Sample Size Computation with r-power control in the context of co-primary endpoints

Philippe Delorme[a], Pierre Lafaye de Micheaux [a], Benoît Liquet [c,d] and **Jérémie Riou** [b,c,d]

[a] *Mathematics and statistics department,University of Montréal, CANADA*
[b] *Danone Research, Clinical Studies Plateform, Centre Daniel Carasso, FRANCE*
[c] *Research Center INSERM U897, ISPED, FRANCE*
[d] *University of Bordeaux Segalen, ISPED, FRANCE*

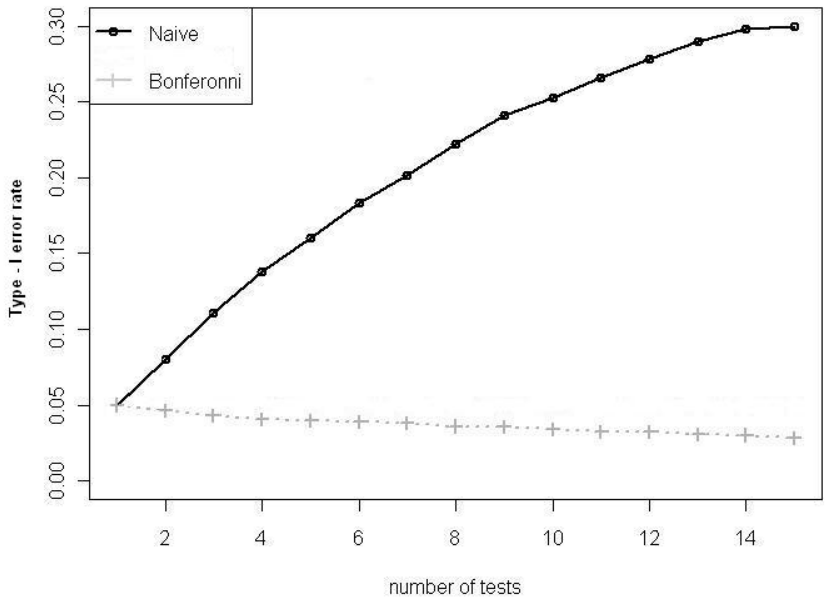
Clinical Context

- The use of **multiple endpoints** to characterize product **safety and efficacy** measures is an increasingly common feature in recent clinical trials;
- Usually, these endpoints are divided into **one** primary endpoint and several secondary endpoints;
- Nevertheless, when we observed a **multi factorial effect** it is necessary to use some **multiple primary endpoint** or a **composite endpoint**.

Multiple Testing Context

Underlying problem

Multiple Co-primary endpoints implies multiple testing problem.



Multiple Testing Context

Table : Possible scenarios for m Tests

Null Hypotheses	Not Rejected	Rejected	Total
True	U	V	m_0
False	T	S	$m - m_0$
Total	W	R	m

In confirmatory context, during data analysis statistician use Type-I FWER control:

$$Type - I\ FWER = \mathbb{P}(V \geq 1).$$

Endpoint definition

The choice of the sample size computation procedure depends on primary endpoint definition.

Primary endpoint definition

- At least one win: At least one test significant among the m;
- At least r win: At least r tests significants among the m,$(1 \leq r \leq m)$;
- All must win: All the m tests significants.

r-Power

Decision rule: At least r wins

At least r tests significant among the m $(1 \leq r \leq m)$;

In this context, we want to control the Type-II gFWER:

$$\beta_{r,m}(P) = \text{pr}(\text{make at least } p-(r-1) \text{ individual Type II errors }),$$

which is defined by 1- "r-power" [1]:

$$1-\beta_{r,m}(P) = \text{pr}(\text{reject at least } r \text{ of the } p \text{ false null hypotheses}).$$

[1] Dunnett, C.W. and Tamhane, A.C.(1992), *JASA*.

