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

Introduction (1)

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

Introduction (3)

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

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) $\times$ 820 classes of adverse events (AE)
  - $551\,040$ cells of which 80% are empty

For a particular couple (AE_i, D_j)

	Drug j	Other Drugs	
Adverse event i	n_{ij}	$n_{i\bar{j}}$	$n_{i.}$
Other adverse events	$n_{\bar{i}j}$	$n_{\bar{i}\bar{j}}$	$n_{\bar{i}.}$
	$n_{.j}$	$n_{.\bar{j}}$	n

- 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)

$$\hat{\psi}_{ij} = \frac{n_{ij}\, n_{\bar{i}\bar{j}}}{n_{\bar{i}j}\, n_{i\bar{j}}}$$

Signal generation

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

$$\widehat{\text{var}}\{\ln(\hat{\psi}_{ij})\} = \frac{1}{n_{ij}} + \frac{1}{n_{\bar{i}\bar{j}}} + \frac{1}{n_{\bar{i}j}} + \frac{1}{n_{i\bar{j}}}$$

- A signal is generated if

$$\ln(\hat{\psi}_{ij}) - 1.96\ \text{var}\{\ln(\hat{\psi}_{ij})\}^{1/2} > 0$$

Proportional Reporting Ratio (PRR) Evans *et al.* (2001)

Association measure

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

$$\hat{\varphi}_{ij} = \frac{n_{ij} / n_{i.}}{n_{\bar{i}j} / n_{\bar{i}.}}$$

- very close to $\hat{\psi}_{ij}$

Signal generation

- Evans *et al.* (2001)
  - $\hat{\varphi}_{ij} > 2$
  - $n_{ij} \geq 3$
  - A χ^2 statistic with 1 df ≥ 4
- van Puijenbroek *et al.* (2002) : same as for the ROR method

Gamma Poisson Shrinkage (GPS) DuMouchel (1999)

Poisson - 2 gamma mixture model

- $n_{ij} | \, e_{ij}, \lambda_{ij} \sim \mathbf{Pn}(\lambda_{ij} \, e_{ij})$ avec $e_{ij} = \dfrac{n_{i.} \, n_{.j}}{n}$
- $\lambda_{ij} \sim \hat{w} \, \mathbf{Ga}(\hat{\alpha}_1, \hat{\beta}_1) + (1 - \hat{w}) \, \mathbf{Ga}(\hat{\alpha}_2, \hat{\beta}_2)$

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

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

Association measure

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

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

Gamma Poisson Shrinkage (GPS) DuMouchel (1999)

Poisson - 2 gamma mixture model

- $n_{ij} | \, e_{ij}, \lambda_{ij} \sim \mathbf{Pn}(\lambda_{ij} \, e_{ij})$ avec $e_{ij} = \dfrac{n_{i.} \, n_{.j}}{n}$
- $\lambda_{ij} \sim \hat{w} \, \mathbf{Ga}(\hat{\alpha}_1, \hat{\beta}_1) + (1 - \hat{w}) \, \mathbf{Ga}(\hat{\alpha}_2, \hat{\beta}_2)$

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

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

Association measure

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

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

Signal generation

- $\mathrm{E}\{\log_2(\lambda^*_{ij})\}$ DuMouchel (1999)

Gamma Poisson Shrinkage (GPS) DuMouchel (1999)

Poisson - 2 gamma mixture model

- $n_{ij} | \, e_{ij}, \lambda_{ij} \sim \mathbf{Pn}(\lambda_{ij} \, e_{ij})$ avec $e_{ij} = \dfrac{n_{i.} \, n_{.j}}{n}$
- $\lambda_{ij} \sim \hat{w} \, \mathbf{Ga}(\hat{\alpha}_1, \hat{\beta}_1) + (1 - \hat{w}) \, \mathbf{Ga}(\hat{\alpha}_2, \hat{\beta}_2)$

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

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

Association measure

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

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

Signal generation

- $\mathrm{E}\{\log_2(\lambda^*_{ij})\}$ DuMouchel (1999)
- $\mathrm{Q}_{0.05}(\lambda^*_{ij})$ DuMouchel *et al.* (2001)

Gamma Poisson Shrinkage (GPS) DuMouchel (1999)

Poisson - 2 gamma mixture model

- $n_{ij} | \, e_{ij}, \lambda_{ij} \sim \mathbf{Pn}(\lambda_{ij} \, e_{ij})$ avec $e_{ij} = \dfrac{n_{i.} \, n_{.j}}{n}$
- $\lambda_{ij} \sim \hat{w} \, \mathbf{Ga}(\hat{\alpha}_1, \hat{\beta}_1) + (1 - \hat{w}) \, \mathbf{Ga}(\hat{\alpha}_2, \hat{\beta}_2)$

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

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

Association measure

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

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

Signal generation

- $\mathrm{E}\{\log_2(\lambda^*_{ij})\}$ DuMouchel (1999)
- $\mathrm{Q}_{0.05}(\lambda^*_{ij})$ DuMouchel *et al.* (2001)
- $\mathrm{Q}_{0.05}(\lambda^*_{ij}) > 2$ Szarfman *et al.* (2002)

Bayesian Confidence Propagation Neural Network (BCPNN) (1)

