Détection automatique de signaux en pharmacovigilance
Approche statistique fondée sur les comparaisons multiples

Ismaïl Ahmed

INSERM UMRS Centre de recherche en Epidémiologie et Santé des Populations - Villejuif

16 novembre 2011
Pharmacovigilance systems

- aim at early detection of adverse drug reaction of marketed drugs
- based on spontaneous reports

The French pharmacovigilance system

- set up in 1979
- currently based on 31 pharmacovigilance centres (CRPV)
- coordinated by the pharmacovigilance unit of the « Agence Française de Sécurité Sanitaire des Produits de Santé » (Afssaps)
Introduction (2)

Large quantity of data

- In France
  - 2000 - 2010: 370,000 spontaneous reports
  - More than 30,000 reports per year
- FDA: 2.6 million, WHO: 3.7 million in 2005

Automatic signal detection methods

- Statistical associations within the pharmacovigilance database
- Potential adverse drug reactions
Existing methods

- Based on different probability models and decision rules

-**Frequentist methods:**
  - Reporting Odds Ratio (ROR, Netherlands)
  - Proportional Reporting Ratio (PRR, UK and EudraVigilance)

-**Bayesian methods:**
  - Gamma Poisson Shrinkage (GPS, USA)
  - Bayesian Confidence Propagation Neural Network (BCPNN, WHO)

France does not use any automatic signal detection method yet
Introduction (4)

Main limit of the existing methods

- Detection rules arbitrarily chosen

Main objective

- Decision rule derived from a statistical error criteria
- Accounting for the multiple comparisons
- False discovery rate (FDR)
Outline

Description of the current methods

Extension to the multiple comparison setting

Simulation study

Application to the French pharmacovigilance data

Discussion
Pharmacovigilance data structure and notations

Large contingency table crossing all the drugs and all the adverse events

- French database (ATC5-HLT)
  - 672 classes of drugs (D) × 820 classes of adverse events (AE)
  - 551,040 cells of which 80% are empty

For a particular couple (AE\_i, D\_j)

<table>
<thead>
<tr>
<th>Adverse event _i</th>
<th>Drug _j</th>
<th>Other Drugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_{ij} )</td>
<td>( n_{i\bar{j}} )</td>
<td>( n_{i.} )</td>
</tr>
<tr>
<td>( n_{\bar{i}j} )</td>
<td>( n_{\bar{i}\bar{j}} )</td>
<td>( n_{\bar{i}.} )</td>
</tr>
<tr>
<td>( n_{.j} )</td>
<td>( n_{.\bar{j}} )</td>
<td>( n )</td>
</tr>
</tbody>
</table>

- \( n_{ij} \): Number of reports involving AE\_i and D\_j
- \( n_{i.} \): Marginal count involving AE\_i
- \( n_{.j} \): Marginal count involving D\_j
- \( n \): Total number of AE-D pairs counts
**Reporting Odds Ratio (ROR)** van Puijenbroek et al. (2002)

**Association measure**

- For the adverse event-drug pair \((i, j)\)

\[
\hat{\psi}_{ij} = \frac{n_{ij} n_{i\overline{j}}}{n_{\overline{i}j} n_{i\overline{j}}}
\]

**Signal generation**

- \(\ln(\hat{\psi}_{ij})\) is assumed to follow a normal distribution with variance:

\[
\text{Var}\{\ln(\hat{\psi}_{ij})\} = \frac{1}{n_{ij}} + \frac{1}{n_{\overline{i}j}} + \frac{1}{n_{\overline{i}j}} + \frac{1}{n_{i\overline{j}}}
\]

- A signal is generated if

\[
\ln(\hat{\psi}_{ij}) - 1.96 \sqrt{\text{Var}\{\ln(\hat{\psi}_{ij})\}} > 0
\]
Proportional Reporting Ratio (PRR)  
Evans et al. (2001)

Association measure

- For the adverse event-drug pair \((i,j)\)

\[
\hat{\phi}_{ij} = \frac{n_{ij}}{n_i} \frac{n_{\bar{i}j}}{n_{\bar{i}}}
\]

- very close to \(\hat{\psi}_{ij}\)

