Adjusting multiplicity using safety data in many-one comparisons

G. Hlavin, P. Bauer, F. Koenig

Section for Medical Statistics - Medical University of Vienna

June 24, 2015

This project has received funding from the European Union's 7th Framework Programme for research, technological development and demonstration under Grant Agreement no 602552.

Introduction

- We consider the problem of multiple comparisons between N treatments and a control.
- Safety is an issue in all phases of clinical trials!
  1. treatments may be dropped due to safety issues
  2. multiplicity adjustment for the remaining treatments
  3. Especially in small populations it is paramount not to "waste" parts of the significance levels for unsafe groups.
- *Do we still have to pay a price as far as multiplicity is concerned due to the dropping of unsafe treatments, which are no longer of interest for efficacy testing?*

Model and Notations

- $N + 1$ groups
  - $i \in \{1, \ldots, N\} =: I$ ... treatment groups
  - $i = 0$ ... control group
- x_{ij} ... efficacy measurement of patient j in group i
  - $x_{ij} \sim N(\mu_i, \sigma^2)$, independent across the i and the j dimension
  - $\mathbf{x}_i$ ... vector of group i efficacy data
  - $\mathbf{X}_K$ ... $\{\mathbf{x}_k : k \in K\}$
- y_{ij} ... toxicity measurement of patient j in group i
  - $\mathbf{y}_i$ ... vector of group i toxicity data
- N potential $H_0^i : \mu_i - \mu_0 \geq 0$

The Naive Test Procedure φ_ν (Safety Selection Step)

Two-Step Procedure: First safety screening, then efficacy testing

We start with N=5 Treatments vs. Control

	Tr. 1	Tr. 2	Tr. 3	Tr. 4	Tr. 5
Safety Selection	$\bar{y}_1 \leq t_1$	$\bar{y}_2 \leq t_2$	$\bar{y}_3 > t_3$	$\bar{y}_4 > t_4$	$\bar{y}_5 \leq t_5$

- treatments with mean toxicity exceeding a pre-fixed threshold t_i are regarded as unsafe and therefore dropped

The Naive Test Procedure φ_ν (Efficacy Testing Step)

Two-Step Procedure: First safety screening, then efficacy testing

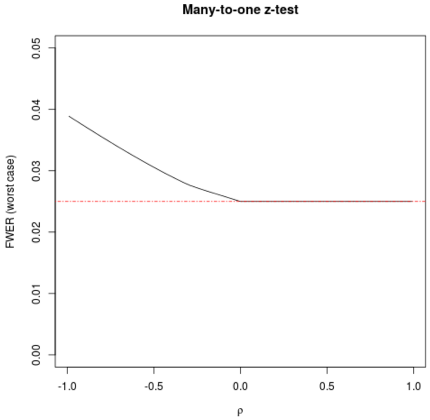
We start with N=5 Treatments vs. Control

	Tr. 1	Tr. 2	Tr. 3	Tr. 4	Tr. 5
Safety Selection	$\bar{y}_1 \leq t_1$	$\bar{y}_2 \leq t_2$	$\bar{y}_3 > t_3$	$\bar{y}_4 > t_4$	$\bar{y}_5 \leq t_5$
	↓	↓	✖	✖	↓
Efficacy Tests	Dunnett-T. $\alpha = \alpha_{nom}$ 3 Tr. vs. 1 Ctrl	Dunnett-T. $\alpha = \alpha_{nom}$ 3 Tr. vs. 1 Ctrl			Dunnett-T. $\alpha = \alpha_{nom}$ 3 Tr. vs. 1 Ctrl

- Only the remaining 3 safe treatments are each tested with a Dunnet-Test for 3 treatment-control comparisons.

wcFWER Control for the Naive Test Procedure

$wcFWER_{\varphi_\nu}$ is defined as the max $FWER_{\varphi_\nu}$ for varying thresholds t_i



- 2 treatments vs control (balanced)
- x_{ij} and y_{ij} bivariate normal with correlation ρ, variance known