Specific aims

1. Find a **power definition** for the interest decision rule (at least r among m), and a given multiple testing procedure;
2. **Compute the Sample Size** for a given multiple testing procedure;
3. Develop an  **Package** to make the work available (rPowerSampleSize).

Reminders

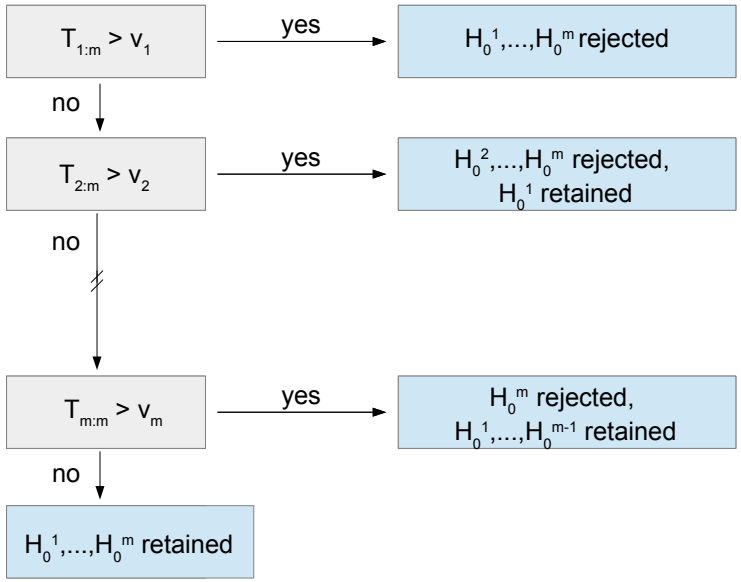
- m co-primary endpoints;
- Success of the trial is defined by: **at least** r co-primary endpoints are significant;
- r-Power control;
- **Single step** and **Stepwise methods**.

Step up methods

We **focus** in this presentation on **Step-up methods**. Nevertheless, the methodology is available for all Single step and StepWise methods.

Step-up methods Principle

1. Let the order statistics: $T_{1:m} \leq T_{2:m} \leq \ldots \leq T_{m:m}$ corresponding respectively to $\mathcal{H}_0^1, \mathcal{H}_0^2, \ldots, \mathcal{H}_0^m$;
2. Algorithm:



Step-up r-Power Formula

$$
\begin{aligned}
1-\beta^{u}_{r,m}(P) &= pr\left(\bigcup_{t=0}^{p-r}(\text{Reject exactly } p-t \text{ false null hypotheses})\right)\\
&= pr\left(\bigcup_{t=0}^{p-r}\left[\left(T_{(t+1):p} > v_{t+1}\right)\cap\left(\bigcap_{j=1}^{t}\left(T_{j:p}\leq v_j\right)\right)\right]\right)\\
&= \sum_{t=0}^{p-r} pr\left(\left(T_{(t+1):p} > v_{t+1}\right)\cap\left(\bigcap_{j=1}^{t}\left(T_{j:p}\leq v_j\right)\right)\right).
\end{aligned}
$$

where the v_j's are critical values for step-up procedures among the false null hypotheses. In the package, we use procedures which control the gFWER.

This formula depends on **order statistics**. We need to use the **Margolin and Maurer Theorem (1976)** [2] in order to obtain a power formula which depends on joint distribution of statistics.

[2] Maurer, W. and Margolin, B.H.(1976), *The Annals of Statistics*.