Signal generation

- Evans et al. (2001)
  - \(\hat{\phi}_{ij} > 2\)
  - \(n_{ij} \geq 3\)
  - A \(\chi^2\) statistic with 1 df \(\geq 4\)

- van Puijenbroek et al. (2002) : same as for the ROR method
**Gamma Poisson Shrinkage (GPS)** DuMouchel (1999)

### Poisson - 2 gamma mixture model

- $n_{ij} | e_{ij}, \lambda_{ij} \sim P_{n}(\lambda_{ij} e_{ij})$ avec $e_{ij} = \frac{n_i \cdot n_j}{n}$

- $\lambda_{ij} \sim \hat{w} \ Ga(\hat{\alpha}_1, \hat{\beta}_1) + (1 - \hat{w}) \ Ga(\hat{\alpha}_2, \hat{\beta}_2)$

  where $(\hat{w}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2)$ maximizes the marginal likelihood

$$\Pi_{ij} \left[ w \ f_{Bn}\{n_{ij}; \alpha_1, \beta_1/(\beta_1 + e_{ij})\} + (1 - w) \ f_{Bn}\{n_{ij}; \alpha_2, \beta_2/(\beta_2 + e_{ij})\} \right]$$

### Association measure

$$\lambda_{ij}^* = \lambda_{ij} | n_{ij}, e_{ij}$$

- $\lambda_{ij}^* \sim w_{ij} \ Ga(\hat{\alpha}_1 + n_{ij}, \hat{\beta}_1 + e_{ij}) + (1 - w_{ij}) \ Ga(\hat{\alpha}_2 + n_{ij}, \hat{\beta}_2 + e_{ij})$

Poisson - 2 gamma mixture model

\( n_{ij} | e_{ij}, \lambda_{ij} \sim \text{Pn}(\lambda_{ij}, e_{ij}) \) avec \( e_{ij} = \frac{n_i \cdot n_j}{n} \)

\( \lambda_{ij} \sim \hat{w} \text{Ga}(\hat{\alpha}_1, \hat{\beta}_1) + (1 - \hat{w}) \text{Ga}(\hat{\alpha}_2, \hat{\beta}_2) \)

where \((\hat{w}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2)\) maximizes the marginal likelihood

\[
\Pi_{ij} \left[ w \cdot f_{\text{Bn}}\left\{ n_{ij}; \alpha_1, \beta_1 / (\beta_1 + e_{ij}) \right\} + (1 - w) \cdot f_{\text{Bn}}\left\{ n_{ij}; \alpha_2, \beta_2 / (\beta_2 + e_{ij}) \right\} \right]
\]

Association measure

\( \lambda^{*}_{ij} = \lambda_{ij} | n_{ij}, e_{ij} \)

\( \lambda^{*}_{ij} \sim w_{ij} \text{Ga}(\hat{\alpha}_1 + n_{ij}, \hat{\beta}_1 + e_{ij}) + (1 - w_{ij}) \text{Ga}(\hat{\alpha}_2 + n_{ij}, \hat{\beta}_2 + e_{ij}) \)

Signal generation

\( \mathbb{E}\{\log_2(\lambda^{*}_{ij})\} \) 

DuMouchel (1999)

Poisson - 2 gamma mixture model

\( n_{ij} | e_{ij}, \lambda_{ij} \sim Pn(\lambda_{ij}, e_{ij}) \) avec \( e_{ij} = \frac{n_i \cdot n_j}{n} \)

\( \lambda_{ij} \sim \hat{w} \text{Ga}(\hat{\alpha}_1, \hat{\beta}_1) + (1 - \hat{w}) \text{Ga}(\hat{\alpha}_2, \hat{\beta}_2) \)

where \((\hat{w}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2)\) maximizes the marginal likelihood

\[
\prod_{ij} \left[ w f_{Bn}\{n_{ij}; \alpha_1, \beta_1/(\beta_1 + e_{ij})\} + (1 - w) f_{Bn}\{n_{ij}; \alpha_2, \beta_2/(\beta_2 + e_{ij})\} \right]
\]

Association measure

\( \lambda^*_{ij} = \lambda_{ij} | n_{ij}, e_{ij} \)

\( \lambda^*_{ij} \sim w_{ij} \text{Ga}(\hat{\alpha}_1 + n_{ij}, \hat{\beta}_1 + e_{ij}) + (1 - w_{ij}) \text{Ga}(\hat{\alpha}_2 + n_{ij}, \hat{\beta}_2 + e_{ij}) \)

