<td>92.5%</td>
</tr>
<tr>
<td>BMI&gt;25</td>
<td>36.9%</td>
<td>91.7%</td>
<td>73.8%</td>
</tr>
</tbody>
</table>
Lessons learned

**Drugs**: PPV ↑, Sensitivity ↓

**Diagnoses**: PPV ↓, Sensitivity ↓

**Parameters**: ??

BUT
Lessons learned (continued)

- Not representative GP sample!
- Great variations (missing, PPV, Sens.) by
  - Practice
  - Software system
- Robustness of the properties?
- Completeness of the extraction!
- Study restricted to coded and structured data
Our message …

Before any secondary usage of data extracted from EPR, we strongly advice assessing properties of the Health Research Information System.
That's all...