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

Multinomial-Dirichlet model

$$\begin{aligned} (n_{ij}, n_{i\bar{j}}, n_{\bar{i}j}, n_{\bar{i}\bar{j}}) &\sim \mathbf{Mu}(n, p_{ij}, p_{i\bar{j}}, p_{\bar{i}j}, p_{\bar{i}\bar{j}}) \\ \text{with } (p_{ij}, p_{i\bar{j}}, p_{\bar{i}j}, p_{\bar{i}\bar{j}}) &\sim \mathbf{Di}(\alpha_{ij}, \alpha_{i\bar{j}}, \alpha_{\bar{i}j}, \alpha_{\bar{i}\bar{j}}) \end{aligned}$$

- 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 :

$$(p_{ij}, p_{i\bar{j}}, p_{\bar{i}j}, p_{\bar{i}\bar{j}})^* \sim \mathbf{Di}(\gamma_{ij}, \gamma_{i\bar{j}}, \gamma_{\bar{i}j}, \gamma_{\bar{i}\bar{j}})$$

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

$$\begin{aligned}(n_{ij}, n_{i\bar{j}}, n_{\bar{i}j}, n_{\bar{i}\bar{j}}) &\sim \mathbf{Mu}(n, p_{ij}, p_{i\bar{j}}, p_{\bar{i}j}, p_{\bar{i}\bar{j}}) \\ \text{with } (p_{ij}, p_{i\bar{j}}, p_{\bar{i}j}, p_{\bar{i}\bar{j}}) &\sim \mathbf{Di}(\alpha_{ij}, \alpha_{i\bar{j}}, \alpha_{\bar{i}j}, \alpha_{\bar{i}\bar{j}})\end{aligned}$$

- 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 :

$$(p_{ij}, p_{i\bar{j}}, p_{\bar{i}j}, p_{\bar{i}\bar{j}})^* \sim \mathbf{Di}(\gamma_{ij}, \gamma_{i\bar{j}}, \gamma_{\bar{i}j}, \gamma_{\bar{i}\bar{j}})$$

  with $\gamma_{kl} = \alpha_{kl} + n_{kl}$

- In particular

$$\begin{aligned}p^*_{ij} &\sim \mathbf{Be}(\gamma_{ij}, \gamma_{i\bar{j}} + \gamma_{\bar{i}j} + \gamma_{\bar{i}\bar{j}}) \\ p^*_{i.} = p^*_{ij} + p^*_{i\bar{j}} &\sim \mathbf{Be}(\gamma_{ij} + \gamma_{i\bar{j}}, \gamma_{\bar{i}j} + \gamma_{\bar{i}\bar{j}}) \\ p^*_{.j} = p^*_{ij} + p^*_{\bar{i}j} &\sim \mathbf{Be}(\gamma_{ij} + \gamma_{\bar{i}j}, \gamma_{i\bar{j}} + \gamma_{\bar{i}\bar{j}})\end{aligned}$$

Bayesian Confidence Propagation Neural Network (BCPNN) (2)

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

Association measure

$$\mathrm{IC}_{ij}^{*} = \log_2 \left(\frac{p_{ij}^{*}}{p_{i.}^{*}\, p_{.j}^{*}} \right)$$

Ratio of beta distributions $\Rightarrow$ No analytic form

Signal generation

$$\mathrm{Q}_{0.025}(\mathrm{IC}_{ij}^{*}) > 0$$

- Normal approximation
  - delta method : Bate *et al.* (1998)
  - exact moments : Gould (2003)
- 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(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

- For each cell, we want to test $H_{0_{ij}} : \psi_{ij} \leq \psi_0$
- 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 Φ denotes the standard normal cdf

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

PRR

- $H_{0_{ij}} : \varphi_{ij} \leq \varphi_0$

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 : $\underbrace{\Pr(N_{ij} \geq n_{ij} | \, n_{i.}, n_{.j}, n; \psi_0)}_{\text{P-values}} - \frac{1}{2} \Pr(N_{ij} = n_{ij} | \, n_{i.}, n_{.j}, n; \psi_0)$

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 $\gamma \in]0, 1]$

$$\text{FDR}(\gamma) = \frac{\pi_0 F_0(\gamma)}{F(\gamma)}$$

Frequentist methods : Proposed approach (4)

FDR estimation : Single null hypothesis (not our case)

- The P-values are uniformly distributed under the null hypothesis

$$\begin{aligned} f(p) &= \pi_0\, f_0(p) &&+ (1-\pi_0)\, f_1(p) \\ &= \pi_0 &&+ (1-\pi_0)\, f_1(p) \end{aligned}$$

- 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 π_0

Frequentist methods : Proposed approach (5)

Location Based Estimator (LBE) Dalmasso *et al.* 2005

$$\mathsf{E}[\varphi(P)] = \pi_0 \mathsf{E}_0[\varphi(P)] + (1-\pi_0)\mathsf{E}_1[\varphi(P)]$$

$$\frac{\mathsf{E}[\varphi(P)]}{\mathsf{E}_0[\varphi(P)]} = \pi_0 + \underbrace{(1-\pi_0)\frac{\mathsf{E}_1[\varphi(P)]}{\mathsf{E}_0[\varphi(P)]}}_{\text{non negative term : Bias}}$$

