Post hoc inference for multiple testing

Pierre Neuvial

Laboratoire de Mathématiques et modélisation d'Evry

Joint (ongoing) work with Gilles Blanchard and Etienne Roquain

MAS 2016, Grenoble

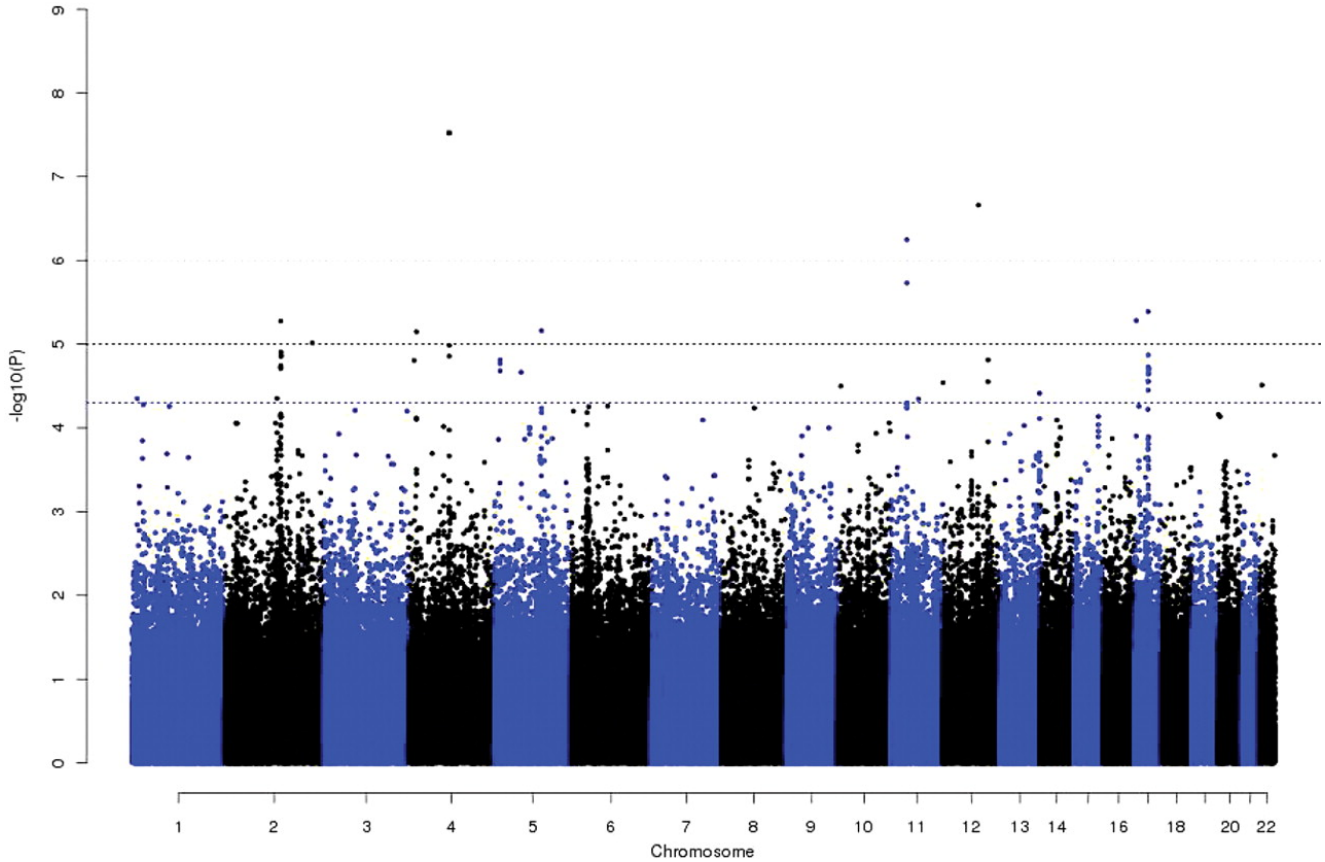
Outline

Outline

1. Post hoc inference
   - Motivation
   - State of the art: Goeman and Solari (2011)
2. Joint Family-Wise Error Rate control for post hoc inference
   - A novel risk measure: JFWER
   - Connection to GS2011
3. Obtaining Joint Family-Wise Error Rate control
   - Setting: known dependency
   - Adjustment of a reference threshold family

Outline

Motivating example: Genome-Wide Association Studies[1]



- $m \sim 10^6$ tests (genomic markers)
- $n \sim 10^3 - 10^4$ observations (individuals)

[1] Saad M, *et al*, *Human molecular genetics* 20.3 (2011), pp. 615–627

Multiple testing

- $\mathcal{H} = \{1, \dots m\}$ m null hypotheses to be tested
- $\mathcal{H}_0 \subset \mathcal{H}$: true null hypotheses, $\mathcal{H}_1 = \mathcal{H} \setminus \mathcal{H}_0$
- $(p_i)_{1 \leq i \leq m}$: p-values

Multiple testing procedures

Aim at building from the data a set R of rejected hypotheses satisfying a statistical guarantee, e.g. controlling:

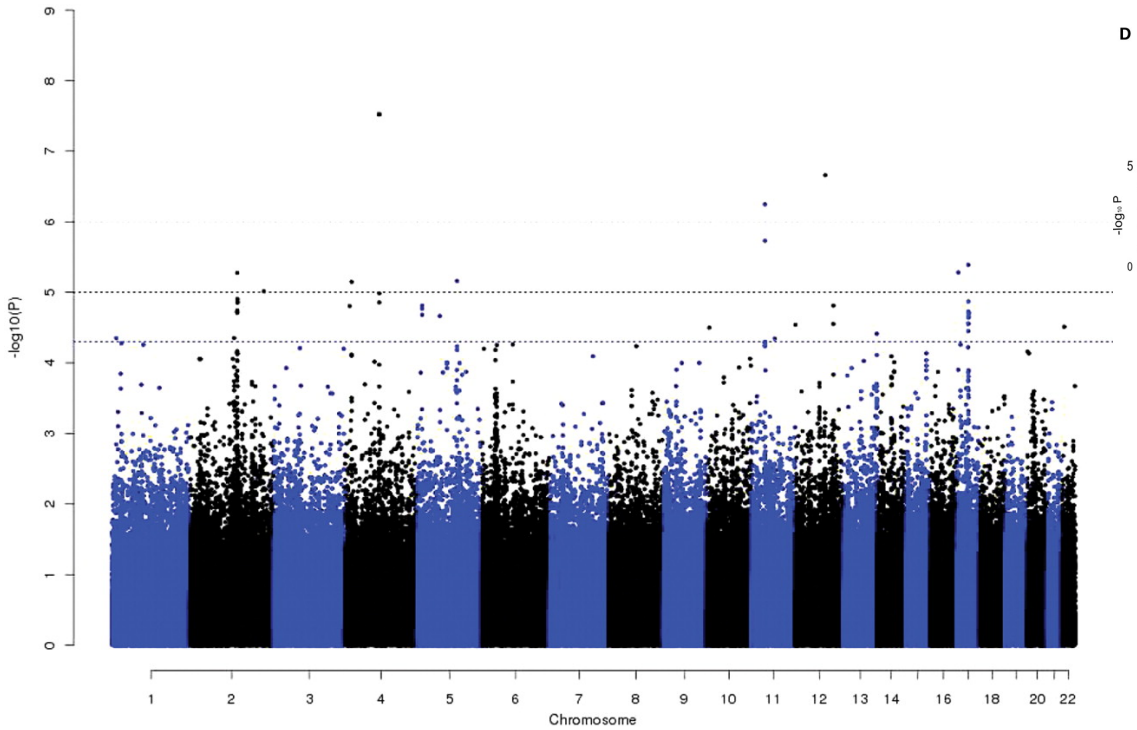
- (k-)Family-Wise Error Rate: $k\text{-FWER} = \mathbb{P}(|\mathcal{R} \cap \mathcal{H}_0| > k - 1)$
- False Discovery Rate[a]: $\text{FDR} = \mathbb{E}\left(\frac{|\mathcal{R} \cap \mathcal{H}_0|}{|\mathcal{R}| \vee 1}\right)$