Positive Relation

Definition: Positive Relation

It holds *positive relation* [König et al., 2006] between efficacy and toxicity, if for all componentwise non-*de*creasing $f \geq 0$ and for all componentwise non-*in*creasing $g \geq 0$ and all $i \in \{1, \ldots, N\}$ it holds:

$$\mathbb{E}[f(\mathbf{x}_i) \cdot g(\mathbf{y}_i)] \leq \mathbb{E}[f(\mathbf{x}_i)] \cdot \mathbb{E}[g(\mathbf{y}_i)]$$

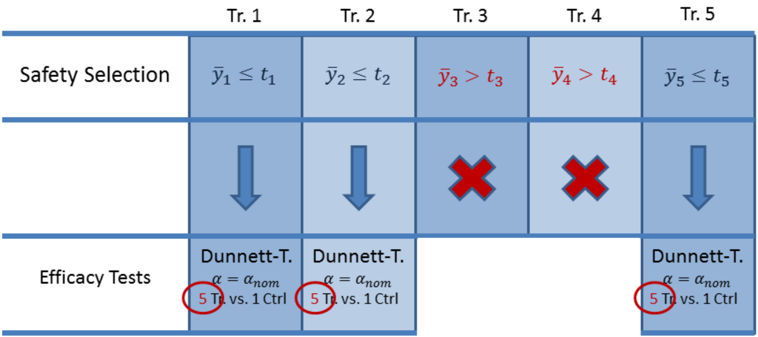
- slight generalization of *association* [Esary et al., 1967]
- two important examples
  1. x_{ij} and y_{ij} are bivariate normal with $\rho_i \geq 0$
  2. y_{ij} is a binary indicator and $P(y_{ij} = 1 | x_{ij})$ is non-decreasing

Theorem:

Under positive relation between efficacy and toxicity, $wcFWER_{\phi_\nu}$ is controlled at level α_{nom}.

Procedures, which control the *wcFWER*:
1. Conservative Two-Step Procedure φ_c

We start with N=5 Treatments vs. Control



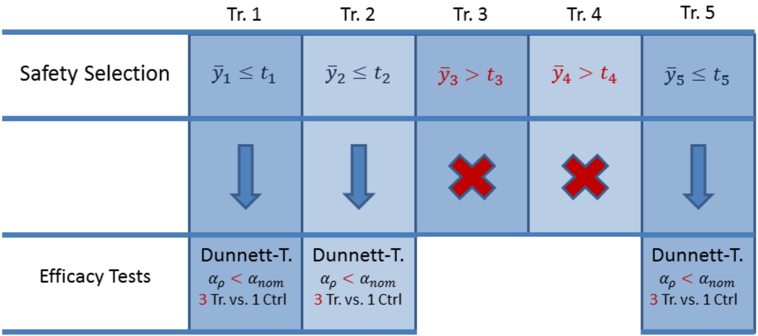
- $wcFWER_{\varphi_c} \leq \alpha_{\text{nom}}$
- severe power-loss can be expected, if N is larger than the number of treatments expected to be selected due to safety

Procedures, which control the *wcFWER*: 2. Naive Two-Step Procedure φ_ν when positive relation holds

- positive relation holds when it is known that $\rho_i \geq 0$ for $\forall i \in \{1, \ldots, N\}$
- In this case $wcFWER_{\varphi_\nu} \leq \alpha_{\mathrm{nom}}$
- Increased power compared to φ_c

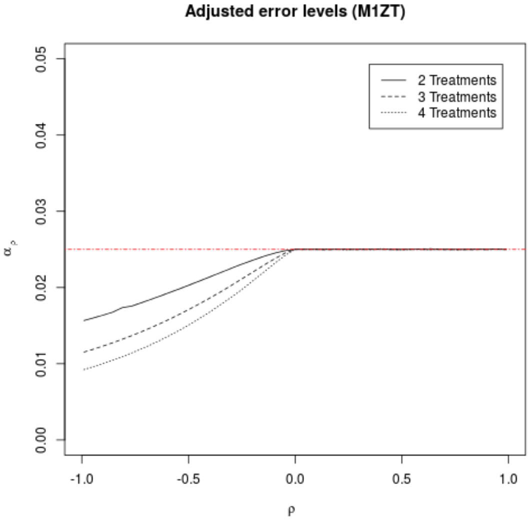
Procedures, which control the *wcFWER*: 3. ρ-Adjusted Two-Step Procedure φ_ρ, when positive relation does not hold