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

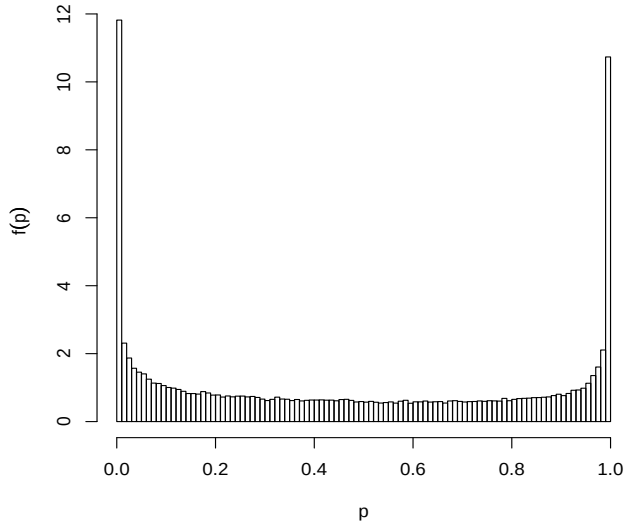
$$\hat{\pi}_0 = \frac{\hat{\mathsf{E}}[\varphi(P)]}{\mathsf{E}_0[\varphi(P)]} = \frac{1/m \sum_{ij} \varphi(p_{ij})}{\mathsf{E}_0[\varphi(P)]} = \frac{1/m \sum_{ij} \{-\ln(1-p_{ij})\}^a}{\Gamma(a+1)}$$

- 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



$$\begin{aligned} f(p) &= \pi_0 \quad f_0(p) \quad + (1-\pi_0) f_1(p) \\ &= \pi_0 \quad \{\pi_{0*} + (1-\pi_{0*}) f_{1*}(p)\} \quad + (1-\pi_0) f_1(p) \end{aligned}$$

- In practice, only small values of γ 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 $f_{P^*}(p^*) = \pi_0\pi_{0*} + (1 - \pi_0\pi_{0*}) f_{1P^*}(p^*)$

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

$$\text{FDP} = \frac{\sum_{ij}(1 - z_{ij})\, d_{ij}}{\sum_{ij} d_{ij}} \longrightarrow \text{FDR} = \text{E}[\text{FDP}]$$

Bayesian FDR estimation : FDR*

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

where $u_{ij} = \Pr(z_{ij} = 0|\,\text{data})$ 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 $\rightarrow H_{0_{ij}} : \lambda_{ij} \leq \text{RR}_0$
  - BCPNN $\rightarrow H_{0_{ij}} : \dfrac{p_{ij}}{p_{i.}p_{.j}} \leq \text{RR}_0$
- and thus to calculate u_{ij} i.e. the posterior probability of $H_{0_{ij}}$
  - GPS :

$$\Pr(\lambda_{ij}^* \leq \text{RR}_0)$$
$$= w_{ij}\, F_{\mathbf{Ga}}(\text{RR}_0; \hat{\alpha}_1 + n_{ij}, \hat{\beta}_1 + e_{ij}) + (1 - w_{ij})\, F_{\mathbf{Ga}}(\text{RR}_0; \hat{\alpha}_2 + n_{ij}, \hat{\beta}_2 + e_{ij})$$

  - BCPNN :

$$\Pr(\text{IC}_{ij}^* \leq \log_2(\text{RR}_0))$$

No analytic form $\rightarrow$ Monte Carlo simulations

Bayesian methods : Proposed approach (3)

Decision rule d_{ij}

- GPS

  DuMouchel (1999) : $d_{ij} = \mathbf{1}_{[\mathsf{E}\{\log_2(\lambda_{ij}^*)\} > \tau]}$

  DuMouchel *et al.* (2001) : $d_{ij} = \mathbf{1}_{[\mathsf{Q}_{0.05}(\lambda_{ij}^*) > \tau]}$

  Our suggestion : $d_{ij} = \mathbf{1}_{[u_{ij} \leq \alpha]}$

  Szarfman *et al.* (2002) : $\mathsf{RR}_0 = 2$ and $\alpha = 0.05$

- BCPNN

  Our suggestion : $d_{ij} = \mathbf{1}_{[u_{ij} \leq \alpha]}$

  Bate *et al.* (1998) : $\mathsf{RR}_0 = 1$ and $\alpha = 0.025$

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 : $\mathrm{E}\{\log_2(\lambda_{ij}^*)\}$, $\mathrm{Q}_{0.05}(\lambda_{ij}^*)$, $\Pr(\mathrm{H}_{0_{ij}}^*)$
  - BCPNN : $\Pr(\mathrm{H}_{0_{ij}}^*)$
- 3 different tested hypotheses based on $\{\psi_0, \mathrm{RR}_0\} = 1, 2$ and 5