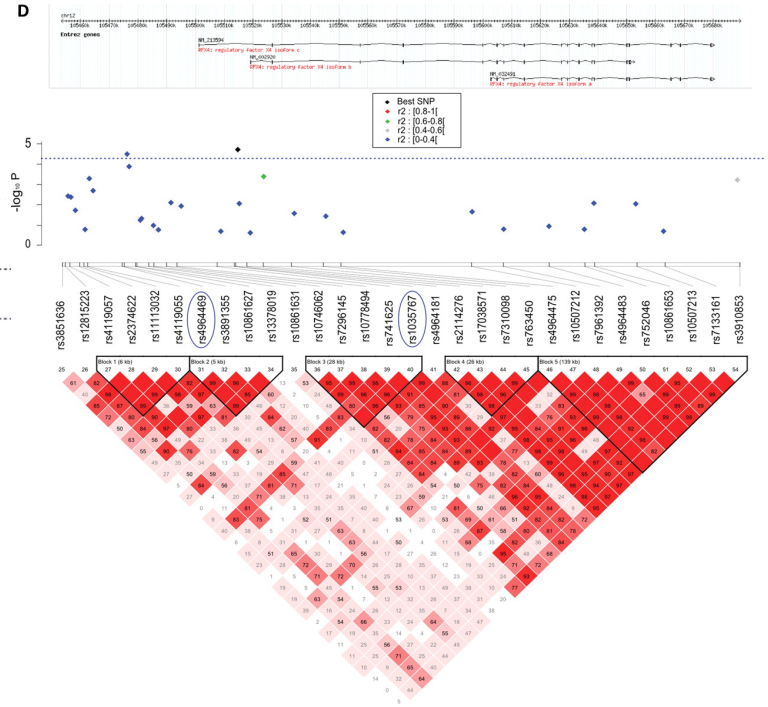
Most procedures used in applications are *thresholding procedures*:

$$\mathcal{R} = \{i \in \mathcal{H}, p_i \leq \hat{t}\}$$

[a] Benjamini and Hochberg, *JRSS B* (1995)

Motivating example: Genome-Wide Association Studies



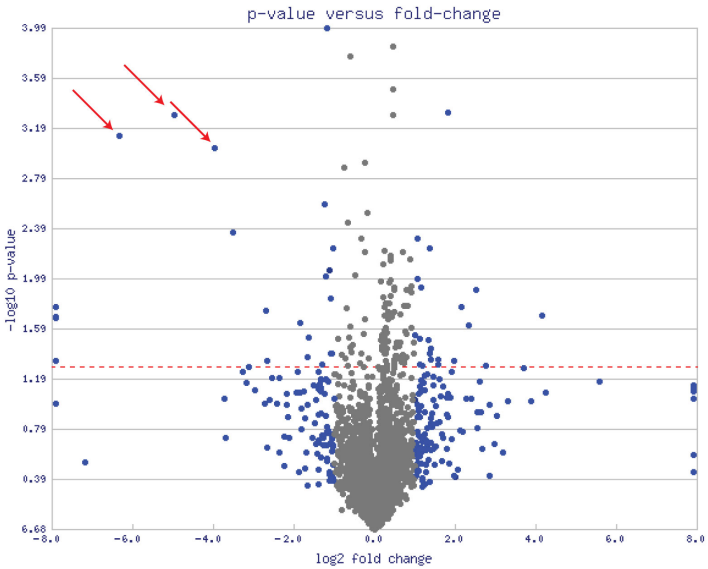


Typical analysis steps

1. define a list of candidates using a *multiple testing procedure*
2. refine this list based on *prior knowledge* (genome regions)

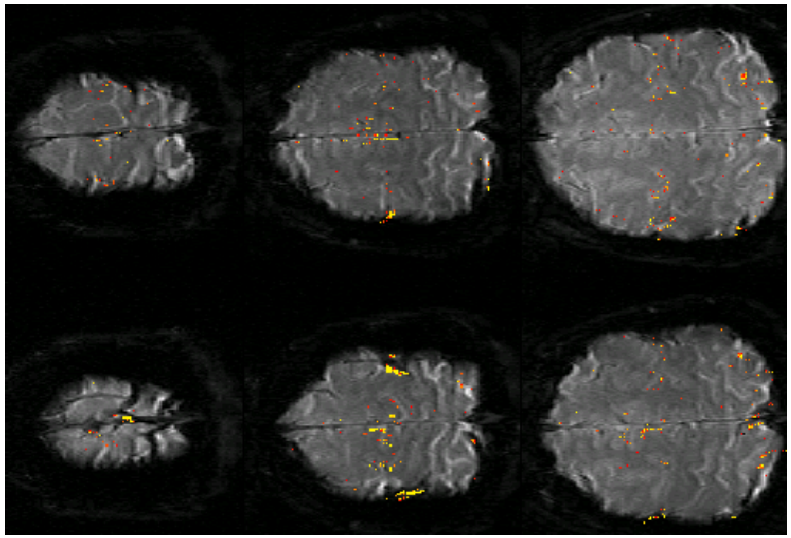
More motivating examples

Cancer studies



Differential gene expression analyses

Neuroimaging



Activation of brain regions

Typical analysis steps

1. define a list of candidates using a *multiple testing procedure*
2. refine this list based on *prior knowledge* (genome regions, gene pathways, brain regions)

Limitations of classical multiple testing procedures

Practical limitation

The initial selection does not take full advantage of available prior knowledge

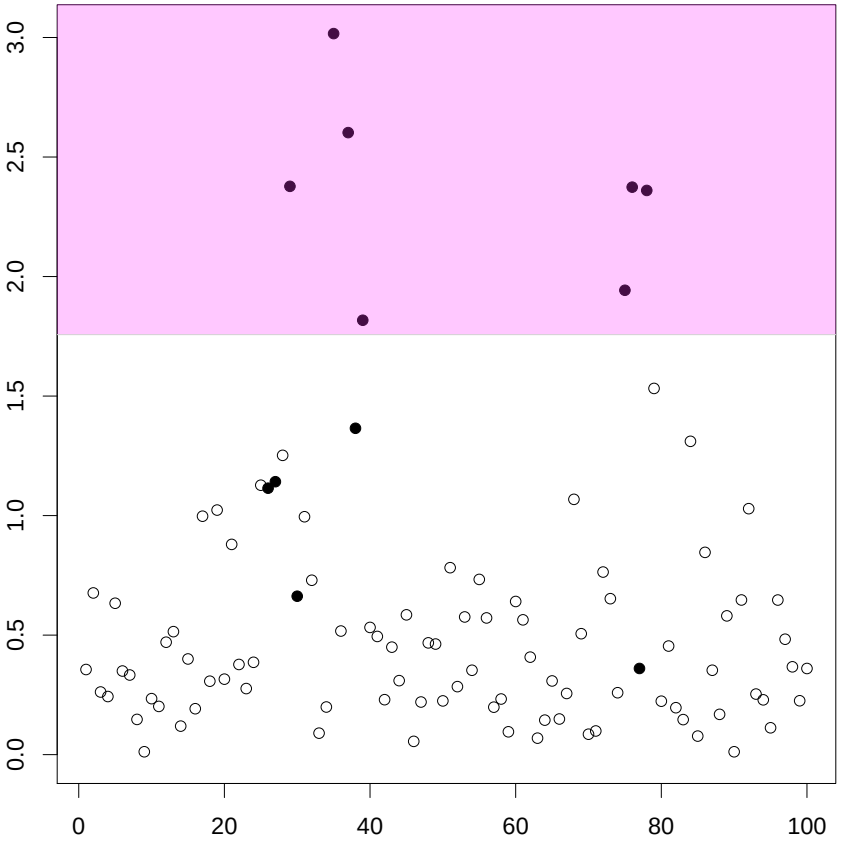
Theoretical limitation

No formal risk assessment can generally be made on the resulting sets of candidates

Can we obtain confidence statements on rejection sets selected after data analysis ?