We start with N=5 Treatments vs. Control



- α_ρ is chosen such, that $wcFWER_{\varphi_\rho} \leq \alpha_{\text{nom}}$

The Choice of α_ρ



Adjusted Test Procedure for Unknown ρ

[Berger and Boos, 1994]

1. Select α_{CI}.
2. Let $wcFWER_{\varphi_\nu}(\alpha'_{\text{nom}}, \rho)$ denote the $wcFWER_{\varphi_\nu}$ for the naive procedure with $FWER_{\varphi_K} = \alpha'_{\text{nom}}$. For given α_{nom} find α'_{nom}, such that

$$wcFWER_{\varphi_\nu}(\alpha'_{\text{nom}}, -1) \cdot \alpha_{\text{CI}} + \alpha'_{\text{nom}} \cdot (1 - \alpha_{\text{CI}}) = \alpha_{\text{nom}}.$$

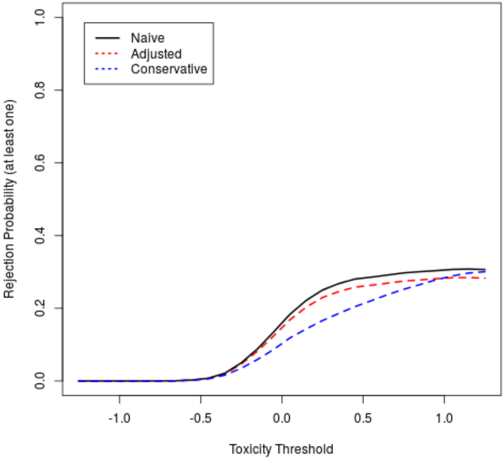
3. Calculate the left boundary $\hat{\rho}_l$ of a one-sided confidence interval for ρ and use the $\hat{\rho}_l$-adjusted test procedure controlled for $wcFWER_{\varphi_{\hat{\rho}_l}} \leq \alpha'_{\text{nom}}$.

wcFWER-Control: Proof in the line of [Tamhane et al., 2012]

Estimation of ρ

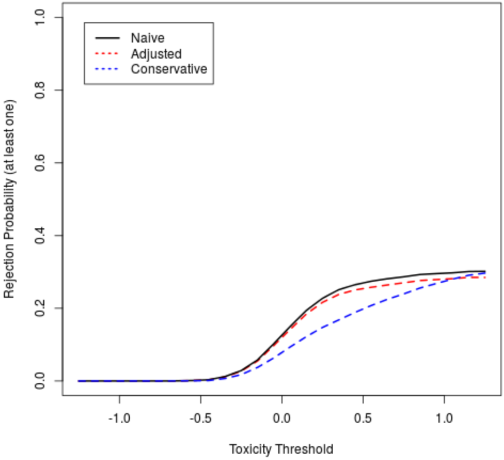
- Fisher transformation
  - arctanh applied on the sample correlation coefficient r_i is approximately normally distributed with mean $\frac{1}{2}\ln(\frac{1+\rho}{1-\rho})$ and sd $\frac{1}{\sqrt{n_i-3}}$.
- $\rho_1 = \ldots = \rho_N = \rho$ assumption $\implies$ narrower CI and higher α'_{nom} are possible.
- (approximate) independence of $\hat{\rho}_l$ and $\varphi_{\hat{\rho}_l}$?

Simulation Study ($\rho = -0.3$)



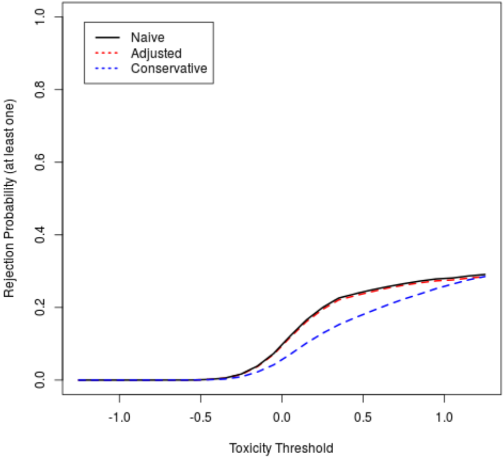
- Effects: $\theta_1 = \theta_2 = \theta_3 = 0.4$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0$)



