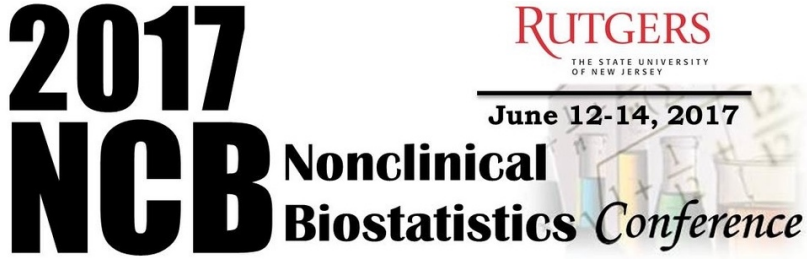
A Selective Inference-based Two-stage Procedure for Clinical Safety Studies

Yalin Zhu

Department of Mathematical Sciences
New Jersey Institute of Technology

yalin.zhu@outlook.com



Examples for Multiple Testing

- Evaluating **efficacy** of a new drug for multiple endpoints.
- Detecting **adverse events** across body system in drug safety analysis.
- Selecting **voxels** on multiple brain regions in fMRI studies.

Multiple families/groups structure are often seen in these experiments!

NJIT New Jersey's Science & Technology University

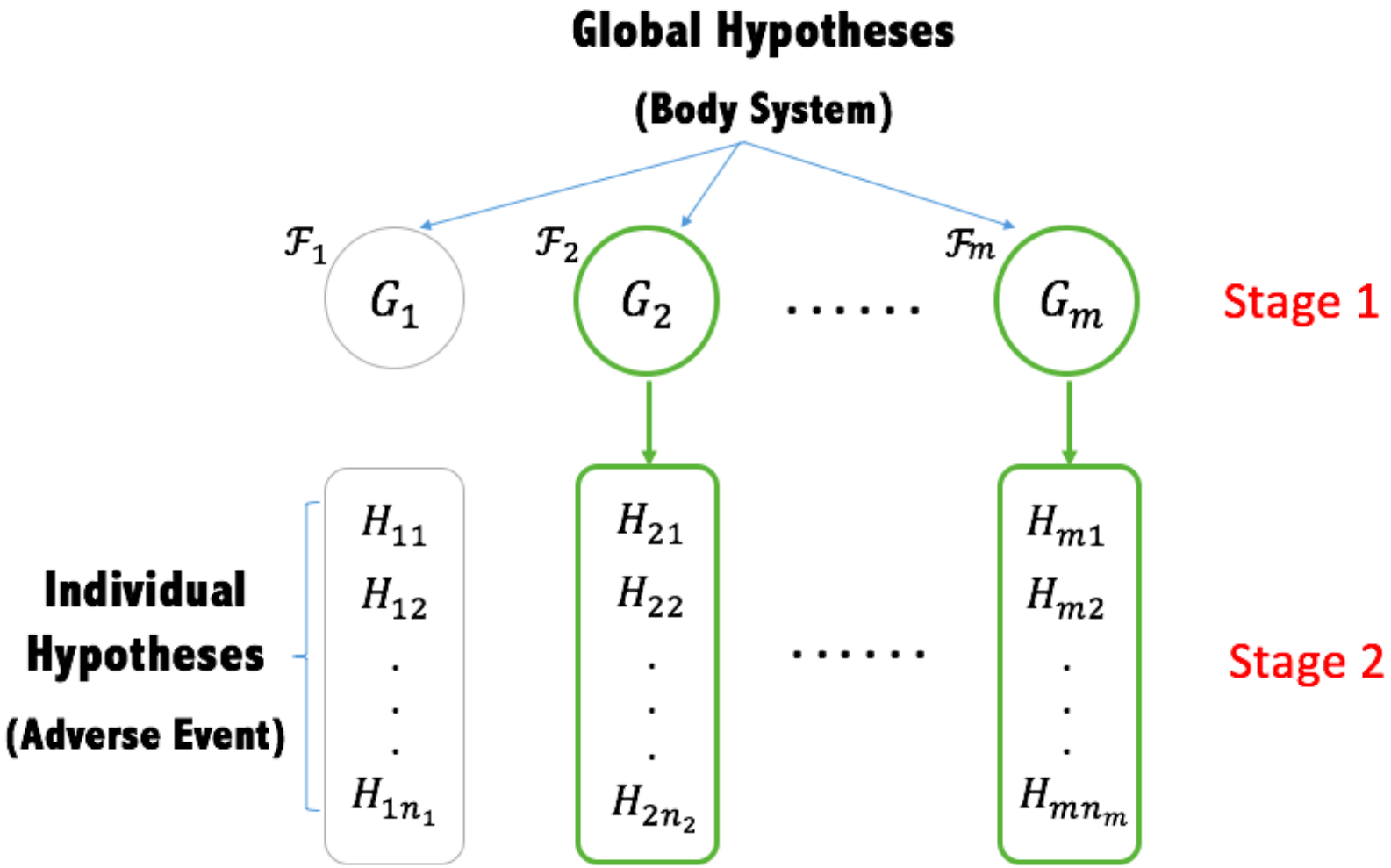
A Motivating Example from Clinical Safety Analysis

- m body systems $BS_1, \ldots, BS_m$.
- n_i adverse event types $AE_1, \ldots, AE_{n_i}$ in the BS_i.
- Goal: **Select** Body Systems of Interest (BSoI).
  Flag significant AEs in BSoI.
- Hypothesis: **Family (BS) level** $G_i : \bigcap_{j=1}^{n_i} H_{ij}$ *vs.* $G_i' : \bigcup_{j=1}^{n_i} H_{ij}'$.
  Individual (AE) level $H_{ij} : \theta_{1ij} = \theta_{2ij}$ *vs.* $H_{ij}' : \theta_{1ij} \neq \theta_{2ij}$.

This safety analysis question can be formulated as a multiple testing problem for multiple families structure!