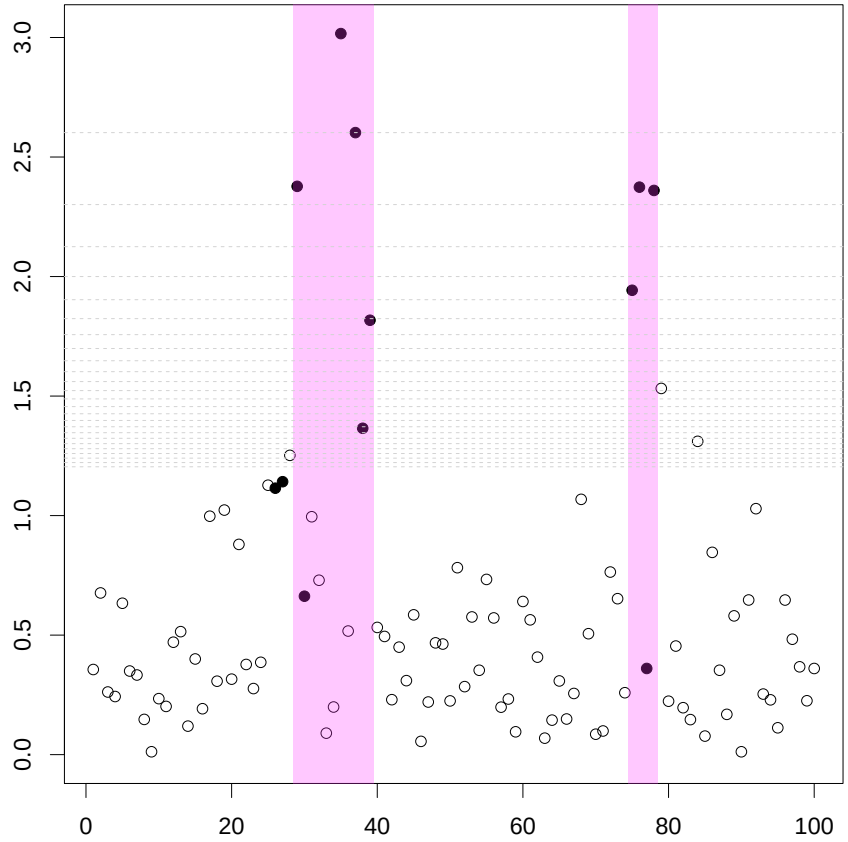
Post hoc inference in a nutshell

Classical multiple testing



FDR $\leq 25\%$

Post hoc inference



With probability $\geq 75\%$

$|\mathcal{H}_1 \cap \mathcal{R}_1| \geq 2$ and $|\mathcal{H}_1 \cap \mathcal{R}_2| \geq 1$

Goeman and Solari (2011)

Existing post hoc procedures[2] are based on *closed testing*[3]

- Require testing all $2^m - 1$ possible intersections between the m original hypotheses!
- Not feasible for $m \geq 20$ or 30.

In practice: "shortcuts"

- computationally efficient procedures (complexity $\sim m \log(m)$)
- increased conservativeness and/or narrower applicability:
- Simes' shortcut: valid under positive dependence between hypotheses (PRDS)

[2] Multiple testing for exploratory research. *Stat. Science* (2011)

[3] Marcus, Peritz and Gabriel, *Biometrika* (1976).

Outline

Joint Family-Wise Error Rate

Definition

A family of nested rejection sets $(R_k)_{k=1\dots m}$ is said to control JFWER at level $\alpha \in [0,1]$ if:

$$\mathbb{P}\big(\forall k \in \{1,\dots,m\}, |\mathcal{H}_0 \cap R_k| \leq k-1\big) \geq 1-\alpha.$$

Interpretation

Simultaneous k-FWER control for all k

Thresholding-based rejection sets

$$R_k = \{1 \leq i \leq m \ : \ p_i \leq t_k(\alpha)\}\,.$$

Post hoc inference through JFWER control

Upper bound on the number of false positives

Given a JFWER controlling family $(R_k)_{k=1\dots m}$, with probability larger than $1-\alpha$, for any rejection set $\mathcal{R}$,

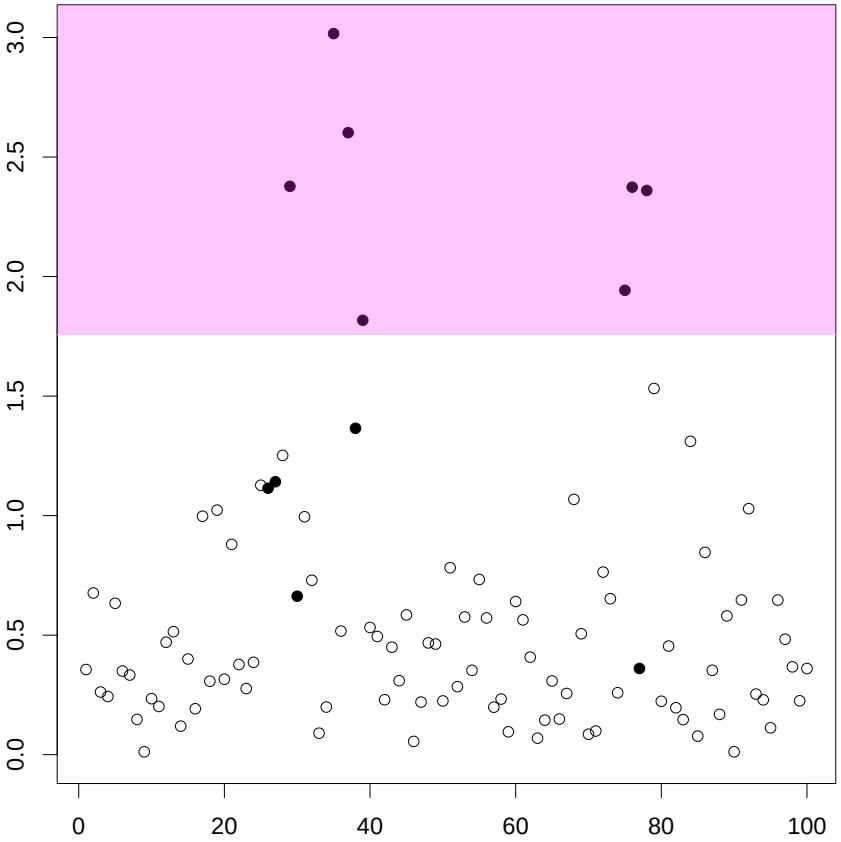
$$|\mathcal{R} \cap \mathcal{H}_0| \leq |\mathcal{R}| \wedge \min_{1 \leq k \leq |\mathcal{R}|} \{|\mathcal{R} \cap (R_k)^c| + k - 1\}$$