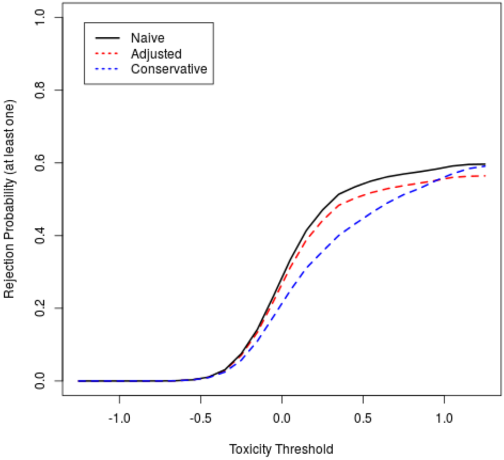
- Effects: $\theta_1 = \theta_2 = \theta_3 = 0.4$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0.3$)



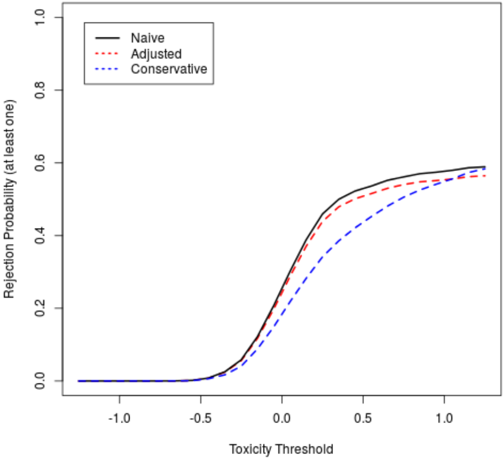
- Effects: $\theta_1 = \theta_2 = \theta_3 = 0.4$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = -0.3$)



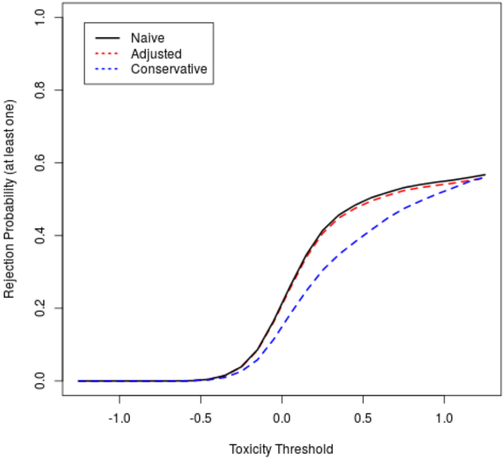
- Effects: $\theta_1 = \theta_2 = \theta_3 = 0.6$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0$)



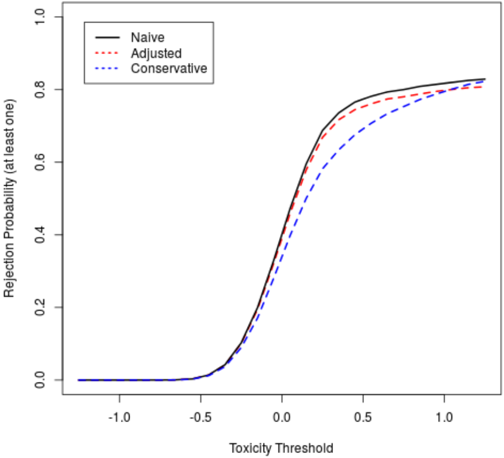
- Effects: $\theta_1 = \theta_2 = \theta_3 = 0.6$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0.3$)



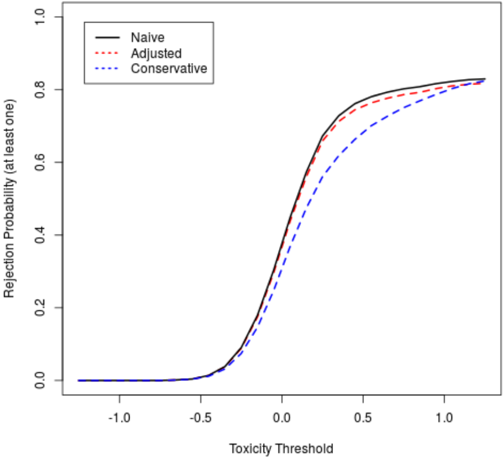
- Effects: $\theta_1 = \theta_2 = \theta_3 = 0.6$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = -0.3$)