Simulation study

Data generation

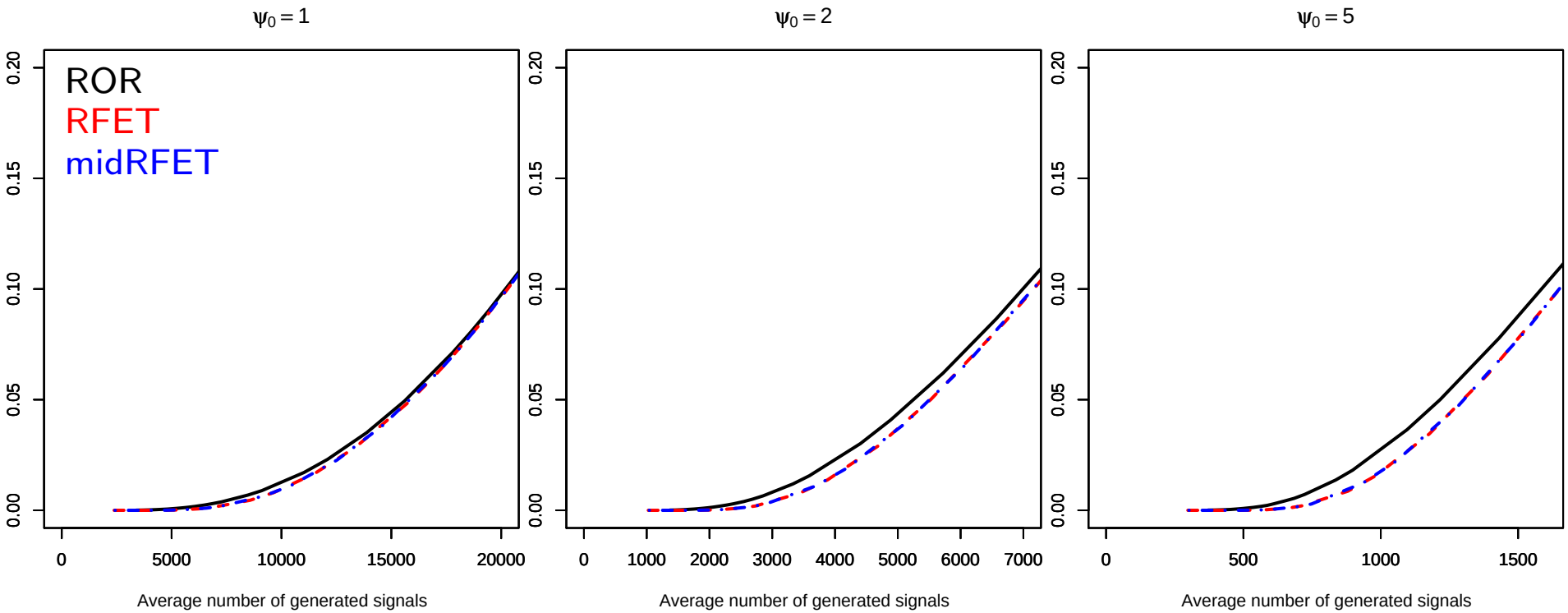
- Model : $\mathbf{n_{ij}} \sim \mathbf{Mu}(n, \mathbf{p_{ij}})$ from the French database
- $\mathbf{p_{ij}}$
  - ${\mathbf{p_{i.}}}^w \sim \mathbf{Di}(\mathbf{n_{i.}})$
  - ${\mathbf{p_{.j}}}^w \sim \mathbf{Di}(\mathbf{n_{.j}}) \quad \Longrightarrow \quad p_{ij} = \frac{r_{ij}^w \; p_{i.}^w \; p_{.j}^w}{\sum_{ij} r_{ij}^w \; p_{i.}^w \; p_{.j}^w}$
  - $\log(r_{ij}^w) \sim \mathbf{Lo}(0, 0.5)$
- From $\mathbf{p_{ij}}$
  - n_{ij}s and the true marginal probabilities : $p_{i.} = \sum_j p_{ij} \quad p_{.j} = \sum_i p_{ij}$
  - real status of the cells according to
    - ψ_{ij}, and ψ_0 for the frequentist methods
    - $\text{RR}_{ij} = \frac{p_{ij}}{p_{i.}\, p_{.j}}$ and RR_0 for the bayesian methods

Simulation plan

- 500 simulated datasets
- « True »FDR estimated by the average of the FDPs over the 500 datasets
- $n_{ij} \geq 3$

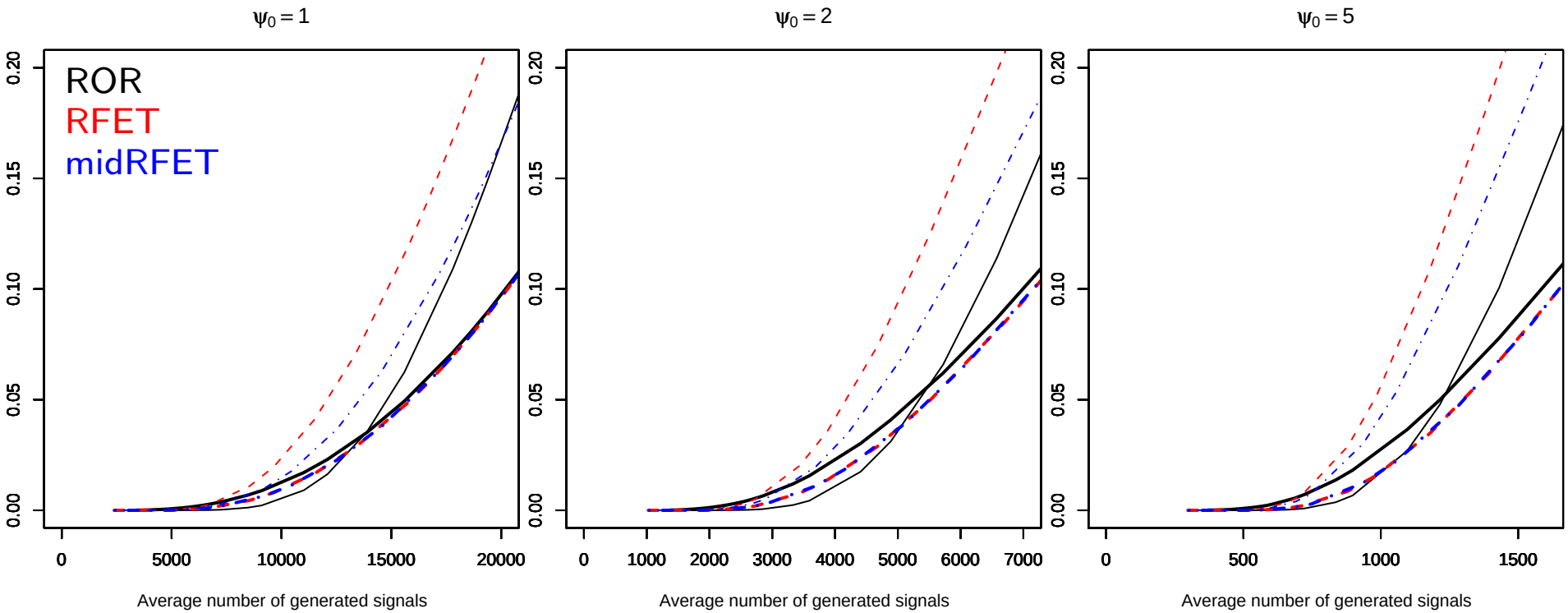
Simulation study : Comparison of the frequentist methods