NJIT New Jersey's Science & Technology University

Addressing Selective Inference



1. **Selection:** Body Systems of Interest (BSoI)
2. **Inference:** Flagging AEs

NJIT New Jersey's Science & Technology University

Research Motivation

- A common two-stage procedure uses the same data for both **selection** and **testing**.
- The selection bias leads to the existing multiple testing procedures in clinical safety studies such as *double FDR* and *modified double FDR* failing to strongly control FDR.
- Most existing approaches for multiple families structure multiple testing only consider **individual hypotheses level** error control, but ignore **family level** error control. *It is natural to consider simultaneously controlling error rates for both levels*

NJIT New Jersey's Science & Technology University

Type 1 Error Rates

- **Family-wise Error Rate:** $FWER = Pr\{reject\ at\ least\ one\ true\ null\}$
  - FWER is suitable for **small scale multiple testing.**
- **Generalized Family-wise Error Rate:**
  $k\text{-}FWER = Pr\{reject\ at\ least\ k\ true\ nulls\}$
  - k-FWER is suitable for **moderate scale multiple testing.**
- **False Discovery Rate:** $FDR = E\left\{\frac{\#\ of\ rejected\ true\ nulls}{\#\ of\ rejected\ all\ hypotheses}\right\}$
  - FDR is suitable for **large scale multiple testing.**

NJIT New Jersey's Science & Technology University

Stepwise Multiple Testing Procedures (MTPs)

Ordered p-value based stepwise MTP, which is described by using a sequence of non-decreasing critical constants $\alpha_1 \leq \ldots \leq \alpha_m$.

- **k-FWER Controlling MTPs**
  - Generalized Bonferroni procedure: $\alpha_i = \dfrac{k\alpha}{m}$
  - Generalized Sidak procedure: α_i satisfies $\sum_{j=k}^{m} \binom{m}{j} \alpha_i^j (1-\alpha_i)^{m-j} = \alpha$
- **FDR Controlling MTPs**
  - **Step-up:** Benjamini-Hochberg (BH) procedure: $\alpha_i = \dfrac{i\alpha}{m}$

    Reject H_i if $p_i \leq \frac{R}{m}\alpha$, where $R = \max\{1 \leq i \leq m : p_{(i)} \leq \dfrac{i\alpha}{m}\}$.

NJIT New Jersey's Science & Technology University

Problem Formulation

For family index $i = 1, \ldots, m$; individual index $j = 1, \ldots, n_i$,

- $\tilde{p}_i$: p-value for the corresponding global hypothesis G_i.
- $\tilde{V}$: number of false selections when selecting families.
- $\mathcal{S} \subseteq \{1, \ldots, m\}$: set of selected families.
- $|\mathcal{S}|$: the number of total selections.

- p_{ij}: p-value for the corresponding individual hypothesis H_{ij}.
- V_i: number of false rejections for i-th family.
- R_i: number of total rejections for i-th family.

NJIT New Jersey's Science & Technology University

Error Rates Guaranteed in This Work

- **Generalized FWER:** (across families)

$$k\text{-}FWER = Pr\{\tilde{V} \geq k\}$$

- **Conditional FDR for i-th family:** (within families)

$$cFDR_i = E\left\{\frac{V_i}{R_i \vee 1} \middle| i \in \mathcal{S}\right\}.$$

- **Average FDR over selected families:** (overall)

$$average\text{-}FDR = E\left\{\frac{\sum_{i\in\mathcal{S}} \frac{V_i}{R_i \vee 1}}{|\mathcal{S}| \vee 1}\right\}$$

NJIT New Jersey's Science & Technology University

Related Literatures

1. **Selective (Post-selection) Inference**
   - Benjamini, Yekutieli (2005) False Discovery Rate: Adjusted Multiple Confidence Intervals for Selected Parameters. JASA 100(469):71-81.
   - Fithian, Sun, Taylor (2015) Optimal inference after model selection. arXiv:1410.2597.
   - Lee, Sun, Sun, Taylor (2013) Exact post-selection inference with the lasso. arXiv:1311.6238.

2. **Multiple Families Multiple Testing**
   - Hu et al. (2010) False discovery rate control with groups. JASA 105(491):1215-27.
   - Benjamini, Bogomolov (2014) Selective inference on multiple families of hypotheses. JRSSB 76(1):297-318.
   - Heller et al. (2016) Post-selection inference following aggregate level hypothesis testing in large scale genomic data. bioRxiv:058404.

NJIT New Jersey's Science & Technology University

A Selective Inference-Based Two-Stage Procedure

Procedure 1 (cFDR-α-minP-k-Sidak-α_1)

1. (a) *For each body system, compute the global p-value* $\tilde{p}_i = n_i \min_{1 \le j \le n_i} \{p_{ij}\}$.
   (b) *Apply generalized Sidak procedure on* $\tilde{p}_1, \ldots, \tilde{p}_m$ *at level* α_1.
2. (a) *In the i-th selected body system, calculate the conditional p-value for* H_{ij}*:* $p'_{ij} = \frac{p_{ij}}{t_i}$ *if* $\min_{1 \le s \le n_i, s \ne j} \{p_{is}\} > t_i$*; otherwise,* $p'_{ij} = p_{ij}$*, where* $t_i = 1 - (1 - \tilde{t})^{\frac{1}{n_i}}$.
   (b) *Apply BH procedure on* $p'_{i1}, \ldots, p'_{in_i}$ *at level* α.

NJIT New Jersey's Science & Technology University

Several Remarks

1. The selection rule: $\tilde{p}_i \leq \tilde{t}$ is equivalent to minP combining function $f(p_{i1}, \ldots, p_{in_i}) = \min_{1 \leq j \leq n_i} \{p_{ij}\} \leq t_i$.
2. The conditional p-value is inflated from original p-value, $P'_{ij} := P_{ij} | \min\{p_{i1}, \ldots, p_{ij-1}, P_{ij}, p_{ij+1}, \ldots, p_{in_i}\} \leq t_i = \frac{P_{ij}}{b_{ij}}$, then observed $p'_{ij} = p_{ij}/b_{ij}$, where the inflation factor
$$b_{ij} = \begin{cases} t_i & \text{if } \min_{1 \leq s \leq n_i, s \neq j} \{p_{is}\} > t_i, \\ 1 & \text{otherwise.} \end{cases}$$
3. If the individual p-value $p_{ij} \sim U(0,1) \longrightarrow$ the conditional p-value $p'_{ij} \sim U(0,1)$

NJIT New Jersey's Science & Technology University

Error Rates Control

Theorem 1 (k-FWER control)

If global p-values $\tilde{P}_1, \ldots, \tilde{P}_m$ are independent p-values with $U(0,1)$ under true null, then Procedure 1 strongly controls the k-FWER at level α_1 across body systems.