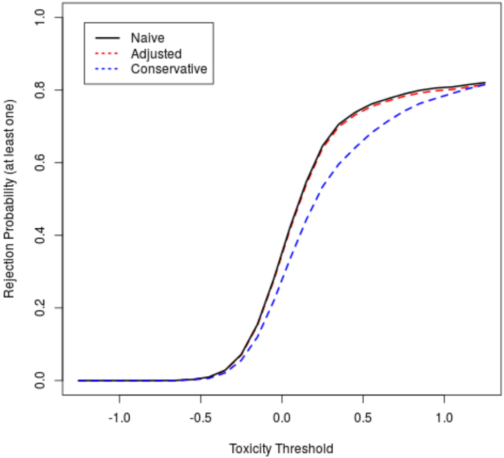
- Effects: $\theta_1 = \theta_2 = \theta_3 = 0.8$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0$)



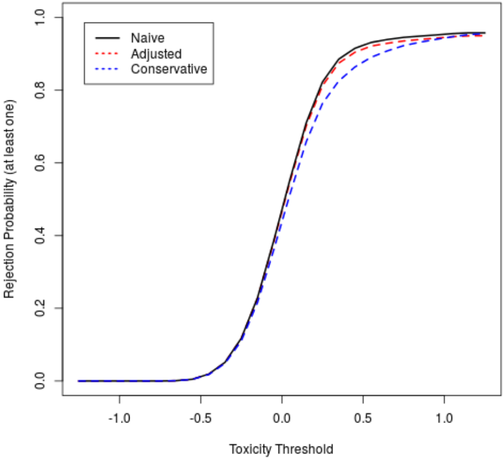
- Effects: $\theta_1 = \theta_2 = \theta_3 = 0.8$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0.3$)



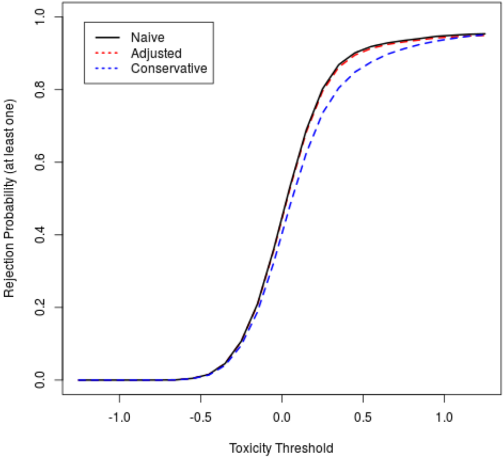
- Effects: $\theta_1 = \theta_2 = \theta_3 = 0.8$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = -0.3$)



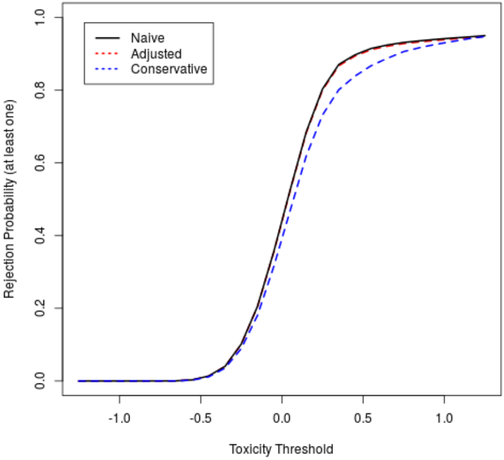
- Effects: $\theta_1 = \theta_2 = \theta_3 = 1$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0$)