FDR



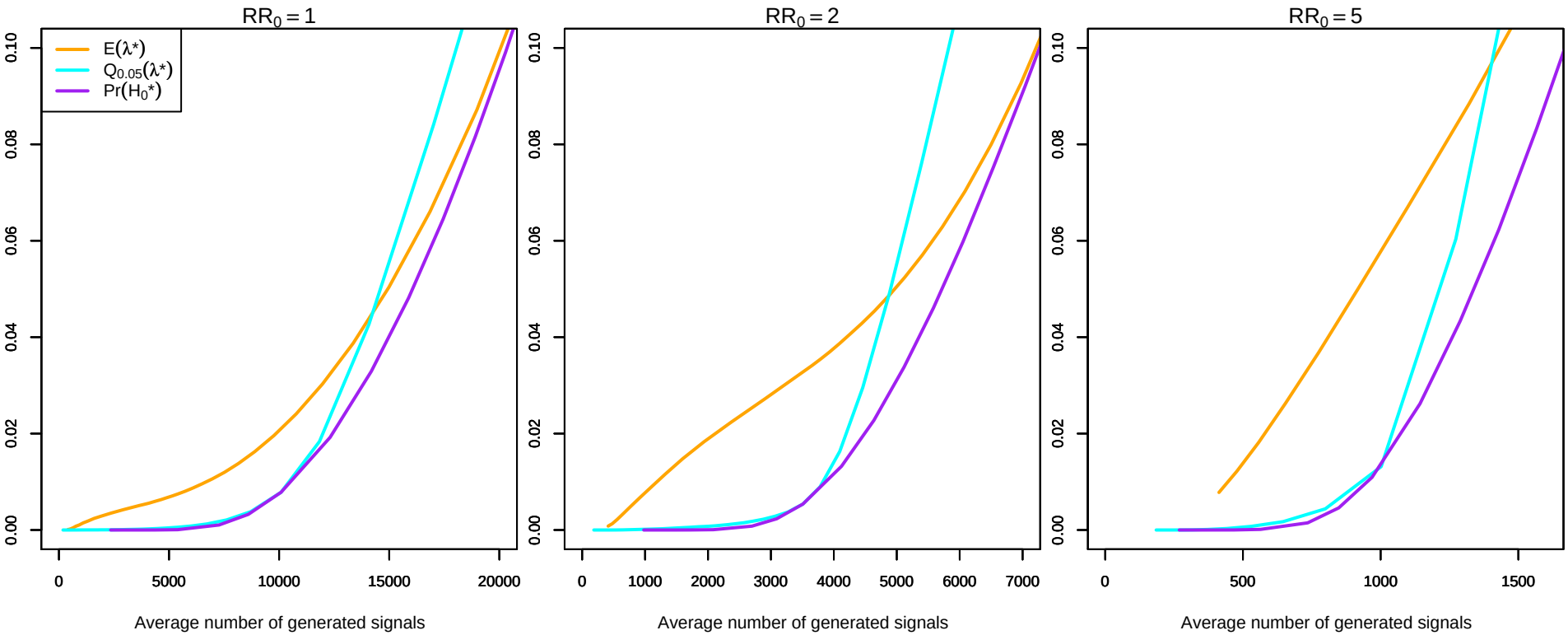
Simulation study : Comparison of the frequentist methods

FDR and estimation



Simulation study : Comparison of the Bayesian methods (1)