Theorem 2 (cFDR control)

For each selected body system $\mathcal{F}_i$, if individual p-values $P_{i1}, \ldots, P_{in_i}$ are independent p-values with $U(0,1)$ under true null, then Procedure 1 strongly controls the conditional FDR at level α.

Corollary 1 (average-FDR control)

Under the assumption of Theorem 1 and 2, Procedure 1 strongly controls the average FDR at level α.

Technology University

Simulation Settings I: Independence

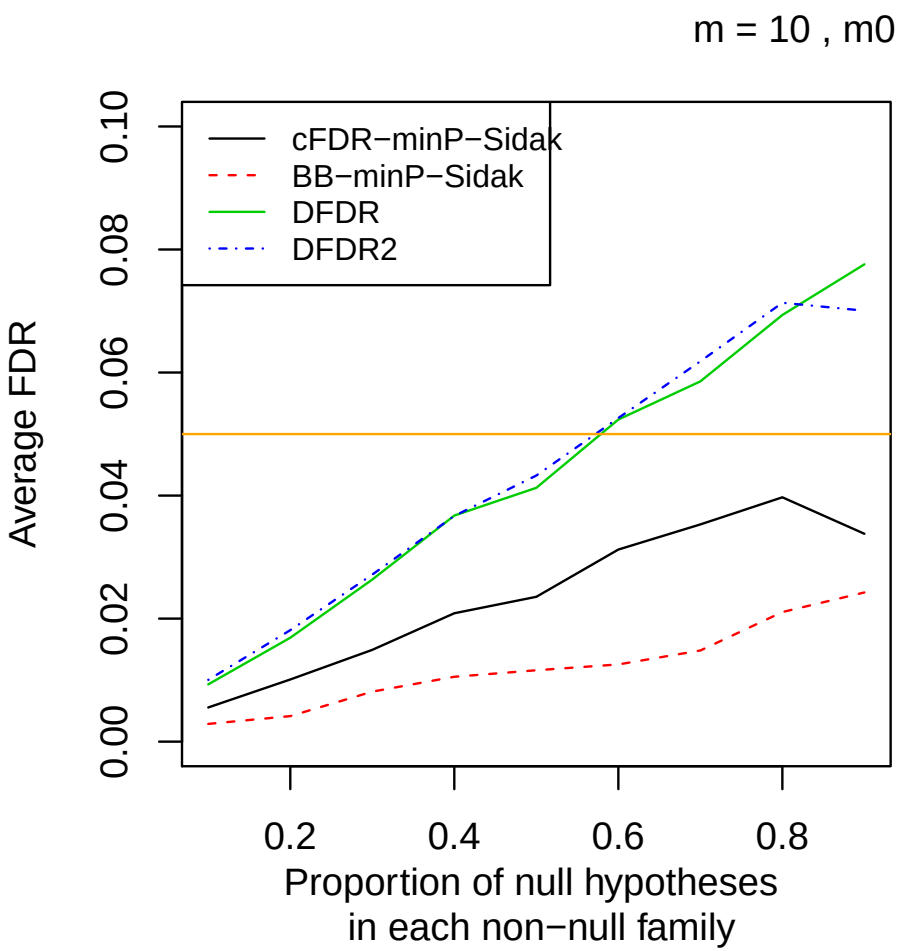
- m: number of BS; n: number of AE.
- Generate $m \times n$ independent normal r.v. matrix with $X_{ij} \sim N(\mu_{ij}, 1)$, where $i = 1, \ldots, m; j = 1, \ldots, n$.
- $m = 10$, $n = 20$, $m_0 = \{2, 4, 6, 8\}$; $n_0 = \{5, 10, 15\}$.

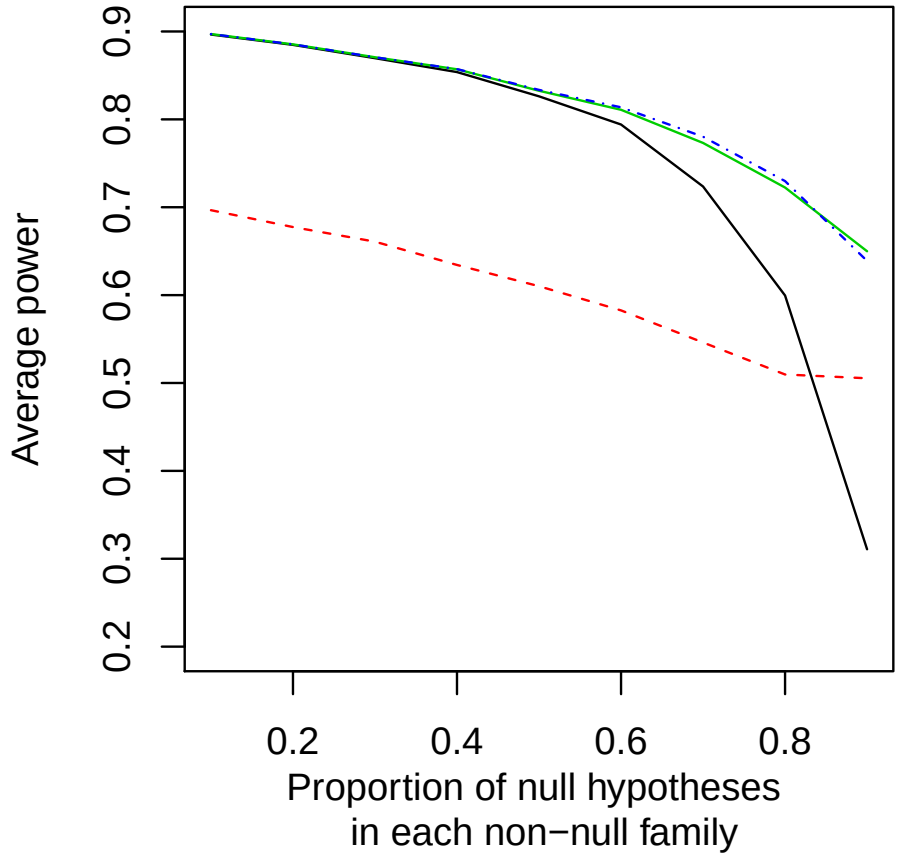
- Test $H_{ij} : \mu_{ij} = 0$ versus $H'_{ij} : \mu_{ij} > 0$.
- Set $\mu_{ij} = 3$ for $i = 1, \ldots, m - m_0; j = 1, \ldots, n - n_0$; the rest of $\mu_{ij} = 0$.