Signal generation

\( E\{\log_2(\lambda^*_{ij})\} \)

\( Q_{0.05}(\lambda^*_{ij}) \)

DuMouchel (1999)

DuMouchel et al. (2001)

Poisson - 2 gamma mixture model

- $n_{ij} | e_{ij}, \lambda_{ij} \sim \mathcal{Pn}(\lambda_{ij}, e_{ij})$ avec $e_{ij} = \frac{n_i \cdot n_j}{n}$
- $\lambda_{ij} \sim \hat{w} \text{Ga}(\hat{\alpha}_1, \hat{\beta}_1) + (1 - \hat{w}) \text{Ga}(\hat{\alpha}_2, \hat{\beta}_2)$

where $(\hat{w}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2)$ maximizes the marginal likelihood

$$\prod_{ij} [w \ f_{\mathcal{Bn}}\{n_{ij}; \alpha_1, \beta_1/(\beta_1 + e_{ij})\} + (1 - w) \ f_{\mathcal{Bn}}\{n_{ij}; \alpha_2, \beta_2/(\beta_2 + e_{ij})\}]$$

Association measure

$$\lambda_{ij}^* = \lambda_{ij} | n_{ij}, e_{ij}$$

- $\lambda_{ij}^* \sim w_{ij} \text{Ga}(\hat{\alpha}_1 + n_{ij}, \hat{\beta}_1 + e_{ij}) + (1 - w_{ij}) \text{Ga}(\hat{\alpha}_2 + n_{ij}, \hat{\beta}_2 + e_{ij})$

Signal generation

- $E\{\log_2(\lambda_{ij}^*)\}$
- $Q_{0.05}(\lambda_{ij}^*)$
- $Q_{0.05}(\lambda_{ij}^*) > 2$

DuMouchel (1999)  
DuMouchel et al. (2001)  
Szarfman et al. (2002)
Bayesian Confidence Propagation Neural Network (BCPNN) (1)

Bate et al. (1998), Noren et al. (2006)

Multinomial-Dirichlet model

\[
(n_{ij}, n_{i\bar{j}}, n_{i\bar{j}}, n_{\bar{i}j}) \sim \text{Mu}(n, p_{ij}, p_{i\bar{j}}, p_{\bar{i}j}, p_{\bar{i}\bar{j}})
\]

with \( (p_{ij}, p_{i\bar{j}}, p_{\bar{i}j}, p_{\bar{i}\bar{j}}) \sim \text{Di}(\alpha_{ij}, \alpha_{i\bar{j}}, \alpha_{i\bar{j}}, \alpha_{\bar{i}\bar{j}}) \)

- The hyperparameters depend on the cell counts
- The posterior distribution of \((p_{ij}, p_{i\bar{j}}, p_{\bar{i}j}, p_{\bar{i}\bar{j}})\) is also a Dirichlet:

\[
(p_{ij}, p_{i\bar{j}}, p_{\bar{i}j}, p_{\bar{i}\bar{j}})^* \sim \text{Di}(\gamma_{ij}, \gamma_{i\bar{j}}, \gamma_{i\bar{j}}, \gamma_{\bar{i}\bar{j}})
\]

with \( \gamma_{kl} = \alpha_{kl} + n_{kl} \)
Bayesian Confidence Propagation Neural Network (BCPNN) (1)

Bate et al. (1998), Noren et al. (2006)

Multinomial-Dirichlet model

\[(n_{ij}, n_{i\bar{j}}, n_{\bar{i}j}, n_{\bar{i}\bar{j}}) \sim \text{Mu}(n, p_{ij}, p_{i\bar{j}}, p_{\bar{i}j}, p_{\bar{i}\bar{j}})\]

with\[(p_{ij}, p_{i\bar{j}}, p_{\bar{i}j}, p_{\bar{i}\bar{j}}) \sim \text{Di}(\alpha_{ij}, \alpha_{i\bar{j}}, \alpha_{\bar{i}j}, \alpha_{\bar{i}\bar{j}})\]