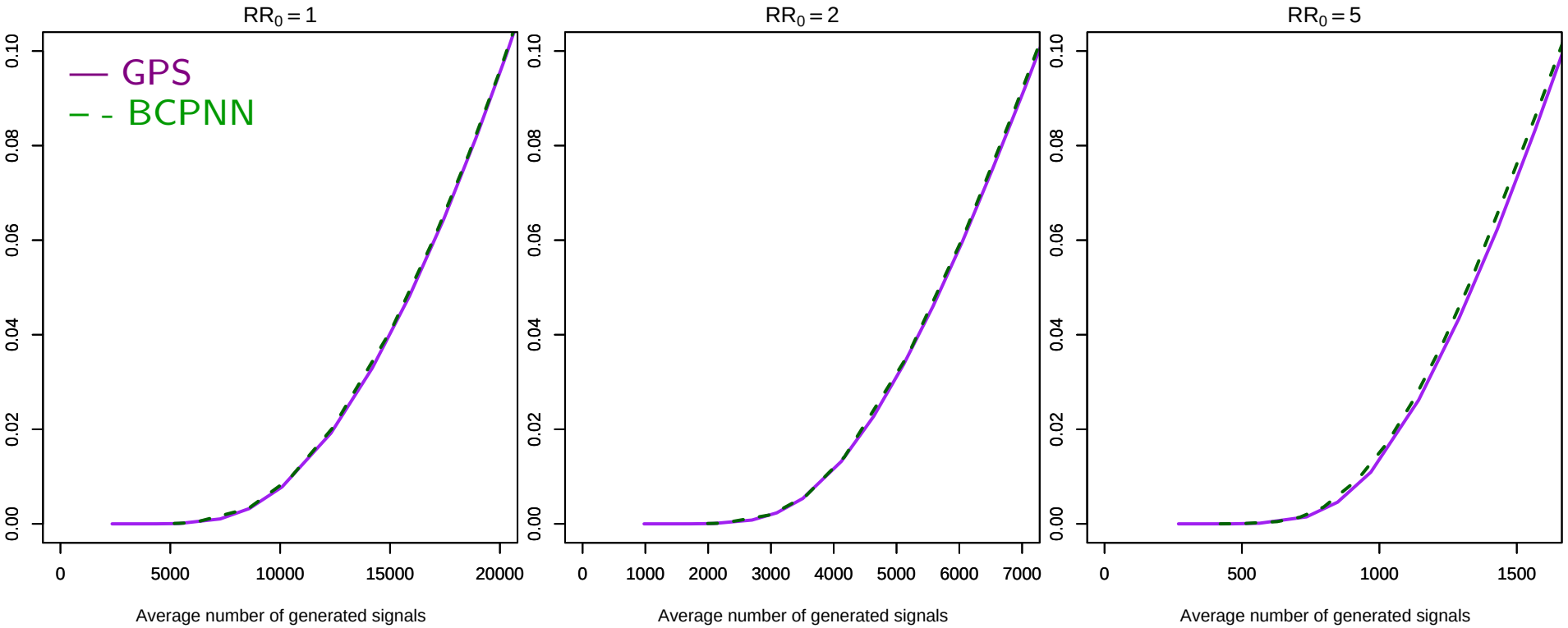
GPS and decision rules



- The proposed decision rule gives better results according to the FDR
- Close performances between $\mathrm{Q}_{0.05}(\lambda^*_{ij})$ and $\mathrm{Pr}(\mathrm{H}^*_{0_{ij}})$ for small values of FDR