Power Formula without order statistics

Let $\underline{a} = (a_1, \ldots, a_q)^T \in \mathbb{N}^q$ and note $\underline{a}^* = (a_2, \ldots, a_{q+1})^T$ with $a_{q+1} = p$ and $a_0 = 0$, $\underline{a}_+ = \sum_{i=1}^q a_i$ and $\Delta a_i = a_i - a_{i-1}$, $i \in \mathcal{I}_{q+1}$. Let introduce the set

$$\mathcal{J}(\underline{a}, p) = \left\{ \underline{j} \in \mathcal{I}_p^{a_q}; j_r < j_{r+1} \text{ for } r \in \{a_{h-1}+1, \ldots, a_h - 1\}, h \in \mathcal{I}_q \text{ and } j_r \neq j_s, 1 \leq r < s \leq a_q \right\}.$$

$$1 - \beta_{r,m}^u(P) \quad \geq \quad 1 - (-1)^{(p-r+1)(p-r+2)/2} \sum_{\underline{a}=\underline{w}}^{\underline{a}^*} (-1)^{\underline{a}_+} \mathbb{P}_{\underline{\mathbf{a}}} \prod_{h=1}^{p-r+1} \begin{pmatrix} (\Delta a_h) - 1 \\ a_h - h \end{pmatrix},$$

where $\underline{w} = (1, \ldots, p-r+1)$ and $\mathbb{P}_{\underline{\mathbf{a}}} = \sum_{\underline{\mathbf{j}} \in \mathcal{J}(\underline{\mathbf{a}}, \mathbf{p})} \text{pr}\left[\cap_{\mathbf{i}=\mathbf{0}}^{\mathbf{p}-\mathbf{r}} \left\{ \cap_{\mathbf{k}=\mathbf{a_i}+\mathbf{1}}^{\mathbf{a_{i+1}}} \left(\mathbf{T_{j_k}} \leq \mathbf{v_{i+1}} \right) \right\} \right]$.

When $p = m$, namely for a weak control of the type-II r-generalized FWER, the equation of power becomes an equality.

Sample Size Computation

Step up methods

The developed formula depends only on the **joint distribution** and the **sample size**, and if the joint distribution is known, the sample size computation is possible.

So, we decided to focus on the **continuous endpoints**.

Joint distribution

Let $X^k \sim N(\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})$ with $k = \{E, C\}$,

- Unstructured Covariance matrix $\rightarrow$ Type-II multivariate non central student distribution
- Asymptotic Context: Multivariate Normal Distribution;
- $\Sigma_k = \begin{pmatrix} \sigma_k^2 & \dots & \rho\sigma_k^2 \\ \vdots & \sigma_k^2 & \vdots \\ \rho\sigma_k^2 & \dots & \sigma_k^2 \end{pmatrix}_{m \times m} \rightarrow$ Type-I multivariate non-central student distribution ;

So, in these two last contexts it is possible to compute **the required Sample Size**.

Context: ANRS 114 Pneumovac Trial

- Endpoints used in this application for the evaluation of immunogenicity in the Vaccine trials are means of log-transformed antibody concentrations for each serotype;
- Data come from ANRS 114 Pneumovac Trial, where the multivalent vaccines yields a response on 7 serotypes;
- We used data from Pedrono et al (2009);
- Covariance matrices are supposed to be the same between groups;
- The analysis will be performed using seven individual superiority Student t-statistics;
- **What is the required sample size for confirmatory trial with different decision rules (r)?**

Results

Parameters:

$d = 0_m, \pi_{r,m} = 0.8, \alpha = 0.05, \delta = (0.55, 0.34, 0.38, 0.20, 0.70, 0.38, 0.86)'$,

and $\Sigma = \begin{pmatrix} 0.124 & 0.134 & 0.137 & 0.075 & 0.140 & 0.128 & 0.161 \\ 0.134 & 0.387 & 0.287 & 0.185 & 0.316 & 0.295 & 0.396 \\ 0.137 & 0.287 & 0.294 & 0.199 & 0.274 & 0.237 & 0.342 \\ 0.075 & 0.185 & 0.199 & 0.369 & 0.192 & 0.156 & 0.238 \\ 0.140 & 0.316 & 0.274 & 0.192 & 0.394 & 0.264 & 0.397 \\ 0.128 & 0.295 & 0.237 & 0.156 & 0.264 & 0.305 & 0.335 \\ 0.161 & 0.396 & 0.342 & 0.238 & 0.397 & 0.335 & 0.651 \end{pmatrix}$