- The hyperparameters depend on the cell counts
- The posterior distribution of \((p_{ij}, p_{i\bar{j}}, p_{\bar{i}j}, p_{\bar{i}\bar{j}})\) is also a Dirichlet:

\[\left(p_{ij}, p_{i\bar{j}}, p_{\bar{i}j}, p_{\bar{i}\bar{j}}\right)^* \sim \text{Di}(\gamma_{ij}, \gamma_{i\bar{j}}, \gamma_{\bar{i}j}, \gamma_{\bar{i}\bar{j}})\]

with \(\gamma_{kl} = \alpha_{kl} + n_{kl}\)

- In particular

\[p_{ij}^* \sim \text{Be}(\gamma_{ij}, \gamma_{i\bar{j}} + \gamma_{i\bar{j}} + \gamma_{\bar{i}\bar{j}})\]

\[p_i^* = p_{ij}^* + p_{i\bar{j}}^* \sim \text{Be}(\gamma_{ij} + \gamma_{i\bar{j}}, \gamma_{i\bar{j}} + \gamma_{\bar{i}\bar{j}})\]

\[p_j^* = p_{ij}^* + p_{\bar{i}j}^* \sim \text{Be}(\gamma_{ij} + \gamma_{\bar{i}j}, \gamma_{\bar{i}j} + \gamma_{\bar{i}\bar{j}})\]
Bayesian Confidence Propagation Neural Network (BCPNN) (2)

Bate et al. (1998), Noren et al. (2006)

Association measure

\[
IC_{ij}^* = \log_2 \left( \frac{\frac{p_{ij}^*}{p_{i.}^*}}{\frac{p_{.j}^*}{p_{.j}^*}} \right)
\]

Ratio of beta distributions \(\Rightarrow\) No analytic form

Signal generation

\[
Q_{0.025}(IC_{ij}^*) > 0
\]

- Normal approximation
  - delta method: Bate et al. (1998)
- Interpolation model built from Monte Carlo simulations: Noren et al. (2006)
Description of the current methods

Extension to the multiple comparison setting

Simulation study

Application to the French pharmacovigilance data

Discussion
False Discovery Rate and Pharmacovigilance

- Automatic signal detection methods are data mining tools
- Extension to the hypothesis testing framework
  - relying on the recent developments in multiple comparison statistical field
  - detection thresholds rule based on statistical criteria
- False Discovery Rate (Benjamini and Hochberg (1995))
  - $E(\text{proportion of false discoveries among the generated signals})$
  - used in the genomic data analysis
  - adapted to massive comparisons and exploratory analysis
Frequentist methods: Proposed approach (1)

Use of the P-values as statistic of interest

ROR

1. For each cell, we want to test $H_{0ij}: \psi_{ij} \leq \psi_0$

2. The corresponding P-values

\[
p_{ij} = 1 - \Phi \left( \frac{\ln(\hat{\psi}_{ij}) - \ln(\psi_0)}{\text{var}[\ln(\hat{\psi}_{ij})]^{1/2}} \right)
\]

where $\Phi$ denotes the standard normal cdf

3. The current decision rule corresponds to
   - choose $\psi_0 = 1$ and
   - generate signals for cells with $p_{ij} \leq 0.025$

PRR

1. $H_{0ij}: \phi_{ij} \leq \phi_0$
Frequentist methods: Proposed approach (2)

Fisher’s exact test

- Simple and exact alternative to ROR and PRR
- Large proportion of cells with small counts
- $P$-values: $\Pr(N_{ij} \geq n_{ij} | n_i, n_j, n; \psi_0)$
- The Fisher’s exact test is known to be conservative
- mid-$P$-values: $\overbrace{\Pr(N_{ij} \geq n_{ij} | n_i, n_j, n; \psi_0) - \frac{1}{2} \Pr(N_{ij} = n_{ij} | n_i, n_j, n; \psi_0)}^{P$-values}$
Frequentist methods: Proposed approach (3)

Marginal distribution of the P-values

P-values are assumed to follow a mixture of two distributions

\[ f(p) = \pi_0 f_0(p) + (1 - \pi_0) f_1(p) \]

- \( f_0(p) \) is the pdf of \( p \) under the null hypothesis
- \( f_1(p) \) is the pdf of \( p \) under the alternative hypothesis

FDR Storey et al. (2002)