- Selection by generalized Sidak with $k = 1, 2, 3$ and $\alpha_1 = \alpha = 0.05$.

NJIT New Jersey's Science & Technology University

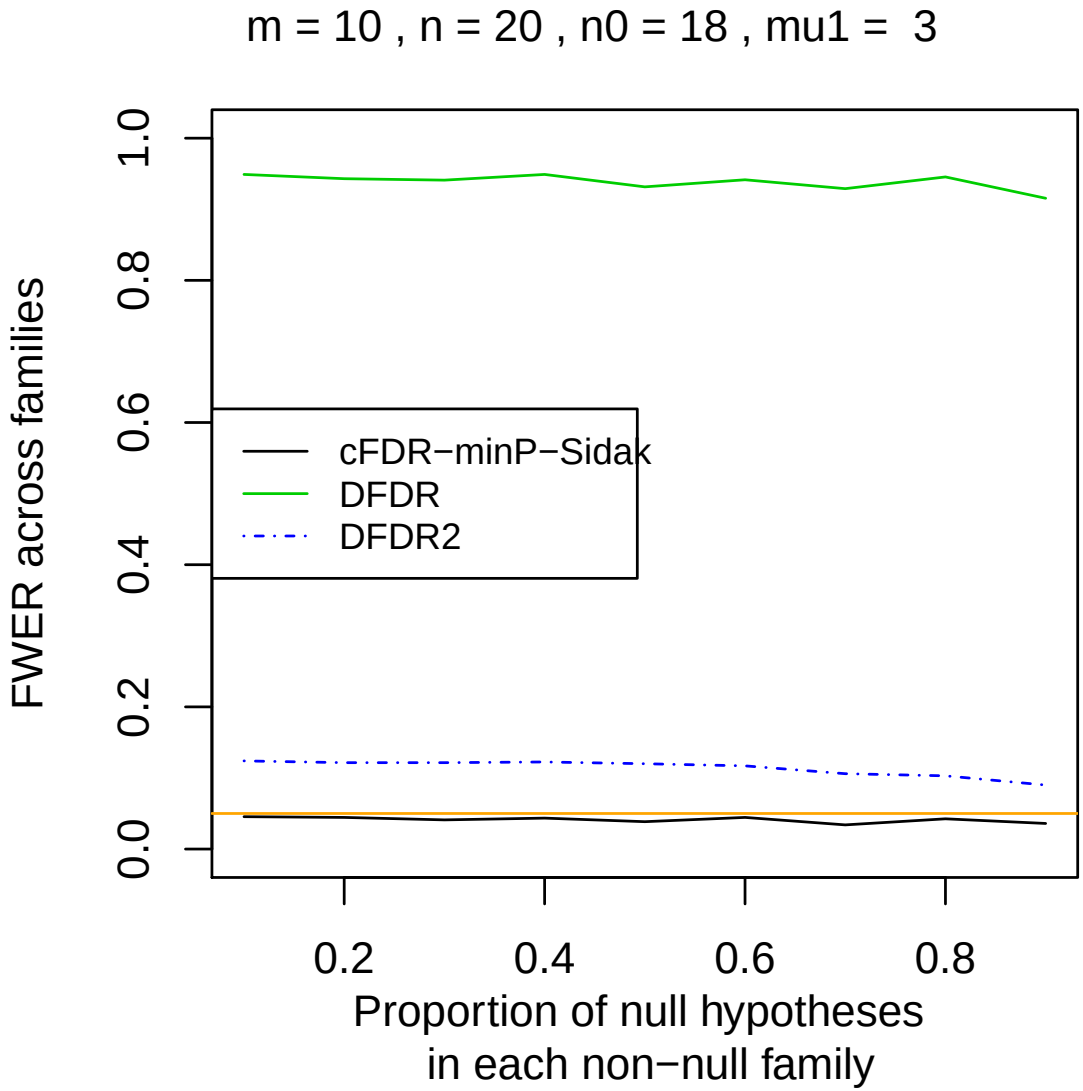
Simulation Results ($k = 1$)