Properties

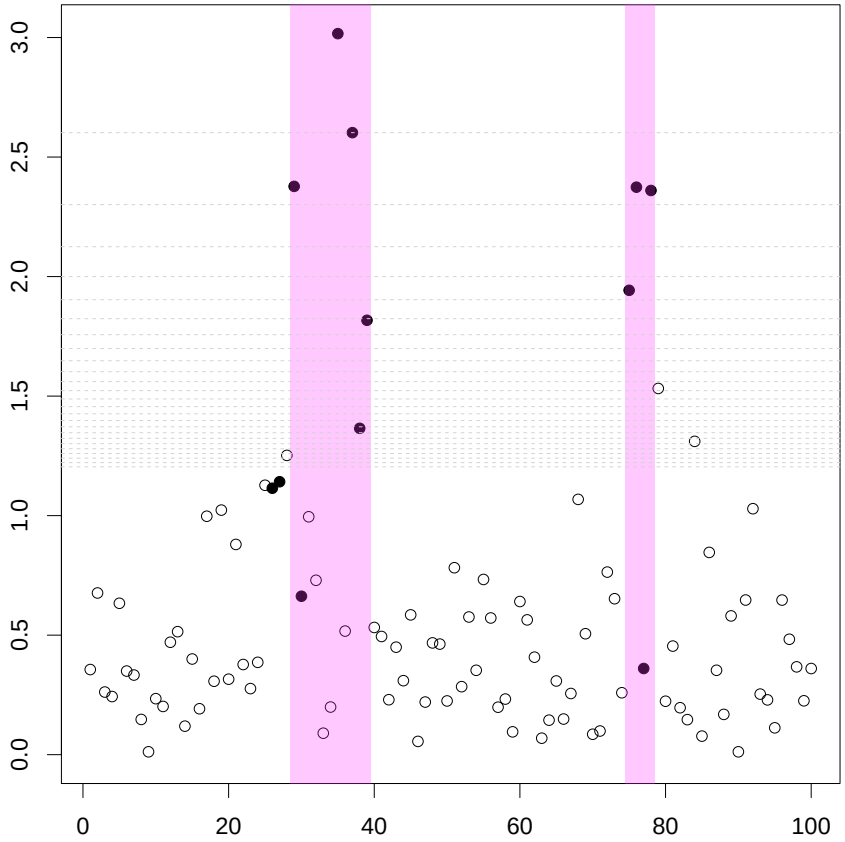
- data-driven rejection sets
- any number of rejection sets

Illustration

Classical multiple testing



Post hoc inference



- data-driven rejection sets
- any number of rejection sets

(How) can JFWER control be achieved?

Simes' inequality [4]

If the p-values (p_i), $1 \leq i \leq m$, are PRDS then

$$\mathbb{P}(\exists k \in \{1, \ldots, m_0\} \ : \ q_{(k)} \leq \alpha k / m_0) \leq \alpha,$$

where $q_{(1)} \leq \cdots \leq q_{(m_0)}$ denote the ordered p-values under H_0

[4] R. J. Simes. *Biometrika* 73.3 (1986), pp. 751–754.

Simes-based JFWER control

Corollary of Simes' inequality

Under PRDS, JFWER control at level α is achieved by the family

$$R_k = \{1 \leq i \leq m \ : \ p_i \leq \alpha k/m\}, 1 \leq k \leq m$$

Proposition (Post hoc bound for the Simes family)

Under PRDS, with probability larger than $1-\alpha$, for any $\mathcal{R}$,

$$|\mathcal{R} \cap \mathcal{H}_0| \leq |\mathcal{R}| \wedge \min_{1 \leq k \leq |\mathcal{R}|} \left\{ \sum_{i \in \mathcal{R}} \mathbf{1}\{p_i > \alpha k/m\} + k - 1 \right\}.$$

- We recover the bound obtained by GS2011
- Easier to interpret (no more closed testing or shortcuts)

Outline

Dependence-free JFWER control?

Under arbitrary dependence, with probability larger than $1-\alpha$, for any $\mathcal{R}$,

$$|\mathcal{R} \cap \mathcal{H}_0| \leq |\mathcal{R}| \wedge \min_{1 \leq k \leq |\mathcal{R}|} \left\{ \sum_{i \in \mathcal{R}} \mathbf{1}\left\{p_i > \alpha / K_m k/m\right\} + k - 1 \right\},$$

$K_m = \sum_{j=1}^{m} j^{-1} \sim \log(m)$: Hommel's correction factor for dependency[5]

Dependence-free adjustment is not a sensible objective

- implies adjusting to a worst case dependency
- very conservative (cf Benjamini-Yekutieli for FDR control)
- we need to be adaptive to dependency

Setting considered here: known dependency

Example: GWAS with pilot data

[5] G Hommel. "Tests of the overall hypothesis for arbitrary dependence structures". Biometrische Zeitschrift 25.5 (1983), pp. 423–430.

JFWER control with λ adjustment

Consider a *reference* family of rejections sets of the form

$$R_k = \{1 \leq i \leq m \ : \ p_i \leq t_k(\alpha)\}, 1 \leq k \leq m$$

Assumption: the joint null distribution of the test statistics is known or can be sampled from

Then it is possible to calibrate $\lambda > 0$ such that the rejection sets associated to $t_k(\lambda\alpha)$ yields tight JFWER control.

Examples of reference family

- Simes family: $t_k(\alpha) = \alpha k/m$ ($\lambda^{-1} = \sum_{k=1}^{m} k^{-1}$ works!)
- Balanced family: $t_k(\alpha)$ such that $\mathbb{P}\big(|\mathcal{H}_0 \cap R_k| \leq k-1\big) \geq 1-\alpha$

Recall: JFWER control under positive dependency

Simes' equality is sharp under independence, but conservative under positive dependence.

Conservativeness of JFWER control under PRDS

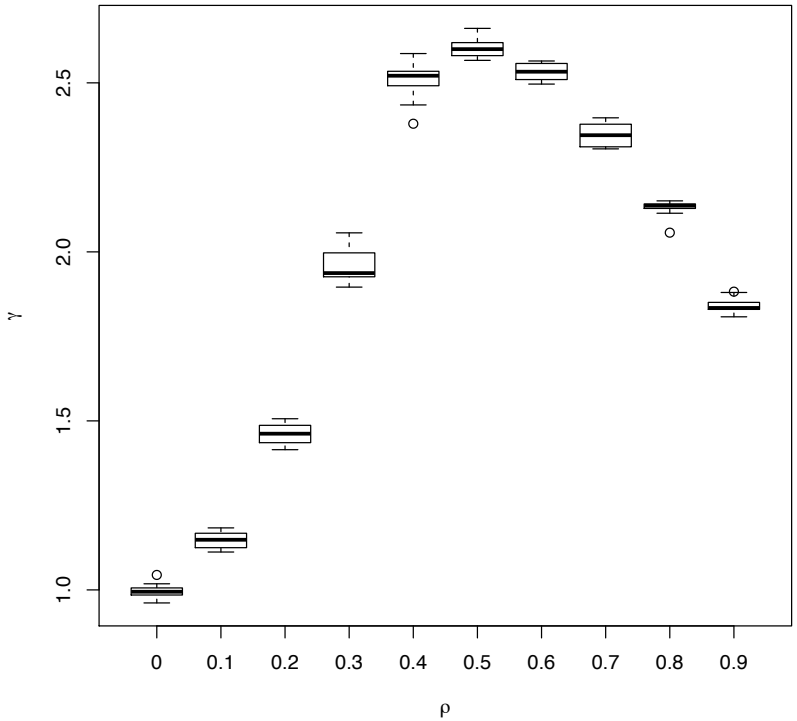
Toy example: Gaussian equi-correlation, white setting ($m_0 = m = 1,000$):
Test statistics $\sim \mathcal{N}(0, \Sigma)$ with $\Sigma_{ii} = 1$ and $\Sigma_{ij} = \rho$ for $i \neq j$.