For a P-value rejection region \([0, \gamma]\) with \( \gamma \in [0, 1] \)

\[ FDR(\gamma) = \frac{\pi_0 F_0(\gamma)}{F(\gamma)} \]
Frequentist methods: Proposed approach (4)

FDR estimation: Single null hypothesis (not our case)

- The P-values are uniformly distributed under the null hypothesis
  \[ f(p) = \pi_0 f_0(p) + (1 - \pi_0) f_1(p) \]
  \[ = \pi_0 + (1 - \pi_0) f_1(p) \]

- For a P-value rejection region \([0, \gamma]\) with \(\gamma \in ]0, 1]\)
  \[ \text{FDR}(\gamma) = \frac{\pi_0 F_0(\gamma)}{F(\gamma)} = \frac{\pi_0 \gamma}{F(\gamma)} \]

- \(F\) is estimated by its empirical estimate: \(\hat{F}(\gamma) = 1/m \sum_{ij} 1(p_{ij} \leq \gamma)\)

- The main difficulty is to estimate \(\pi_0\)
Frequentist methods: Proposed approach (5)

Location Based Estimator (LBE) Dalmasso et al. 2005

$$E[\varphi(P)] = \pi_0 E_0[\varphi(P)] + (1 - \pi_0) E_1[\varphi(P)]$$

$$\frac{E[\varphi(P)]}{E_0[\varphi(P)]} = \pi_0 + (1 - \pi_0) \frac{E_1[\varphi(P)]}{E_0[\varphi(P)]}$$

non negative term: Bias

- $\varphi$ is chosen to minimize the non negative term: $\varphi(p) = - \ln(1 - p)^a$

$$\hat{\pi}_0 = \frac{\hat{E}[\varphi(P)]}{E_0[\varphi(P)]} = \frac{1/m \sum_{ij} \varphi(p_{ij})}{E_0[\varphi(P)]} = \frac{1/m \sum_{ij} \{- \ln(1 - p_{ij})\}^a}{\Gamma(a + 1)}$$

- LBE only assumes that $f$ is a non increasing function
- $\hat{\pi}_0$ is a biased estimator $\Rightarrow$ upper bound for the FDR estimation
Frequentist methods: Proposed approach (6)

FDR estimation: One-sided null hypothesis

- The P-values are not calculated under $H_{0_{ij}} : \psi_{ij} \leq \psi_0$ but under $H_{0_{ij^*}} : \psi_{ij} = \psi_0$
- The P-values are not uniformly distributed under $H_{0_{ij}}$
- $f_0$ expressed as a mixture of a uniform $f_{0^*}$ and a non-decreasing $f_{1^*}$ function

\[
f(p) = \pi_0 f_0(p) + (1 - \pi_0) f_1(p)
= \pi_0 \{\pi_{0^*} + (1 - \pi_{0^*}) f_{1^*}(p)\} + (1 - \pi_0) f_1(p)
\]

- In practice, only small values of $\gamma$ are of interest:

\[
\text{FDR}(\gamma) = \frac{\pi_0 F_{0}(\gamma)}{F(\gamma)} = \frac{\pi_0 \{\pi_{0^*} \gamma + (1 - \pi_{0^*}) F_{1^*}(\gamma)\}}{F(\gamma)} = \frac{\pi_0 \pi_{0^*} \gamma}{F(\gamma)}
\]

- The issue is thus to estimate $\pi_0 \pi_{0^*}$

- LBE on $P^* = 1 - 2|P - \frac{1}{2}|$ which has a non-increasing pdf expressed as

\[
f_{P^*}(p^*) = \pi_0 \pi_{0^*} + (1 - \pi_0 \pi_{0^*}) f_{1^*}(P^*(p^*))
\]
Bayesian methods: Proposed approach (1)

Based on the Bayesian decision theory framework - Müller et al. (2004)

**Status**

\[ z_{ij} \in \{0, 1\} \]

**Decision**

\[ d_{ij} \in \{0, 1\} \]

**FDR**

\[
FDP = \frac{\sum_{ij} (1 - z_{ij}) d_{ij}}{\sum_{ij} d_{ij}} \quad \rightarrow \quad FDR = E[FDP]
\]