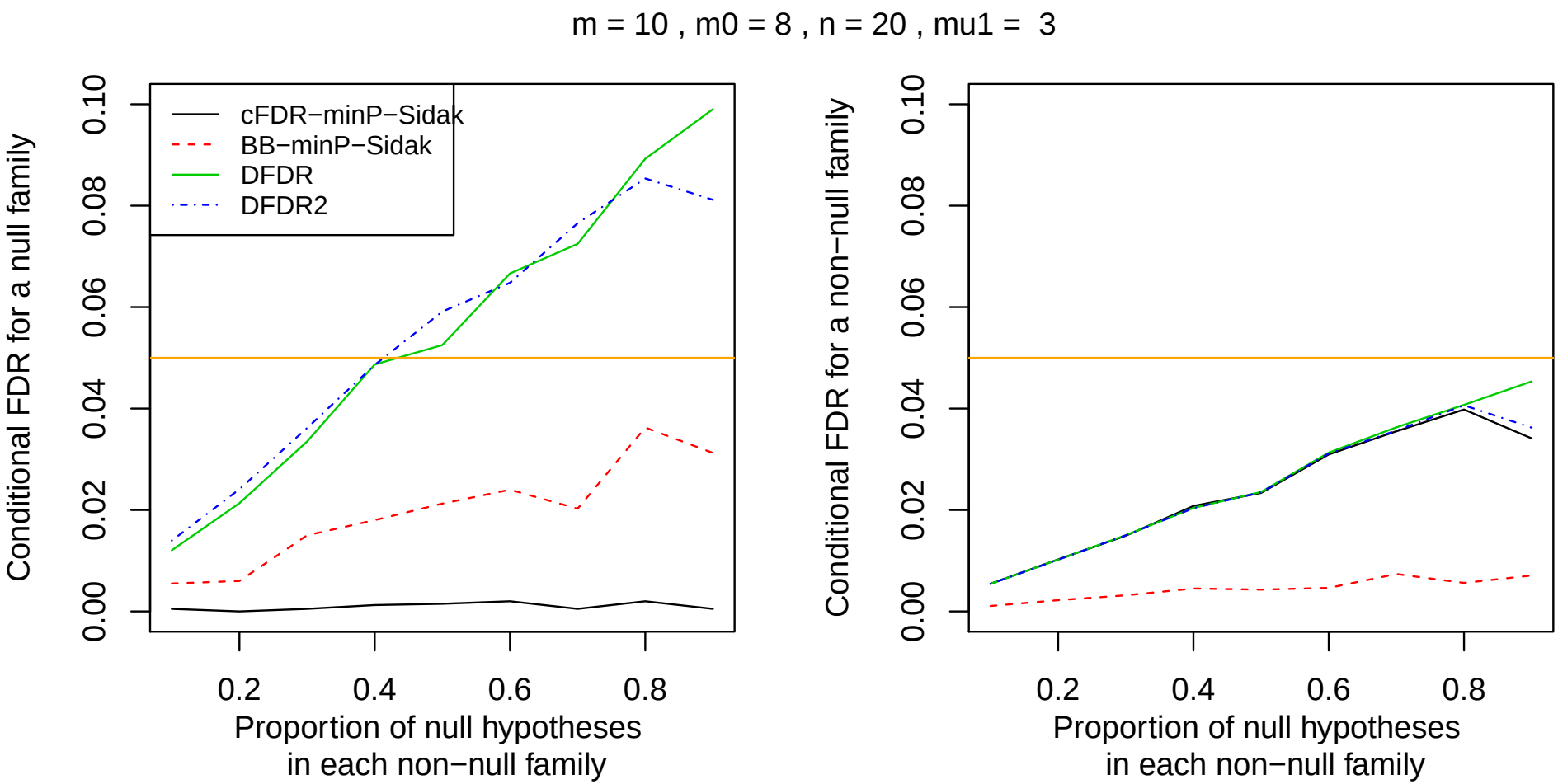


Technology University

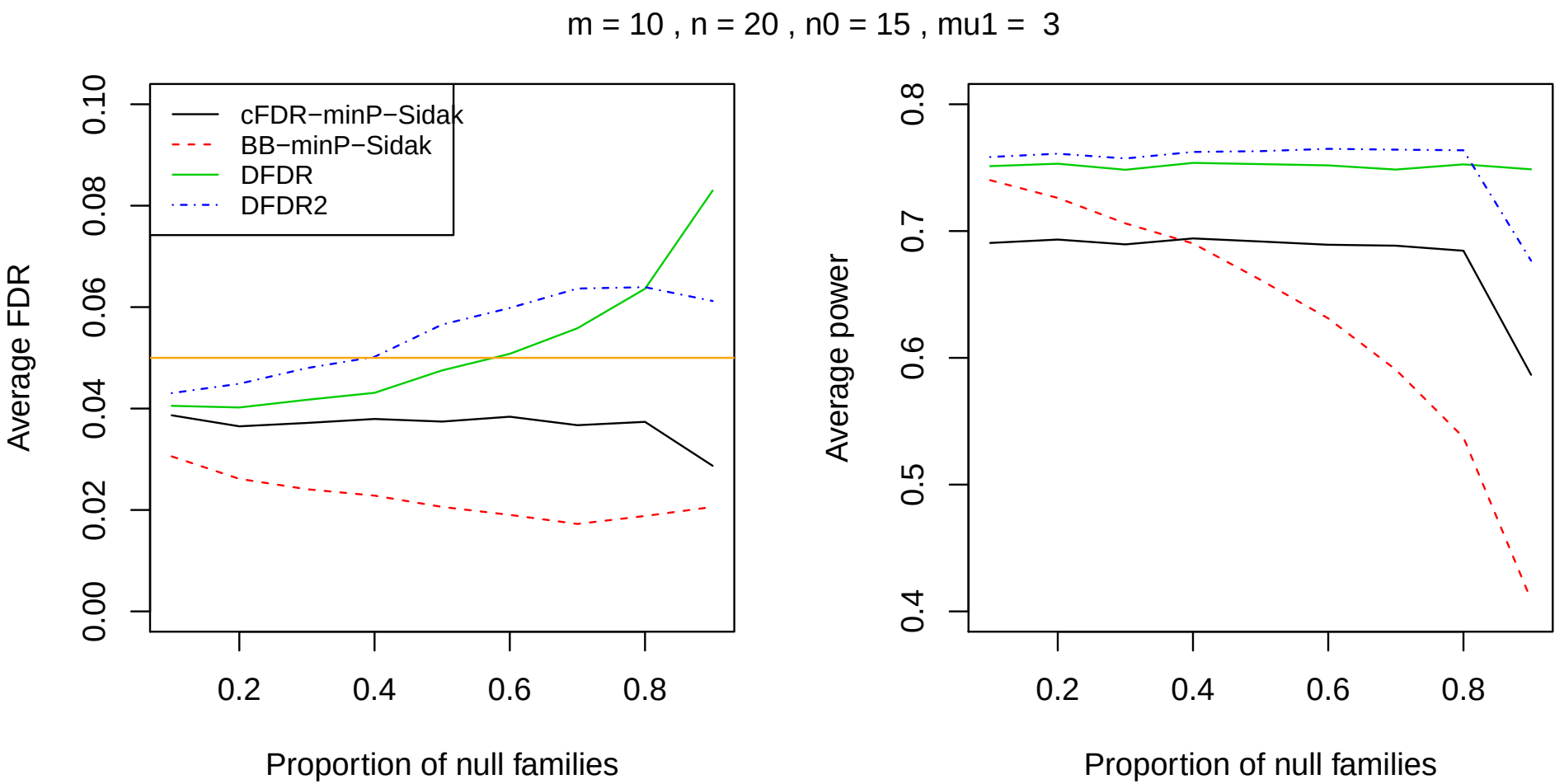
Simulation Results ($k = 1$)