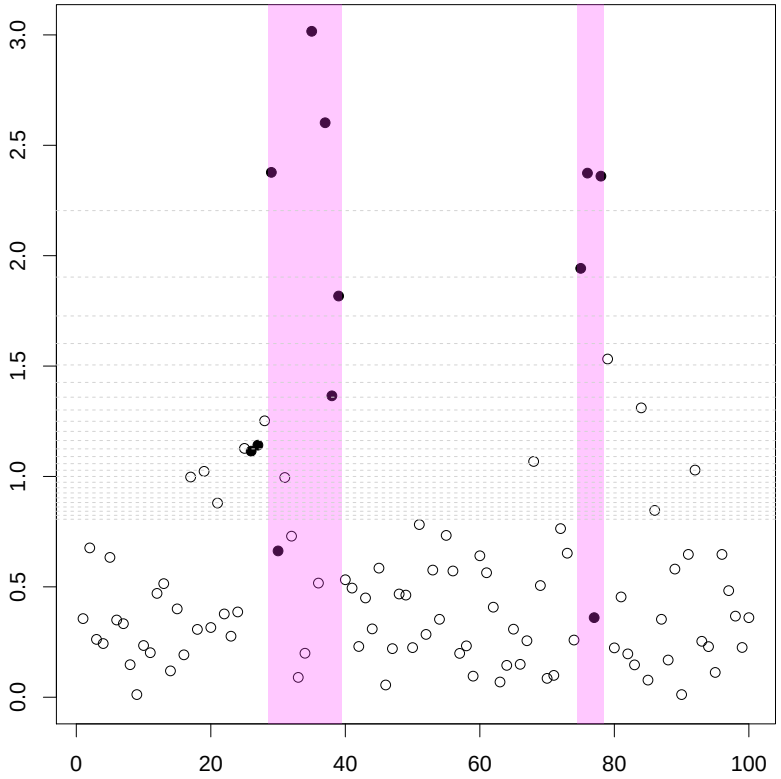
Equi-correlation level: ρ	0	0.1	0.2	0.4	0.8
Achieved JFWER $\times \alpha^{-1}$	0.99	0.85	0.72	0.42	0.39

Can we build a family achieving sharper JFWER control?

JFWER control with λ adjustment for Simes' family

Example under positive dependency (Gaussian equi-correlation)





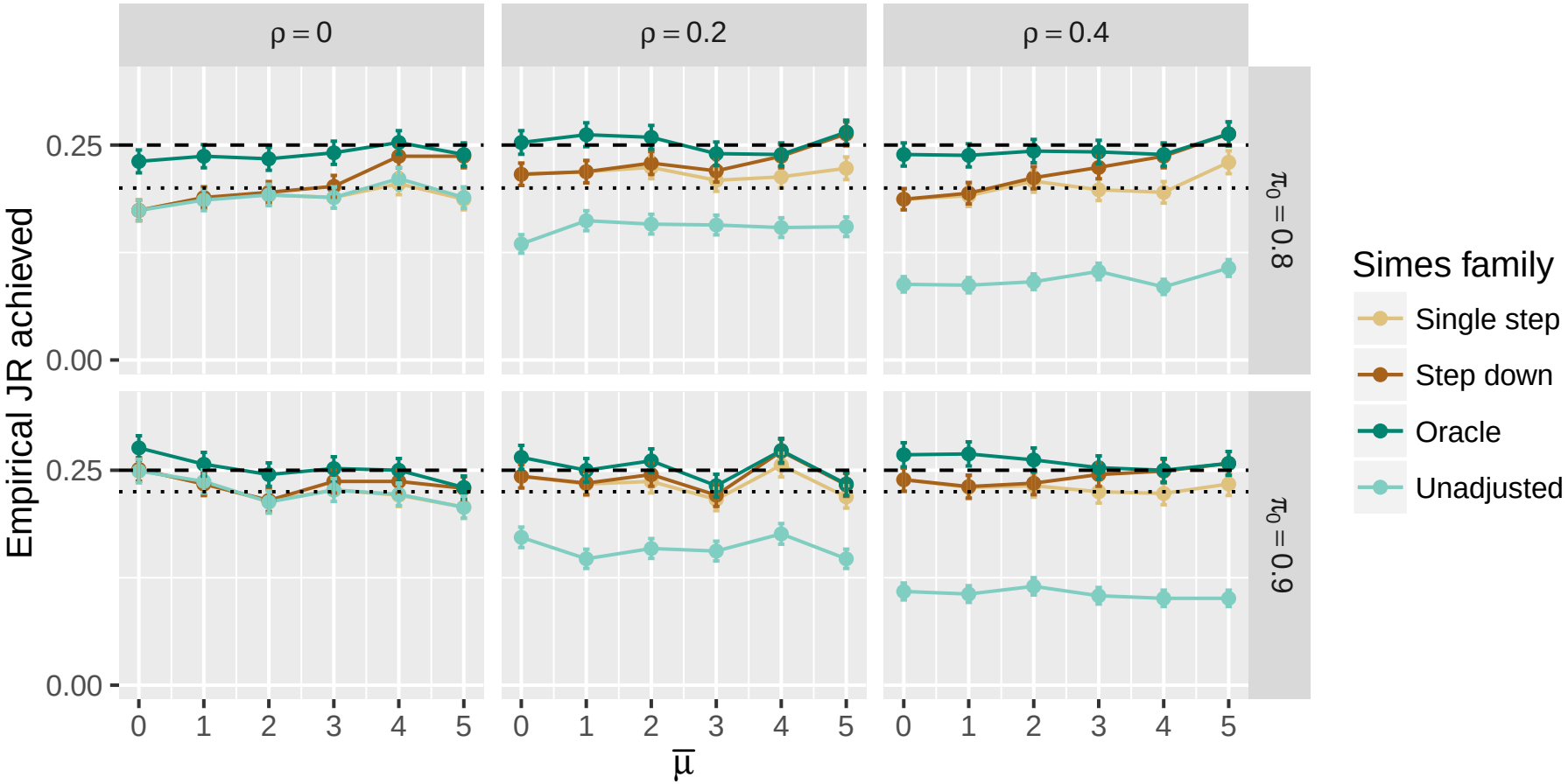
With probability $\geq 1 - \alpha = 75\%$:

$t_k(\alpha)$	Lower bound on $\lvert\mathcal{R} \cap \mathcal{H}_1\rvert$
$\alpha k/m$	$\lvert\mathcal{R}_1 \cap \mathcal{H}_1\rvert \geq 2$ and $\lvert\mathcal{R}_2 \cap \mathcal{H}_1\rvert \geq 1$
$\alpha\lambda k/m$	$\lvert\mathcal{R}_1 \cap \mathcal{H}_1\rvert \geq 3$ and $\lvert\mathcal{R}_2 \cap \mathcal{H}_1\rvert \geq 2$

JFWER control with λ adjustment for Simes' family

Numerical results under Gaussian equi-correlation:

- $X_i \sim \mathcal{N}(0,1)$ under H_0, and $X_i \sim \mathcal{N}(\bar{\mu},1)$ under H_1
- $\mathrm{cor}(X_i, X_j) = \rho$ for $i \neq j$



Conclusions

Summary

- JFWER: a new risk measure for multiple testing
- can be used to build post hoc inference procedures
- generalizes existing post hoc procedures

Results not discussed here

- Control of $\mathbb{P}(\forall k \in \{1, \ldots, \mathrm{Kmax}\}, |\mathcal{H}_0 \cap R_k| \leq k - 1)$
- Data-driven reference families for *balanced* JFWER control
- *Step-down* JFWER control in order to adapt to $|\mathcal{H}_0|$
- Power

Acknowledgements and future works

Thanks!