**Bayesian FDR estimation: FDR**

\[
FDR^* = E[FDP|data] = \frac{\sum_{ij} u_{ij} d_{ij}}{\sum_{ij} d_{ij}}
\]

where \( u_{ij} = \Pr(z_{ij} = 0|data) \) i.e. the posterior Pr. of \( H_{0_{ij}} : z_{ij} = 0 \)
Bayesian methods: Proposed approach (2)

Status $z_{ij} - u_{ij}$

- For each cell, we want to test
  - GPS $\rightarrow H_{0ij} : \lambda_{ij} \leq RR_0$
  - BCPNN $\rightarrow H_{0ij} : \frac{p_{ij}}{p_i p_j} \leq RR_0$

- and thus to calculate $u_{ij}$ i.e. the posterior probability of $H_{0ij}$

- GPS:
  \[
  \Pr(\lambda_{ij}^* \leq RR_0) = w_{ij} F_{Ga}(RR_0; \hat{\alpha}_1 + n_{ij}, \hat{\beta}_1 + e_{ij}) + (1 - w_{ij}) F_{Ga}(RR_0; \hat{\alpha}_2 + n_{ij}, \hat{\beta}_2 + e_{ij})
  \]

- BCPNN:
  \[
  \Pr(\text{IC}^*_{ij} \leq \log_2(RR_0))
  \]
  No analytic form $\rightarrow$ Monte Carlo simulations
Bayesian methods: Proposed approach (3)

Decision rule $d_{ij}$

- **GPS**
  
  DuMouchel (1999): $d_{ij} = 1[E\{\log_2(\lambda^*_{ij})\} > \tau]$
  
  DuMouchel et al. (2001): $d_{ij} = 1[Q_{0.05}(\lambda^*_{ij}) > \tau]$
  
  Our suggestion:
  
  Szarfman et al. (2002): $RR_0 = 2$ and $\alpha = 0.05$

- **BCPNN**
  
  Our suggestion: $d_{ij} = 1[u_{ij} \leq \alpha]$
  
  Bate et al. (1998): $RR_0 = 1$ and $\alpha = 0.025$
Description of the current methods

Extension to the multiple comparison setting

Simulation study

Application to the French pharmacovigilance data

Discussion
Simulation study

Objectives

- To compare the performances of the methods
- To evaluate the quality of the FDR estimators

Methods

- Frequentist methods :
  - ROR « new »
  - RFET
  - midRFET

- Bayesian methods :
  - GPS : $E\{\log_2(\lambda_{ij}^*)\}$, $Q_{0.05}(\lambda_{ij}^*)$, $Pr(H_{0ij}^*)$
  - BCPNN : $Pr(H_{0ij}^*)$

- 3 different tested hypotheses based on $\{\psi_0, RR_0\} = 1, 2$ and $5$
Simulation study

Data generation

- Model: \( n_{ij} \sim \text{Mu}(n, p_{ij}) \)

- \( p_{ij} \)
  - \( p_{i.}^w \sim \text{Di}(n_{i.}) \)
  - \( p_{.j}^w \sim \text{Di}(n_{.j}) \)
  - \( \log(r_{ij}^w) \sim \text{Lo}(0, 0.5) \)

- \( p_{ij} = \frac{r_{ij}^w p_{i.}^w p_{.j}^w}{\sum_{ij} r_{ij}^w p_{i.}^w p_{.j}^w} \)

Simulation plan

- 500 simulated datasets

- « True » FDR estimated by the average of the FDPs over the 500 datasets

- \( n_{ij} \geq 3 \)
Simulation study: Comparison of the frequentist methods

FDR

ψ₀ = 1

ψ₀ = 2

ψ₀ = 5

ROR

RFET

midRFET
Simulation study: Comparison of the frequentist methods

FDR and estimation

\[ \psi_0 = 1 \]

\[ \psi_0 = 2 \]

\[ \psi_0 = 5 \]

ROR

RFET

midRFET

Average number of generated signals

0 5000 10000 15000 20000
0.00 0.05 0.10 0.15 0.20

Average number of generated signals

0 1000 2000 3000 4000 5000 6000 7000
0.00 0.05 0.10 0.15 0.20

Average number of generated signals

0 500 1000 1500
0.00 0.05 0.10 0.15 0.20

Average number of generated signals

0 1000 1500 2000 2500 3000 3500 4000
0.00 0.05 0.10 0.15 0.20
Simulation study : Comparison of the Bayesian methods (1)