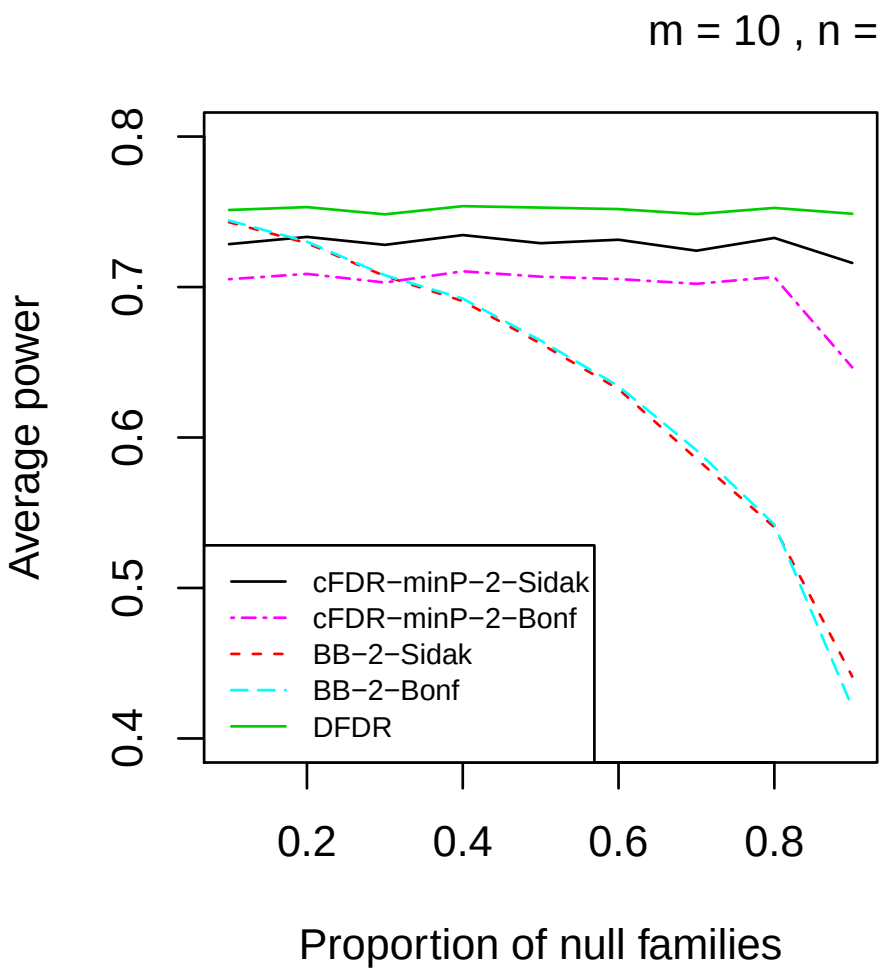
Simulation Results ($k = 1$)

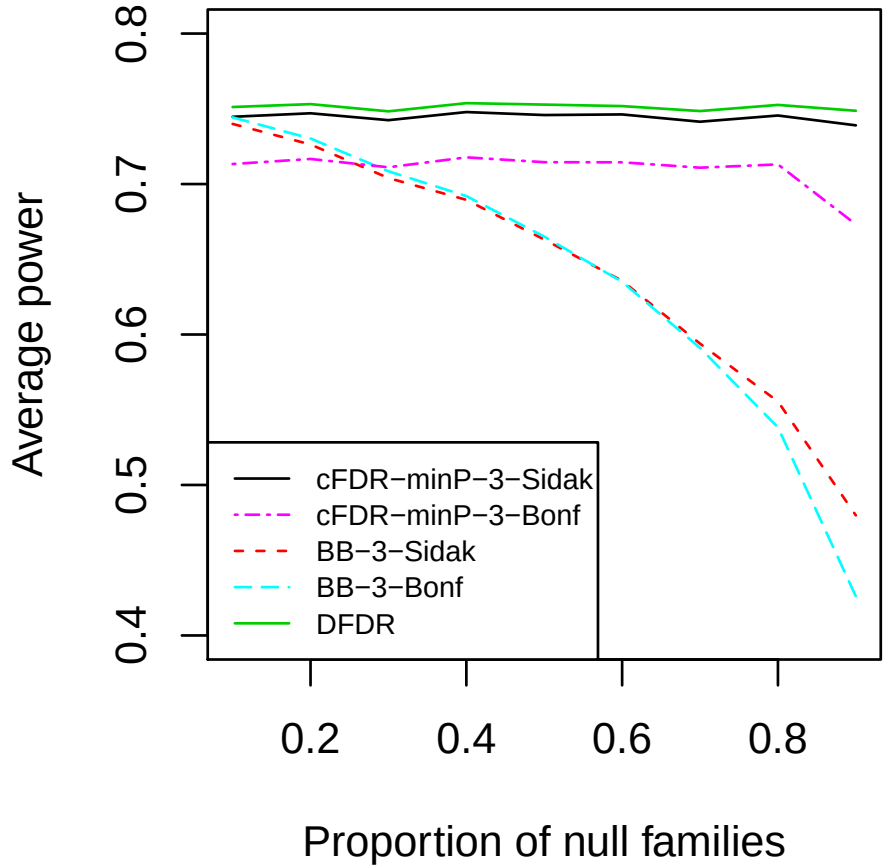


Simulation Results ($k = 1$)



Selection Rule: Bonferroni v.s. Sidak ($k = 2, k = 3$)