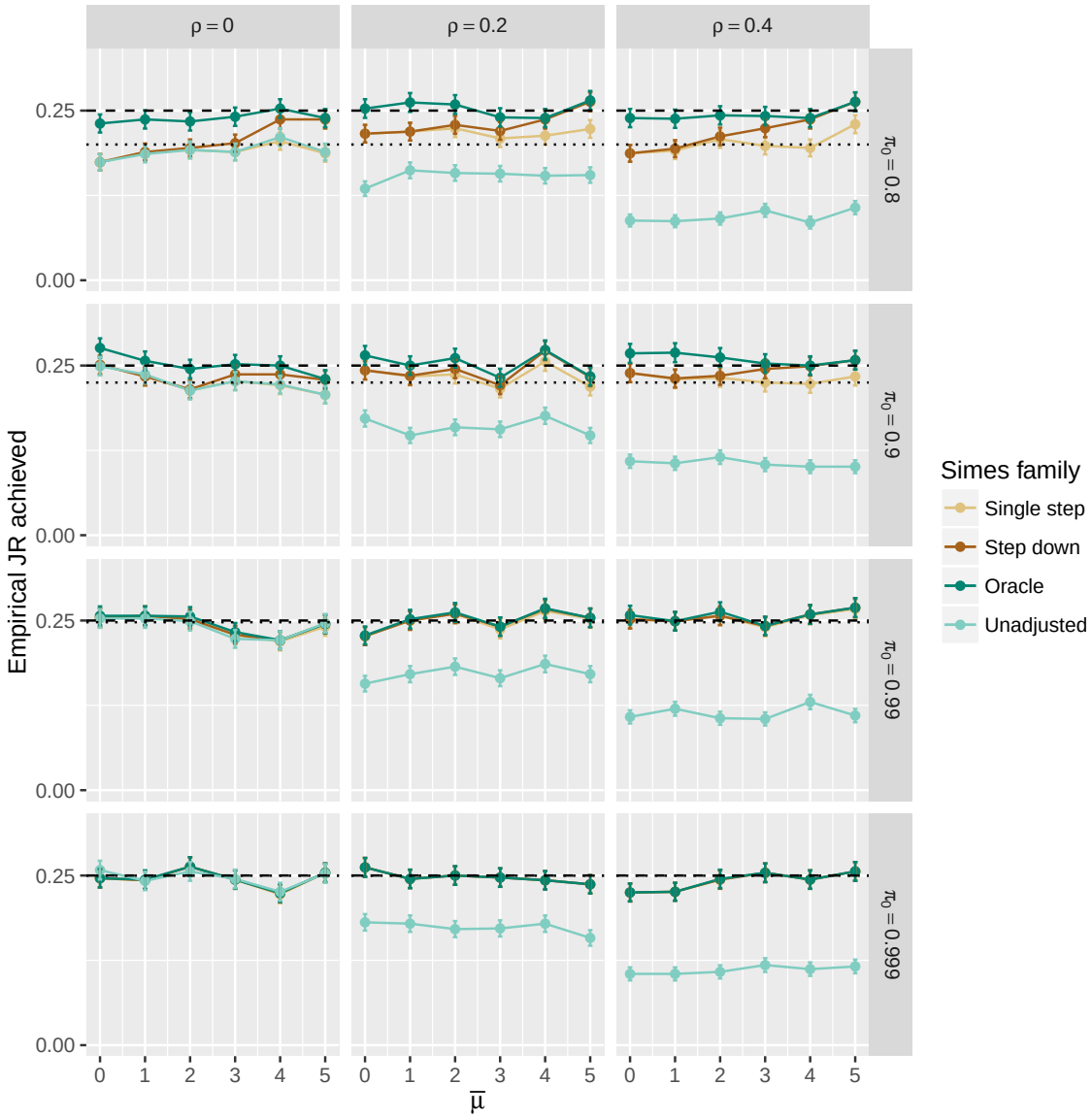
- Etienne Roquain, Gilles Blanchard
- Funding by CNRS: PEPS FaSciDo (Fondements et Applications de la Science des Données)

Future works: ANR project JCJC SansSouci (2016-2019)

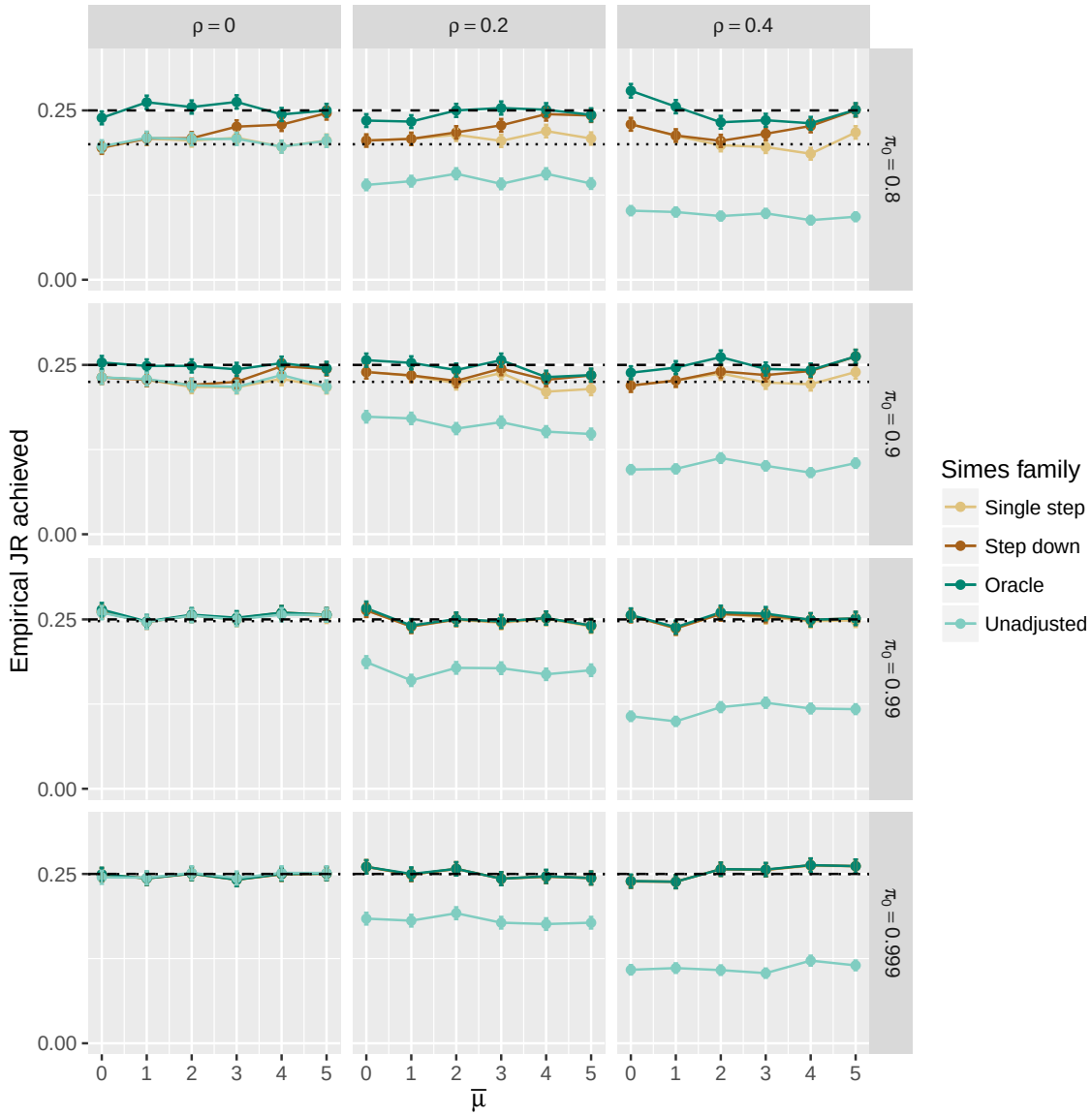
with G. Blanchard, C. Dalmasso, S. Delattre, JF Deleuze, G. Durand, E. Le Floch, M. Martinez, G. Rigaill, E. Roquain, F. Samson.