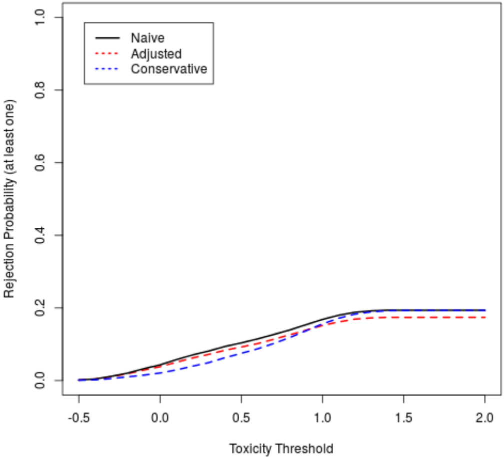
- Effects: $\theta_1 = \theta_2 = \theta_3 = 1$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0.3$)



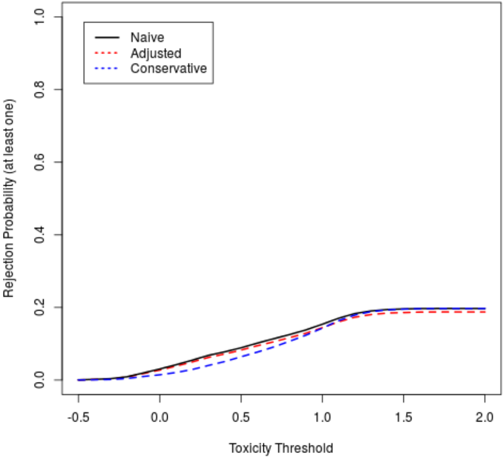
- Effects: $\theta_1 = \theta_2 = \theta_3 = 1$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = -0.3$)



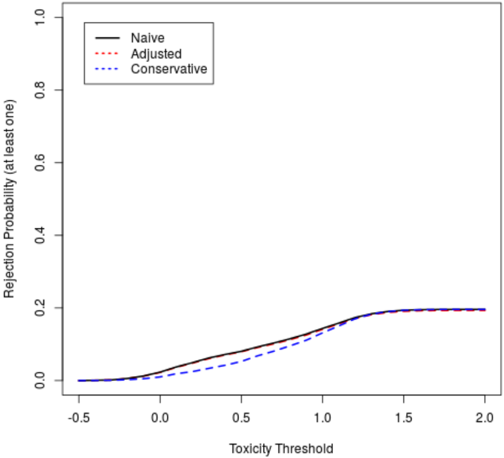
- Effects: $\theta_1 = \frac{1}{3}\theta_3,\ \theta_2 = \frac{2}{3}\theta_3, \theta_3 = 0.4$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0$)



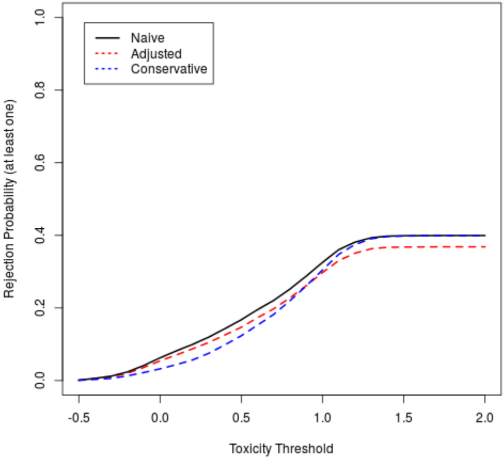
- Effects: $\theta_1 = \frac{1}{3}\theta_3,\ \theta_2 = \frac{2}{3}\theta_3, \theta_3 = 0.4$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0.3$)



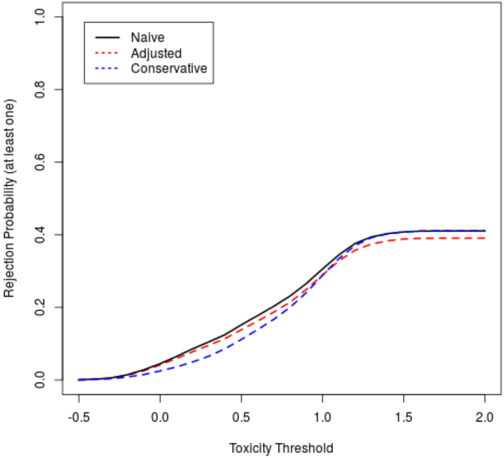
- Effects: $\theta_1 = \frac{1}{3}\theta_3,\ \theta_2 = \frac{2}{3}\theta_3, \theta_3 = 0.4$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = -0.3$)



