Data-Mining-Based Detection of Adverse Drug Events

Emmanuel CHAZARD, Cristian PREDA, Béatrice MERLIN, Grégoire FICHEUR, the PSIP consortium, Régis BEUSCART
Consortium: 13 partners

1/ **Hospitals**
   France, Denmark
   With / without CPOE

2/ **Industry:**
   Oracle, IBM, Medasys
   (CPOE editors)
   Vidal (pharmaceutical Kbase)

3/ **Academic teams**
   Data & Semantic mining,
   Decision Support Systems,
   Human Factors Engineering

Duration: 40 months
(Jan 08 → April 2011)
• Adverse Drug events (ADEs)
  – Almost 10% of stays
  – Less than 2% would be declarated
  – Responsible of 98000 deaths each year in USA

• Objective
  – Propose new methods to prevent ADEs
  – Develop automated rules to detect them
  – Integrate rules in a CDS system generating relevant alerts to the physician
Why use data-mining to detect ADEs?

• Chart review is time-consuming
• With data-mining, we are able to analyze 10,000 records in few minutes
• Data-mining may overcome the inevitable limits of expert knowledge: thousands ADEs described in the literature
• Data-mining may detect complex and sometimes combined ADEs that an expert may not necessarily identify
Data-mining based rules generation

**What we have**

- Data
- Events
  - potential causes and effects
- Statistical associations
  - effects linked to causes
- Drug linked events = rules
- Confidence of rules
  - In each medical department

**How we get it**

- Data aggregation
- Statistical analysis (trees…)
- Bibliographic analysis
- Evaluation of rules

2009-08-31
Main reasoning to detect ADE

- **ADE**: *injury caused by medical management rather than the underlying condition of the patient*

- The ADE event:
  - Is not declared in the stay
  - Requires a specific human case review
  - Is hidden in the data

- **Data-mining**: three-steps procedure
  - Identify a kind of traceable incident = “effect”
  - Automatically find a statistical association with some drugs in combination with other causes
  - The coincidence of causes and effects generate a rule which allow us to detect stays with ADE
Data-mining based rules generation

**What we have**

- Data
- Events
  - potential causes and effects
- Statistical associations
  - effects linked to causes
- Drug linked events = rules
- Confidence of rules
  - In each medical department

**How we get it**

- Data aggregation
- Statistical analysis (trees…)
- Bibliographic analysis
- Evaluation of rules
Data can be considered as:
- **Cause** or context of an ADE
- **Effect** which is a potential manifestation of an ADE

Example on diagnosis:
- **Cause**: chronic diseases, reason of the admission
- **Effect**: acute events during the stay

Example on lab results:
- **Cause**: abnormality existing at admission
- **Effect**: abnormality got during the stay

Example on drugs:
- **Cause**: prescription of the day
- **Effect**: antidotes…
Medical Data Bases
25,000 records

- Copenhagen Hospitals (University Hospitals, DK)
  - Cardiology & internal medicine: 2,700 records

- Rouen hospital (University Hospital, FR)
  - Cardiology & internal medicine: 800 records

- Denain hospital (General Hospital, FR)
  - Surgery: 2,600 records
  - Gynecology obstetrics: 1,800 records
  - Medicine A: 1,700 records
  - Medicine B: 900 records

- Lille hospital (University Hospital, FR)
  - Geriatry: 15,000 records
Variables have to be interpreted

Data:
values of natremia
min = 135
max = 145

Event:
hyponatremia = 1
start on day 2
stop on day 4
hyponatremia = 0
elsewhere
Example on drugs

Data:
- antiepileptic day 4
- antiepileptic day 5
- antiepileptic day 6

Event:
- antiepileptic = 1
  - start on day 4
  - stop on day 6
- antiepileptic = 0
  - before day 4
Data aggregation: overview

• Available data:
  – Complex data scheme with 7 tables, 91 fields
  – Potentially more than 30000 different variables
  – Too numerous and redundant codes
    • E.g. Diagnosis (ICD 10): 18 000 possible codes
    • E.g. Drugs (ATC): 5 400 possible codes
  • => Need to simplify the data

• Aggregated data:
  – One flat table containing one row per stay
  – 576 “cause or context” variables
  – 55 “possible effect” variables (lab values++)
Data-mining based rules generation

What we have

- Data
- Events
  - potential causes and effects
- Statistical associations
  - effects linked to causes
- Drug linked events = rules
- Confidence of rules
  - In each medical department

How we get it

- Data aggregation
- Statistical analysis (trees, …)
- Bibliographic analysis
- Evaluation of rules

2009-08-31
Data Mining Methods

• General principle of our analysis:
  1) investigate the causes associated with the effect
  2) Is there any drug among these causes?