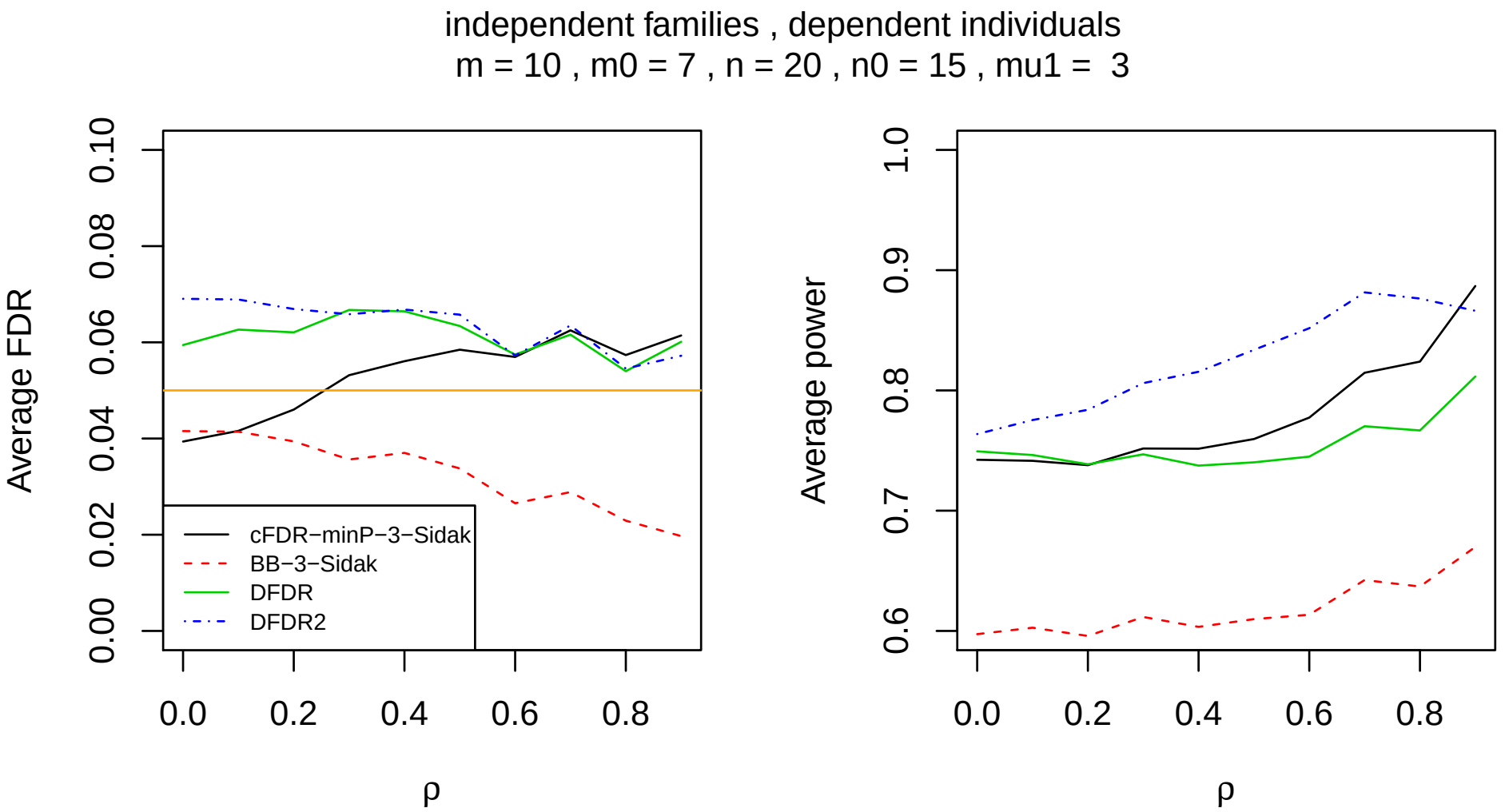
Simulation Settings II: Dependence

- The p-values are **dependent within one body system** with equal correlation and **independent** of the p-values in **other body systems**.

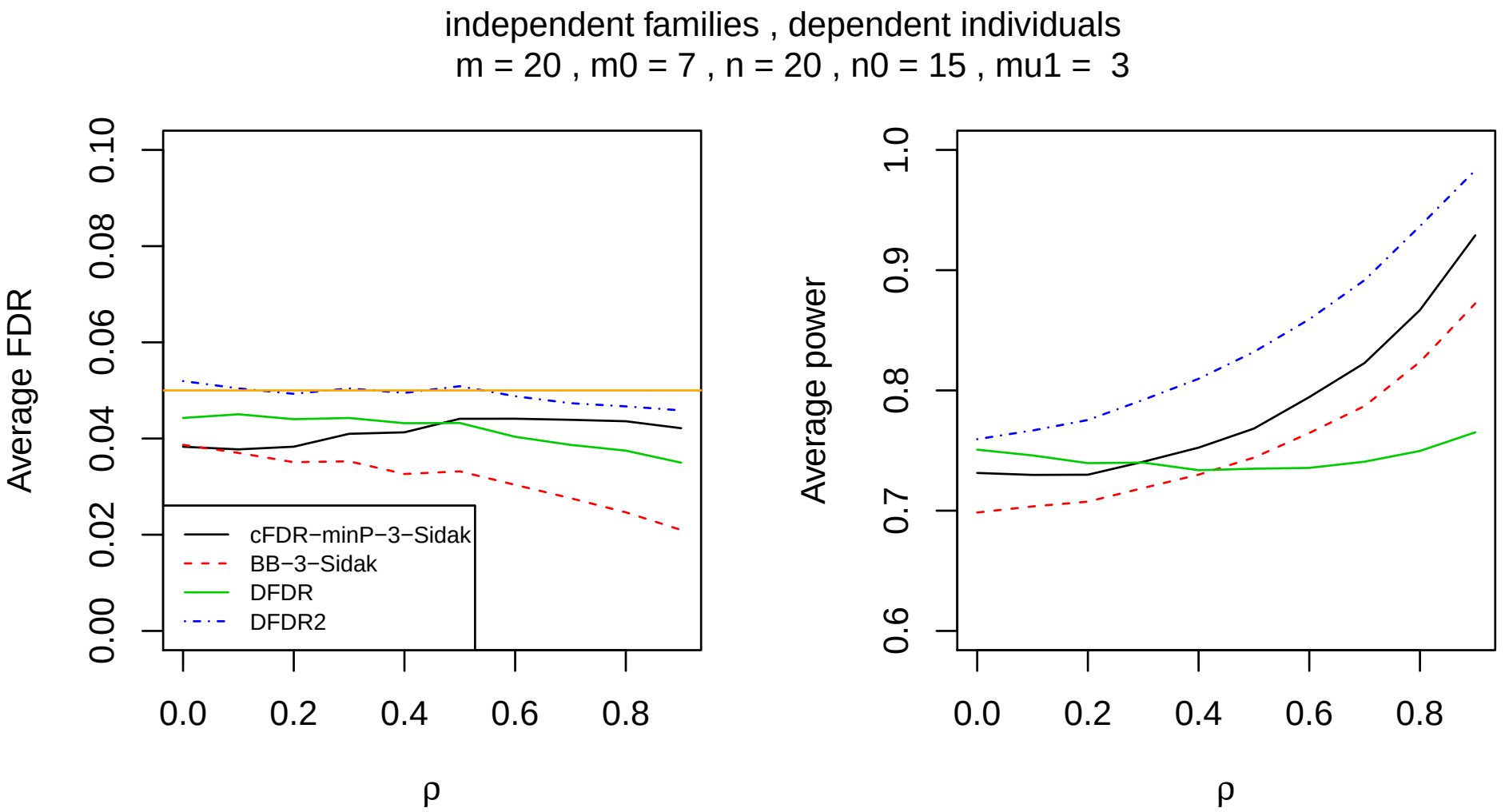
- $m = \{10, 20\}$, $m_0 = 7$, $n = 20$, $n_0 = 15$, $\rho = \{0, 0.1, \ldots, 0.9\}$.
- $\alpha_1 = \alpha = 0.05$.
- $B = 2000$ iterations.