- Mathematical statistics: asymptotic and finite sample
- Algorithmics: structured rejection sets
- Applications to genomics and neuro-imaging
- Software and visualization tools

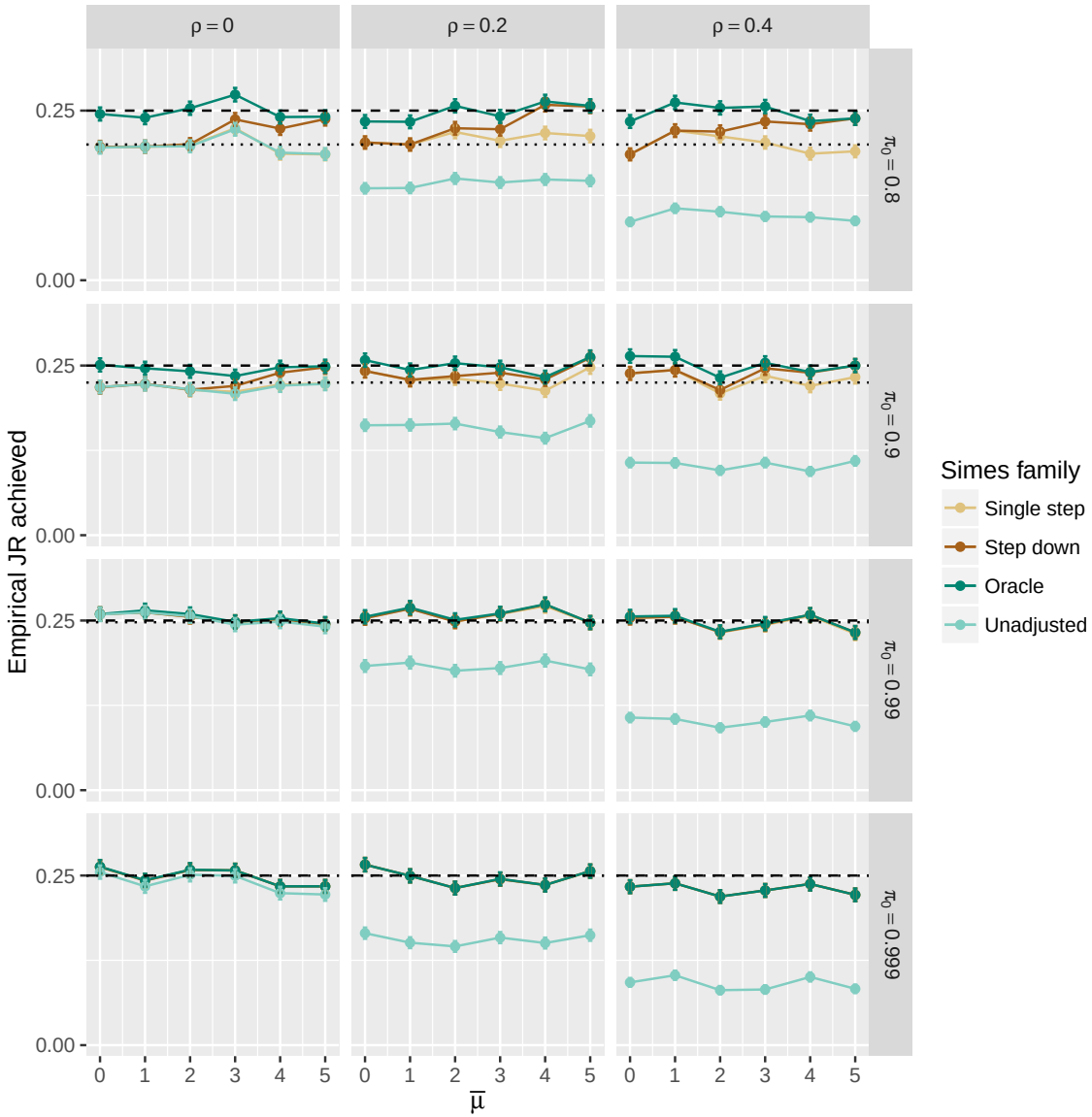
Simes Family, kMax=m



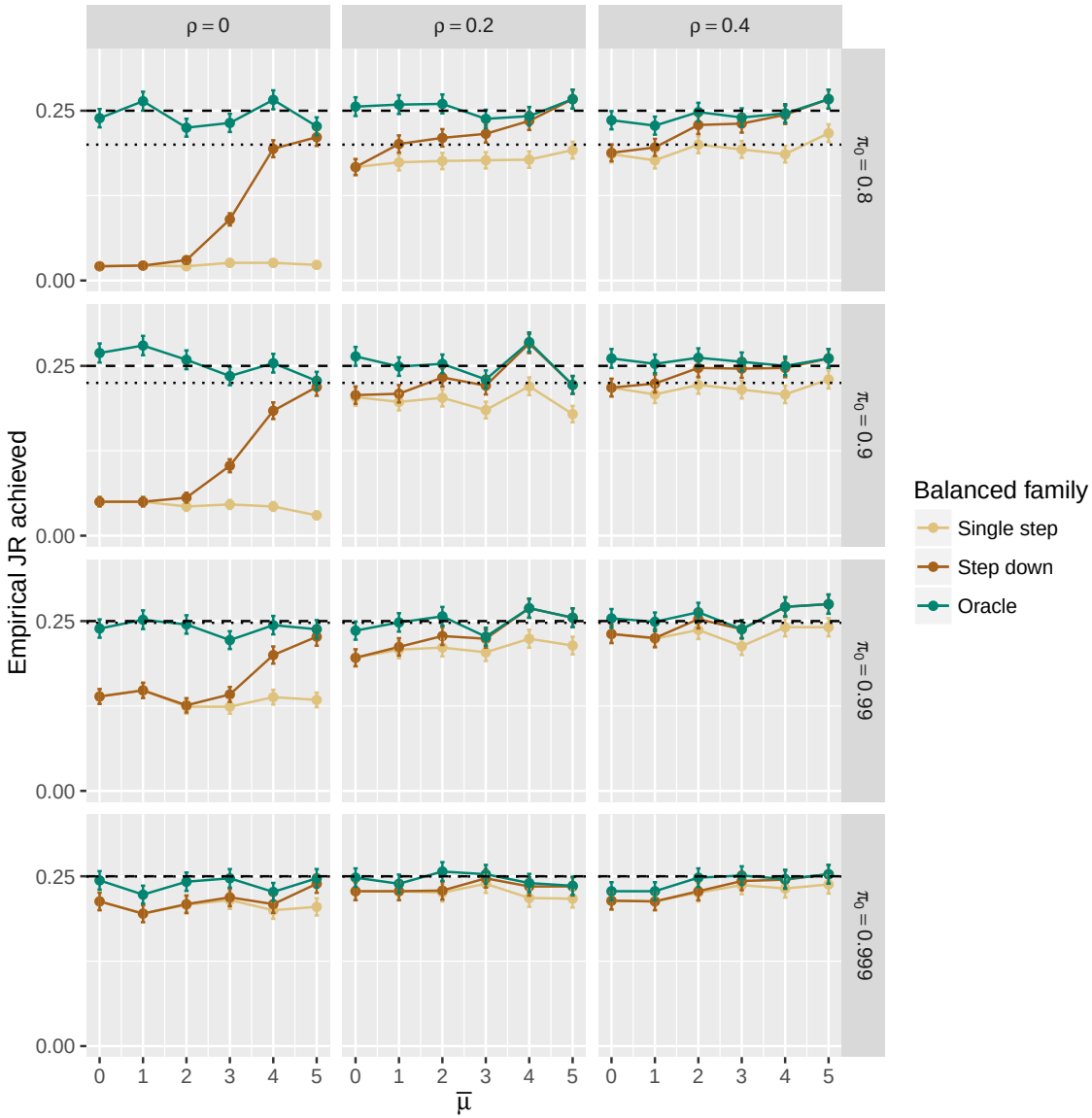
Simes Family, kMax=200



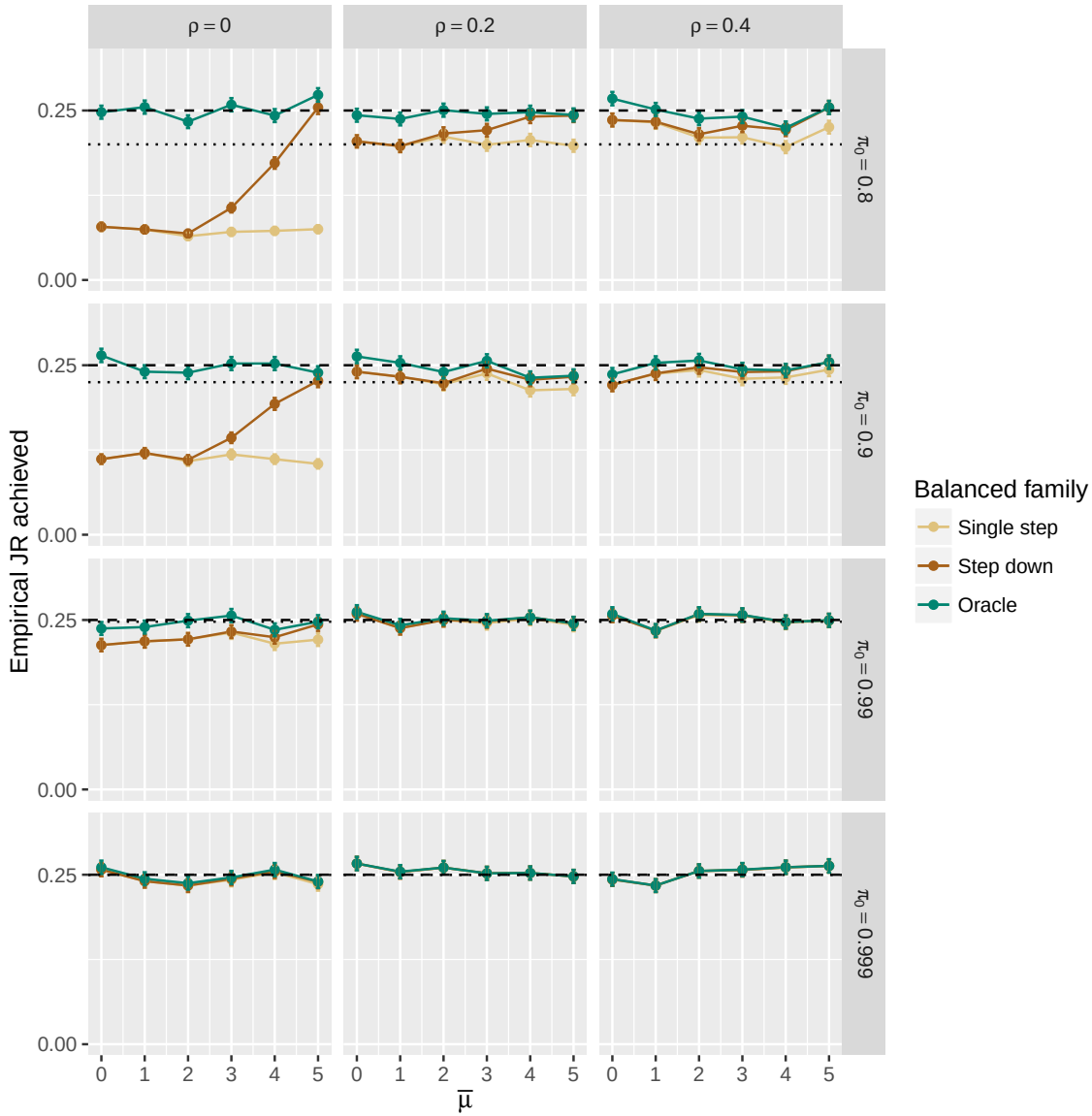
Simes Family, kMax=$2m_1$



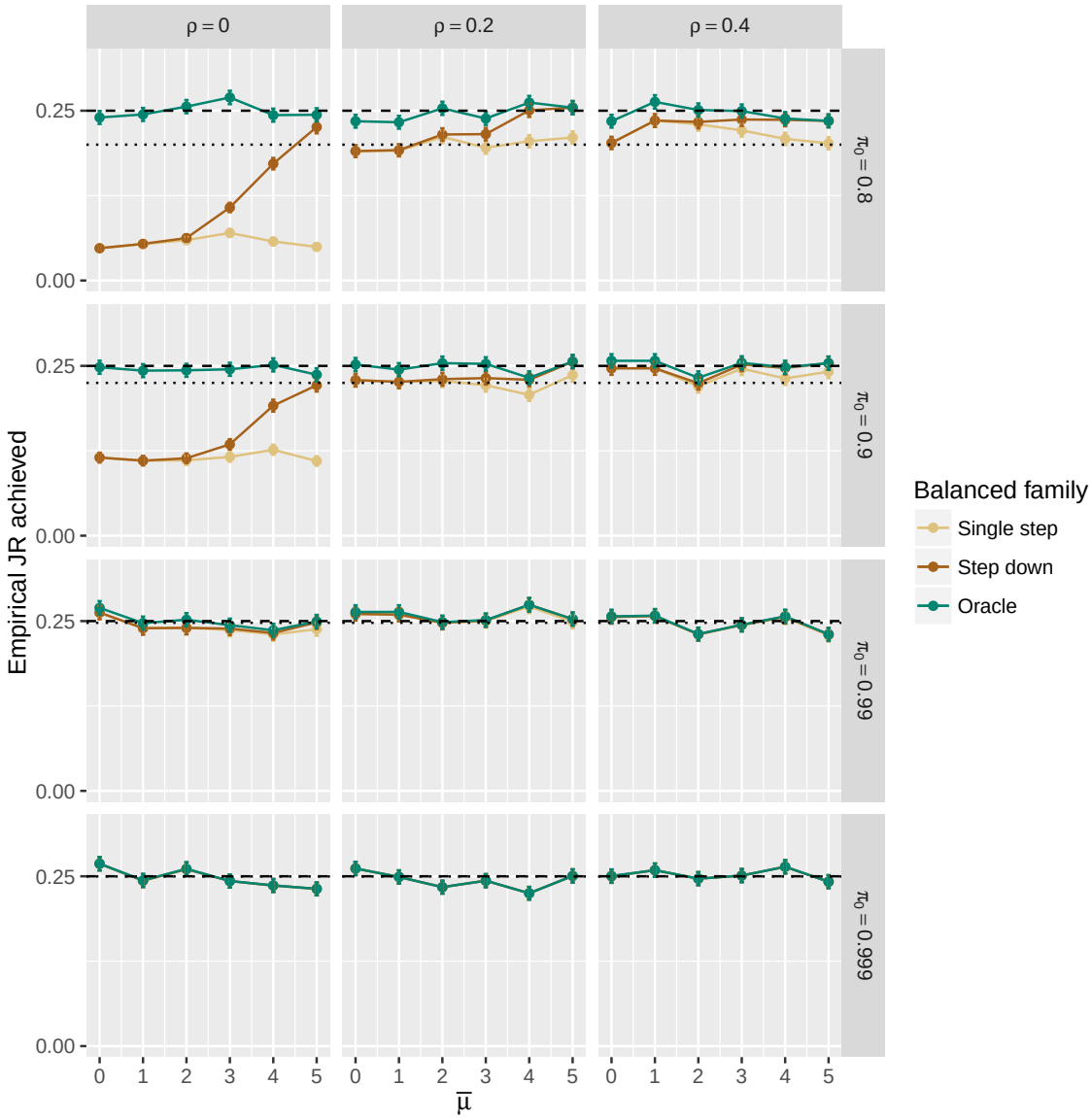
Balanced Family, kMax=m



Balanced Family, kMax=200



Balanced Family, kMax=$2m_1$



Outline

Outline

1. Post hoc inference
   - Motivation
   - State of the art: Goeman and Solari (2011)
2. Joint Family-Wise Error Rate control for post hoc inference
   - A novel risk measure: JFWER
   - Connection to GS2011
3. Obtaining Joint Family-Wise Error Rate control
   - Setting: known dependency
   - Adjustment of a reference threshold family