• Statistical methods already used in PSIP:
  – Regression Trees (CART)
  – Multiple Correspondence Analysis
  – Logistic Regression Analysis
  – Principal Component Analysis
  – Association Rules
Appearance of a too low INR (INR < 2)
Risk of Thrombosis

- Ex. of tree obtained from the method CART (classification and regression tree)
- We are you looking for the variables (causes) associated with an effect
- Can be read at each node:
  - the name of each variable used in the regression
  - The confidence at this level of the regression
Appearance of a too low INR (patient with anticoagulation) - Rule N°1

Rule enunciation:
Lab(previous too high INR)=1 & MedInfo(age)>78.68 & Lab(previous hypoalbuminemia)=1 ⇒ Appearance of a too low INR

Rule characteristics:
Support: 6
Confidence: 86%

7 stays match the conditions, 6 of them present the effect (86%=6/7)

Outcomes:
0% death
avg duration: 13.4 days
Appearance of a too low INR (patient with anticoagulation) - Rule N°2

Rule enunciation:
Lab(previous too high INR)=0
& Drug(vitamin K antagonist)=1
& Drug(prokinetic)=1
⇒ Appearance of a too low INR

Rule characteristics:
Support: 4
Confidence: 67%

6 stays match the conditions, 4 of them present the effect (67%=4/6)

Outcomes:
16.67% death
avg duration: 15 days
## Data-mining based rules generation

### What we have
- Data
- **Events**
  - potential causes and effects
- **Statistical associations**
  - effects linked to causes
- **Drug linked events = rules**
- **Confidence of rules**
  - In each medical department

### How we get it
- Data aggregation
- **Statistical analysis** (trees, …)
- **Bibliographic analysis**
- **Evaluation of rules**
Too high INR means Hypocoagulation (risk of bleeding): INR > 5

Interpretation: When a patient is admitted for a too high INR (risk of bleeding), if age > 78 and hypoalbuminemia, then a too low INR (risk of thrombosis) appears with a 86% probability.

Outcomes:
0% death
avg duration: 13.4 days
Appearance of a too low INR (patient with anticoagulation) - Rule N°2

Rule enunciation:
Lab(previous too high INR)=0 & Drug(vitamin K antagonist)=1 & Drug(prokinetic)=1 ⇒ Appearance of a too low INR

Rule characteristics:
Support: 4
Confidence: 67%

INR at entry is correct and patients have Vitamin K antagonists

When they are prescribed simultaneously a prokinetic drug, the risk of a too low INR (and consequently the risk of thrombosis) is 67%
Data-mining based rules generation

What we have

- Data
- Events
  - potential causes and effects
- Statistical associations
  - effects linked to causes

Drug linked events = rules

How we get it

- Data aggregation
- Statistical analysis (trees,…)
- Bibliographic analysis
- Evaluation of rules

Confidence of rules

In each medical department
### Interest of multi-site data-mining and rules evaluation

<table>
<thead>
<tr>
<th>Rules of detection</th>
<th>Copenhagen</th>
<th>Denain Medicine A</th>
<th>Denain Medicine B</th>
<th>Denain Surgery</th>
<th>Rouen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug(diuretic)=1 &amp; Drug(angiotensin conversion enzyme inhibitor)=1 =&gt; Acute renal failure</td>
<td>26/341=8%</td>
<td>63/304=21%</td>
<td>25/181=14%</td>
<td>23/145=16%</td>
<td>4/152=3%</td>
</tr>
<tr>
<td>Drug(beta lactams[antibiotic])=1 &amp; Drug(vitamin K antagonist (antithrombotic))=1 =&gt; Appearance of high inr</td>
<td>2/9=22%</td>
<td>9/35=26%</td>
<td>9/54=17%</td>
<td>0/8=0%</td>
<td>6/31=19%</td>
</tr>
<tr>
<td>Drug(paracetamol[anilides])=1 &amp; Lab(kidney insufficiency)=1 &amp; Drug(vitamin K antagonist (antithrombotic))=1 =&gt; Appearance of high inr</td>
<td>2/16=13%</td>
<td>5/22=23%</td>
<td>7/21=33%</td>
<td>0/10=0%</td>
<td>4/37=11%</td>
</tr>
</tbody>
</table>
• 242 rules of detection have already been validated
• We compared our rules with other available rules databases (Vidal, David Bates' team): several rules in common
• 2 chart reviews in Denain and Copenhagen to evaluate the ability of rules to detect stays with ADEs
• Implementation of rules in CDSS in progress
Thank you for your attention