GPS vs BCPNN with $\Pr(H^*_{0_{ij}})$

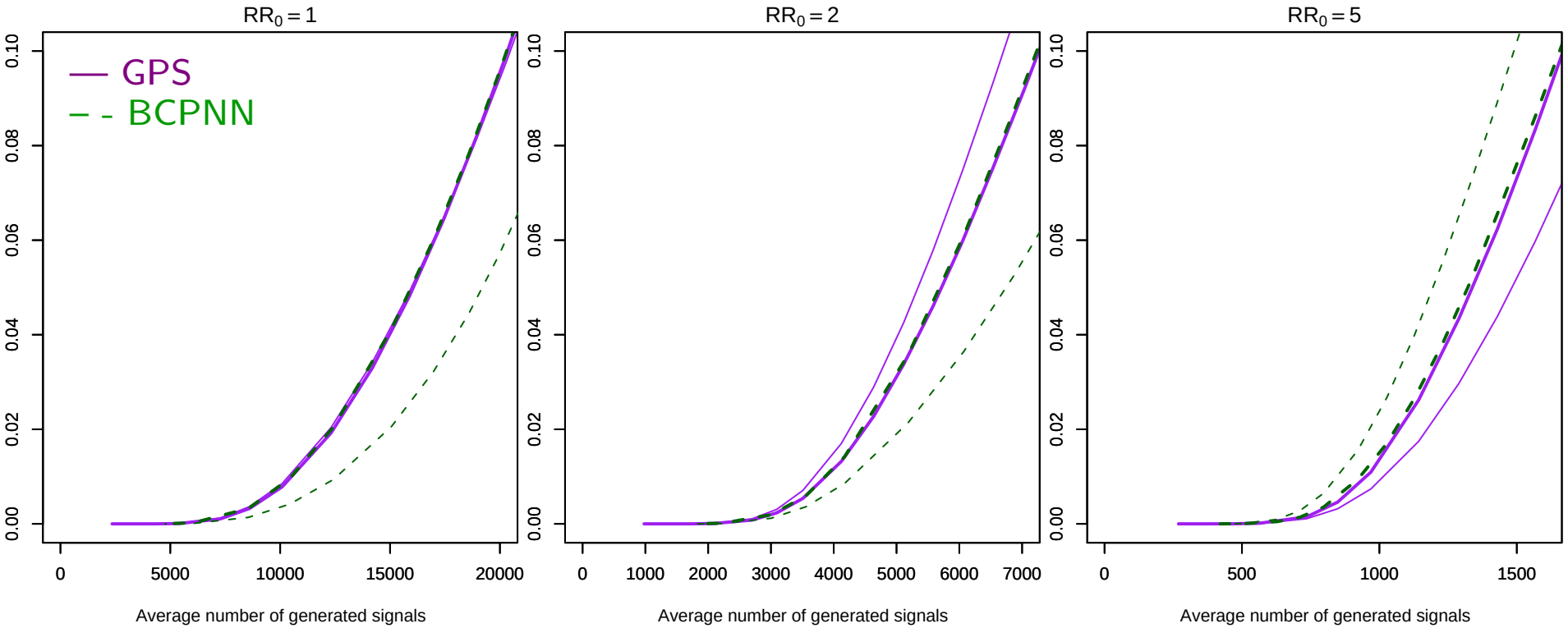
FDR and estimation



- Very close performances according to the FDR

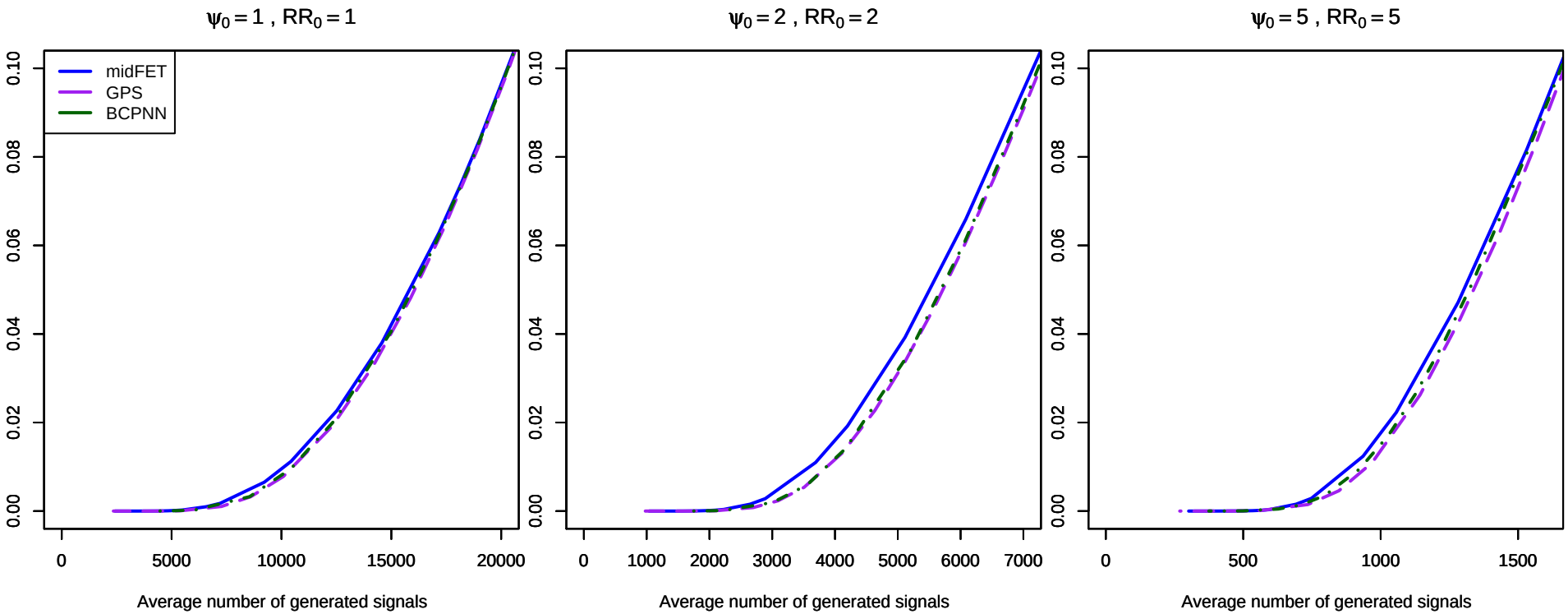
GPS vs BCPNN with $\Pr(H^*_{0_{ij}})$

FDR and estimation



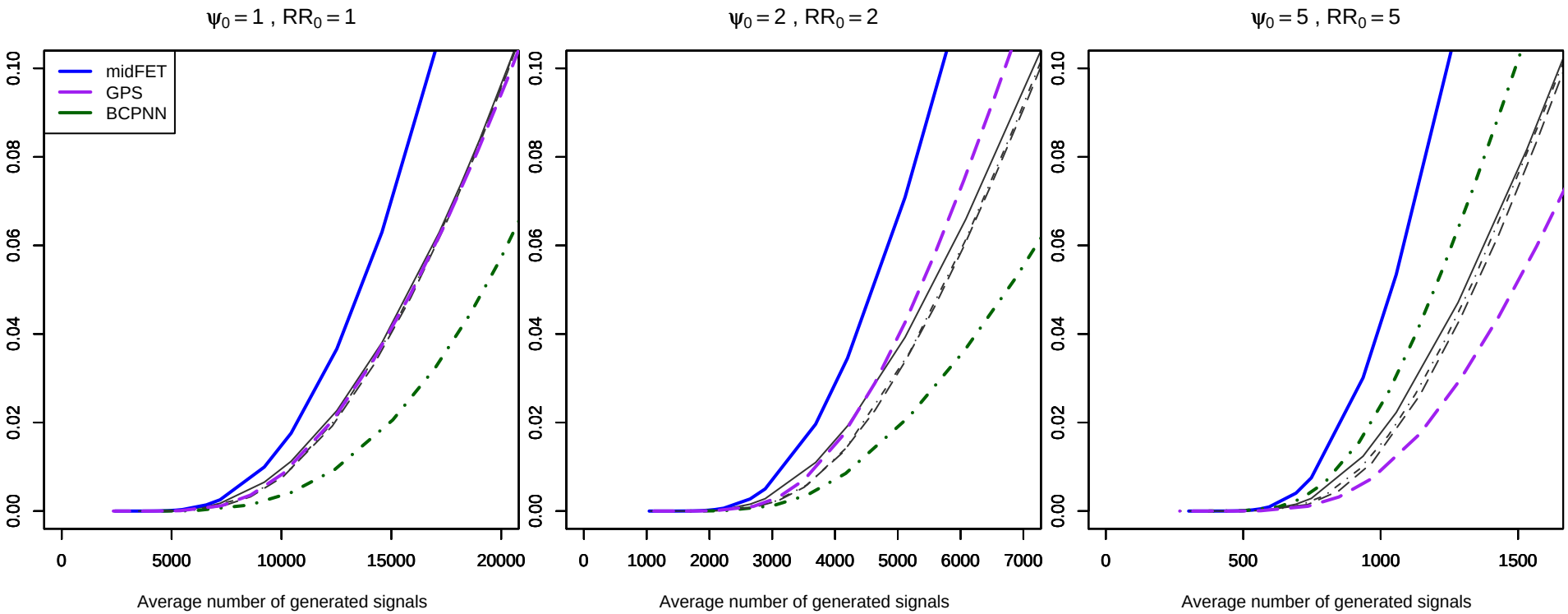
- Very close performances according to the FDR
- Large differences in the FDR estimation