NJIT New Jersey's Science & Technology University

Simulation Results ($m = 10$)



Simulation Results ($m = 20$)



Simulation Summary

- The proposed procedure controls **average FDR** along with k-**FWER across families** and **conditional FDR within a family** under all independent and moderate dependent scenarios.

- DFDR and DFDR2 **cannot** control *cFDR for null families* and *average-FDR* under all scenarios! BB procedure **cannot** guarantee *cFDR control for null families* when using k-FWER selection rule with $k > 1$.

- The proposed procedure is more powerful than BB procedure with the same selection rule.

NJIT New Jersey's Science & Technology University

Back to the Motivating Example

Revisit the safety analysis of a candidate vaccine against measles, mumps, rubella and varicella (MMRV)[1].

- Monitor 40 **Tier 2 AE types** from $m = 8$ **body systems**
- $N_1 = 148 \longrightarrow \text{MMR} \longrightarrow X_{1i}$
  $N_2 = 132 \longrightarrow \text{MMRV} \longrightarrow X_{2i}$
- Two-sided **Fisher's Exact Test**
- $\alpha_1 = 0.05$, $\alpha = 0.1$

[1] The paper of Mehrotra and Heyse (2004)

NJIT New Jersey's Science & Technology University

A Safety Analysis Example

Table 1: The example from Mehrotra and Heyse (2004) under $\alpha_1 = 0.05$ for selecting BSoI and $\alpha = 0.1$ for flagging AEs.

Approach	BSoI	Flagging AE
Naive BH	NA	503 (1)
DFDR	5 (1)	503 (1)
DFDR2	5 (1)	503 (1)
GBH	NA	503 (1)
Original BB	5 (1)	503 (1)
cFDR-minP-0.05	2,5,7 (3)	0
cFDR-minP-Sidak-0.05	0	0
cFDR-minP-3-Sidak-0.05	5 (1)	0

NJIT New Jersey's Science & Technology University

Another Safety Analysis Example

Table 2: The Example 4.2 from Mehrotra and Adewale (2012) under $\alpha_1 = 0.05$ for selecting BSoI and $\alpha = 0.1$ for flagging AEs. The study monitored 49 AE types across 9 body systems, $N_1 = N_2 = 1616$.

Approach	BSoI	Flagging AE
Naive BH	NA	703 (1)
DFDR	3, 7 (2)	301, 704 (2)
DFDR2	7 (1)	703, 704 (2)
GBH	NA	301, 401, 703, 704 (4)
Original BB	7 (1)	703 (1)
cFDR-minP-0.05	1, 2, 3, 4, 6, 7 (6)	301, 703, 704 (3)
cFDR-minP-Sidak-0.05	7 (1)	703, 704 (2)
cFDR-minP-3-Sidak-0.05	3, 7 (2)	703, 704 (2)

NJIT New Jersey's Science & Technology University

Comparisons for Different Approaches

Table 3: Error rates control for different MTPs with multiple families structure

Approach	Family level	Within family	Overall
DFDR	×	×	×
DFDR2	×	×	×
GBH	×	×	global FDR
Original BB	FDR	×	average FDR
cFDR-minP-t (fixed)	×	conditional FDR	average FDR
cFDR-minP-k-Sidak	k-FWER	conditional FDR	average FDR

Discoveries across **body systems selection** (BSoI) is considerable important and should be given more attentions for future research!

NJIT New Jersey's Science & Technology University

R Package: MHTmult

▸ R package: MHTmult a convenient computing tool to help users make decisions more efficiently.

- Calculating adjusted (conditional) *p*-values for **multiple families MTPs**:
  `cFDR.cp.adjust(); DFDR.p.adjust(); DFDR2.p.adjust(); avgFDR.p.adjust(); GBH.p.adjust()`
- Calculating adjusted *p*-values for **k-FWER controlling MTPs**:
  `gbonf.p.adjust(); gsidak.p.adjust()`
- Options:
  1. Selection rule (combining method, selecting procedure)
  2. Decision-making
  3. Visualization

NJIT New Jersey's Science & Technology University

Summary

1. We developed a two-stage multiple testing procedure using minimum p-value to selecting body systems, and efficiently flag the AE types.
2. We found out the procedure can guarantee desired type 1 error rate control **across family level**, **within selected families** and **overall average on selected families**.
3. Simulation studies and real data analysis illustrate that the proposed procedures outperform the existing procedures in terms of the error rates control and power. *DFDR procedure cannot control the FDR!*
4. Computing tools such as R Package ▸ package: MHTmult can be easily used, one web application with visualization function is still on work.

NJIT New Jersey's Science & Technology University

Key References

[1] Benjamini, Y. and Bogomolov, M. (2014) Selective inference on multiple families of hypotheses. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76, 297–318.

[2] Guo, W. and Romano, J. (2007) A generalized Sidak-Holm procedure and control of generalized error rates under independence. Statistical Applications in Genetics and Molecular Biology, 6, 1.

[3] Heller, R., Chatterjee, N., Krieger, A., and Shi, J. (2016) Post-selection inference following aggregate level hypothesis testing in large scale genomic data. bioRxiv, 058404.

[4] Mehrotra, D. V. and Adewale, A. J. (2012) Flagging clinical adverse experiences: reducing false discoveries without materially compromising power for detecting true signals. Statistics in Medicine, 31, 1918–1930.

[5] Mehrotra, D. V. and Heyse, J. F. (2004). Use of the false discovery rate for evaluating clinical safety data. Statistical Methods in Medical Research, 13, 227-238.

NJIT New Jersey's Science & Technology University

Acknowledgment

- Dr. Wenge Guo, New Jersey Institute of Technology
- Dr. Steven Novick, MedImmune Inc.

NJIT New Jersey's Science & Technology University