GPS and decision rules

- The proposed decision rule gives better results according to the FDR
- Close performances between $Q_{0.05}(\lambda_{ij}^*)$ and $Pr(H_{0,ij}^*)$ for small values of FDR
GPS vs BCPNN with $\Pr(H_{0ij}^*)$

FDR and estimation

- Very close performances according to the FDR
GPS vs BCPNN with \( \Pr(H^*_{0ij}) \)

### FDR and estimation

- Very close performances according to the FDR
- Large differences in the FDR estimation
Global Comparison

- Very close performances according to the FDR
- Large differences in FDR estimation
Global Comparison

ψ₀ = 1, RR₀ = 1

ψ₀ = 2, RR₀ = 2

ψ₀ = 5, RR₀ = 5

Average number of generated signals

Average number of generated signals

Average number of generated signals

- Very close performances according to the FDR
- Large differences in FDR estimation

Results in favor of the use of the GPS model with $\Pr(H_{0_{ij}}^*)$
Description of the current methods

Extension to the multiple comparison setting

Simulation study

Application to the French pharmacovigilance data

Discussion
Global analysis (1)

- 1984-2003
- $n_{ij} \geq 3 \rightarrow 47\,520$ cells

FDR estimation

Results close to that obtained in the simulation study
Feasibility Study

Afssaps is interested in implementing an automatic signal detection tool in France

Objective

A feasibility study was recently conducted with two aims

- Relevance of the signals generated by the method
- Time necessary for the evaluation of the signals
Feasibility Study

Methodology

- Use of $\text{GPS}_{\text{pH0}}$ with FDR at 5%
- A first analysis was performed on the data (Jan 1 2000 - dec 31 2008)
- A second analysis was performed three months later (Jan 1 2000 - March 31 2009)
- Selection of the 1414 signals that were highlighted by the second analysis but not by the first one
- These signals were dispatched among the 31 regional pharmacovigilance centres and the pharmacovigilance department at Afssaps (about 40 signals per centre)
Feasibility Study

Methodology

The centres were given one month to

- categorize the signals
  - known/nothing further
  - known/further follow up
  - unknown/further follow up
  - unknown/outlier
  - no time to assess

- Evaluate the time necessary for the analysis
Feasibility Study

Results

- 28 centres participated (sent back the questionnaire) → 1294 signals
- 1170 signals were categorised
  - 35.7% known/nothing further
  - 6.9% known/further follow up
  - 16.8% unknown/further follow up
  - 36.6% unknown/outlier
  - 4% no time to assess
- Large variability in the time necessary for analysis
  - between 2 and 26 hours per centre
  - median: 6 hours

Conclusion

- 277 signal worth following up
- At least 60% of relevant signals
- The lack of precision was often invoked for the outliers: too vague signals
  → need to work out the coding dictionaries
Description of the current methods

Extension to the multiple comparison setting

Simulation study

Application to the French pharmacovigilance data

Discussion
Discussion (1)

Extension of the existing methods to the multiple comparison framework

- No modification of the model
- New decision rules

The GPS model with $\Pr(H_{0_{ij}}^*)$ provides the best results

- in the simulation study
- in the sequential evaluation study
Discussion (2)

Limits of the GPS model

- Data represented as a contingency table
  - Common to all the methods currently used
  - co-prescription

Use of penalized regressions

- Logistic regression in which
  - One studies one adverse event at a time
  - Several hundred of predictors: the drugs (coded in 1/0)
  - Several hundred thousand (or millions) of individuals

- Lasso type algorithm / Stability Selection (Meinshausen & Bühlmann JRSS B 2010)

- Much more intensive

- Should make it possible to better account for the correlation structure between the drugs
  - co-prescription

- Direct extension to the study of the interaction drug-drug
References

False discovery rate estimation for frequentist pharmacovigilance signal detection methods.
PMID: 19432790.

Pharmacovigilance data mining with methods based on false discovery rates: a comparative simulation study.
PMID: 20811349.

Bayesian pharmacovigilance signal detection methods revisited in a multiple comparison setting.
PMID: 19360795.