Global Comparison



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

Global Comparison



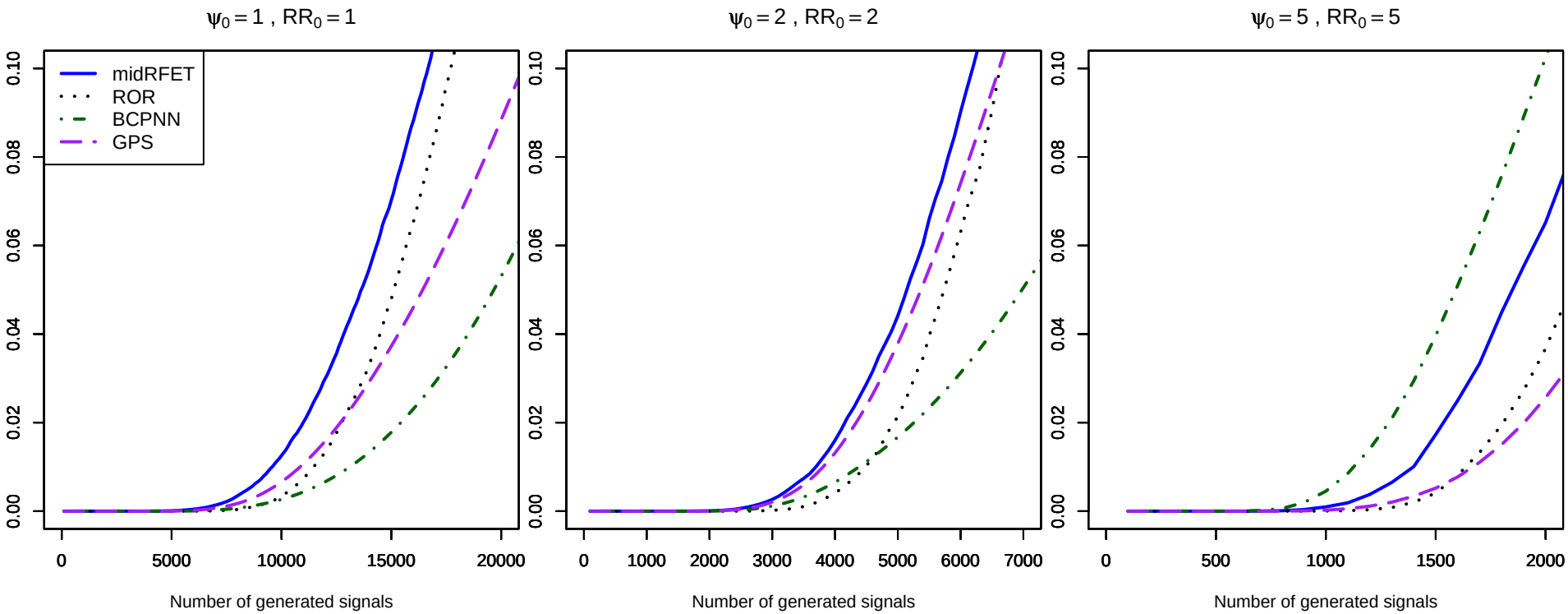
- Very close performances according to the FDR
- Large differences in FDR estimation

Results in favor of the use of the GPS model with $\Pr(\mathrm{H}^*_{0_{ij}})$

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 GPS_{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) $\Longrightarrow$ 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
  $\Longrightarrow$ 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 $\Pr(\mathrm{H}^*_{0_{ij}})$ 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

I Ahmed, C Dalmasso, F Haramburu, F Thiessard, P Broët, and P Tubert-Bitter.

False discovery rate estimation for frequentist pharmacovigilance signal detection methods.
Biometrics, 66(1) :301–309, March 2010.
PMID : 19432790.

I Ahmed, F Thiessard, G Miremont-Salamé, B Bégaud, and P Tubert-Bitter.

Pharmacovigilance data mining with methods based on false discovery rates : a comparative simulation study.
Clinical Pharmacology and Therapeutics, 88(4) :492–498, October 2010.
PMID : 20811349.

Ismaïl Ahmed, Françoise Haramburu, Annie Fourrier-Réglat, Frantz Thiessard, Carmen Kreft-Jais, Ghada Miremont-Salamé, Bernard Bégaud, and Pascale Tubert-Bitter.

Bayesian pharmacovigilance signal detection methods revisited in a multiple comparison setting.
Statistics in Medicine, 28(13) :1774–1792, June 2009.
PMID : 19360795.