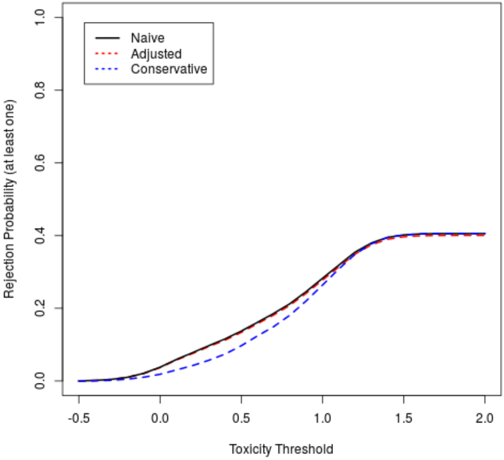
- Effects: $\theta_1 = \frac{1}{3}\theta_3,\ \theta_2 = \frac{2}{3}\theta_3, \theta_3 = 0.6$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0$)



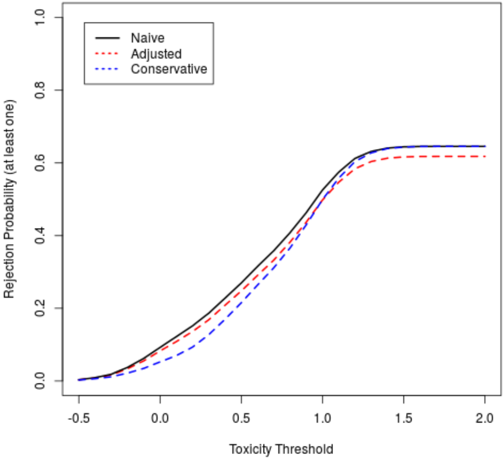
- Effects: $\theta_1 = \frac{1}{3}\theta_3,\ \theta_2 = \frac{2}{3}\theta_3, \theta_3 = 0.6$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0.3$)



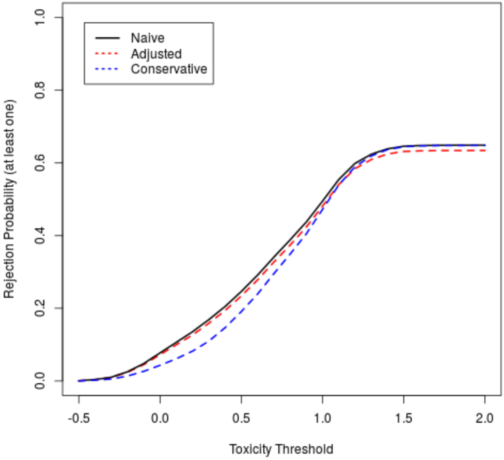
- Effects: $\theta_1 = \frac{1}{3}\theta_3,\ \theta_2 = \frac{2}{3}\theta_3, \theta_3 = 0.6$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = -0.3$)



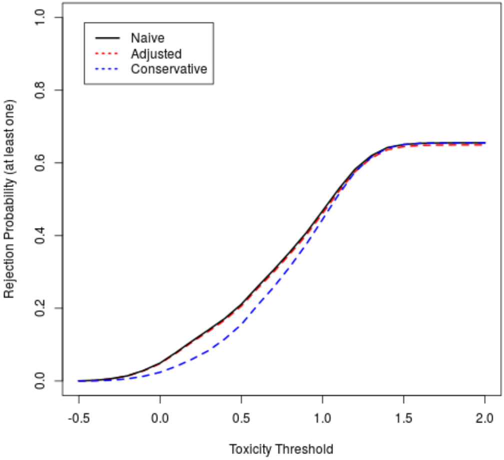
- Effects: $\theta_1 = \frac{1}{3}\theta_3,\ \theta_2 = \frac{2}{3}\theta_3, \theta_3 = 0.8$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0$)



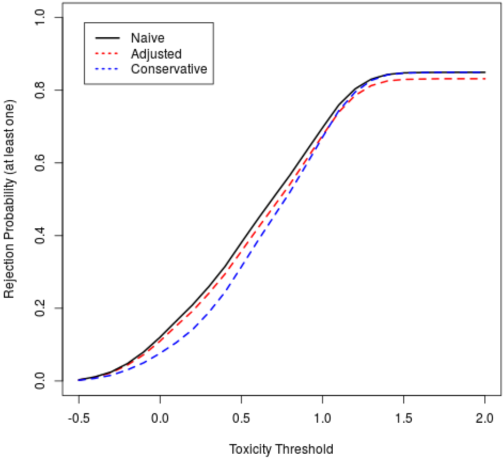
- Effects: $\theta_1 = \frac{1}{3}\theta_3,\ \theta_2 = \frac{2}{3}\theta_3, \theta_3 = 0.8$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0.3$)



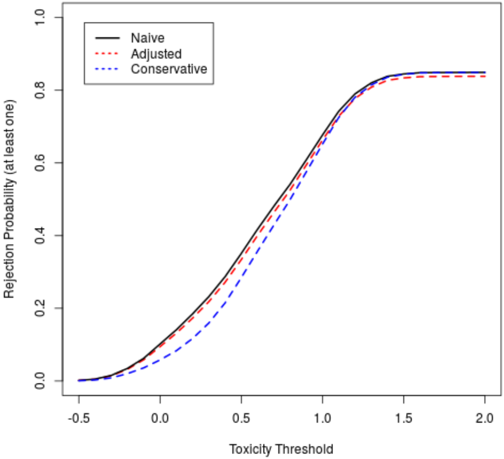
- Effects: $\theta_1 = \frac{1}{3}\theta_3,\ \theta_2 = \frac{2}{3}\theta_3, \theta_3 = 0.8$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = -0.3$)



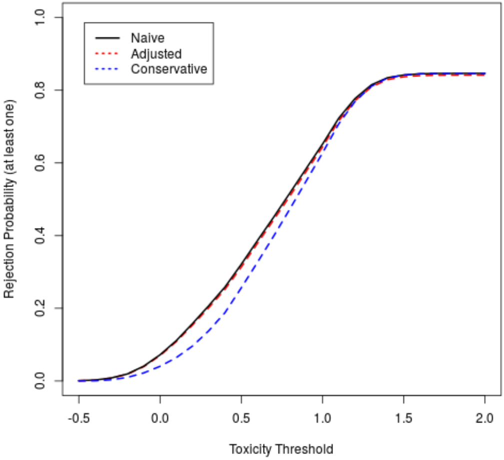
- Effects: $\theta_1 = \frac{1}{3}\theta_3,\ \theta_2 = \frac{2}{3}\theta_3, \theta_3 = 1$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0$)



- Effects: $\theta_1 = \frac{1}{3}\theta_3,\ \theta_2 = \frac{2}{3}\theta_3, \theta_3 = 1$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Simulation Study ($\rho = 0.3$)



- Effects: $\theta_1 = \frac{1}{3}\theta_3,\ \theta_2 = \frac{2}{3}\theta_3, \theta_3 = 1$
- Tox. Means: $\tau_1 = 0, \tau_2 = 0.5, \tau_3 = 1$

Conclusion

- The naive test procedure is adequate, if there is no doubt about positive relation.
- The conservative test controls the *wcFWER*, regardless of how safety selection is done.
- The ρ-adjusted procedure can be a good alternative to the conservative procedure, if it can be expected that there are safety issues and
  - ρ is known, or
  - there is at least a lower boundary for ρ.
- For unknown ρ, an approach that incorporates lower boundary estimation, may be adequate.

Berger, R. L. and Boos, D. D. (1994).
P values maximized over a confidence set for the nuisance parameter.
Journal of the American Statistical Association, 89(427):1012–1016.

Esary, J. D., Proschan, F., Walkup, D. W., et al. (1967).
Association of random variables, with applications.
The Annals of Mathematical Statistics, 38(5):1466–1474.

König, F., Bauer, P., and Brannath, W. (2006).
An adaptive hierarchical test procedure for selecting safe and efficient treatments.
Biometrical journal, 48(4):663–678.

Tamhane, A. C., Wu, Y., and Mehta, C. R. (2012).
Adaptive extensions of a two-stage group sequential procedure for testing primary and secondary endpoints (i): unknown correlation between the endpoints.
Statistics in medicine, 31(19):2027–2040.