Table : Sample Size Computation for various definitions of immunogenicity:

	$r = 3$	$r = 5$	$r = 7$
Bonferroni	22	51	-
Holm	21	41	-
Hochberg	20	40	116

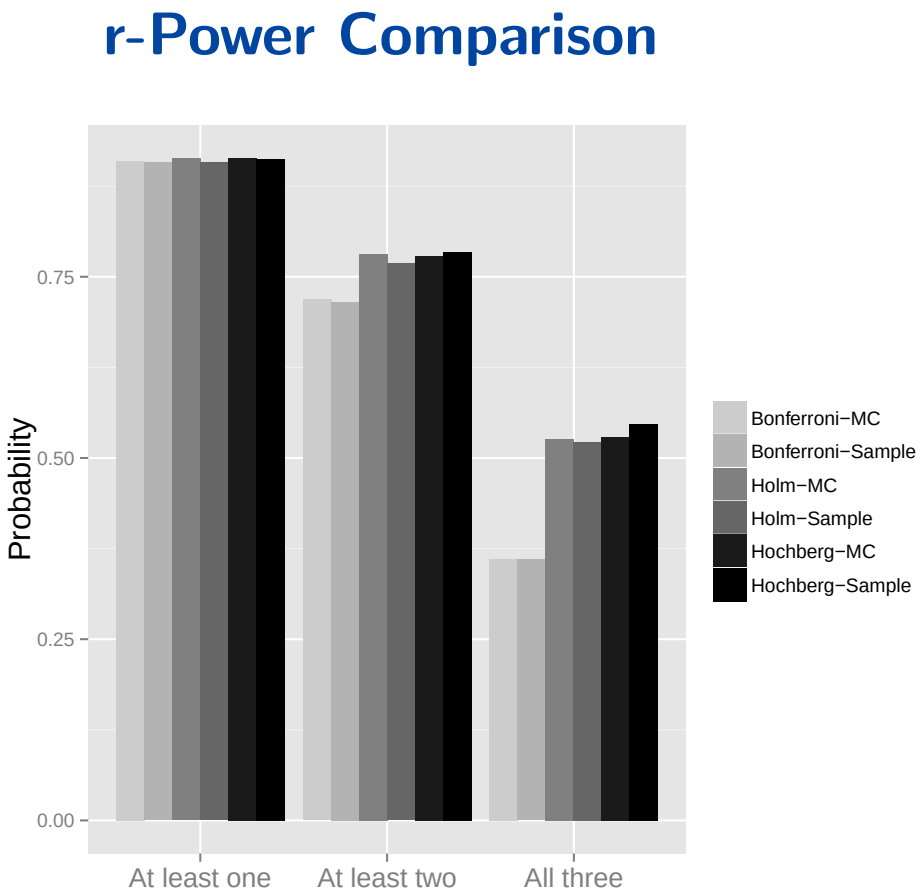
Performance (1/2)

Recently, authors have used a **Monte-Carlo simulation** in order to compute the **r-power** of a procedure in a clinical trial [3]. The aim of these slides is to compare it with our approach, in terms of **power** and **computation time**.

- New treatment against schizophrenia with a primary endpoint based on change from baseline for **three dosing groups**;
- **Continuous endpoints**, true mean changes are expected to be given by vector $\boldsymbol{\delta} = (5.0, 5.0, 3.5)^T$;
- We considered $\alpha = 0.025$, $n = 260$, the same standard deviation for each endpoint ($\sigma_k = 18$) and each group, and the same correlation between all tests ($\rho = 0.5$) for each group;
- We considered **Bonferroni, Holm and Hochberg Procedures**, and N=100,000 Monte-Carlo simulations.

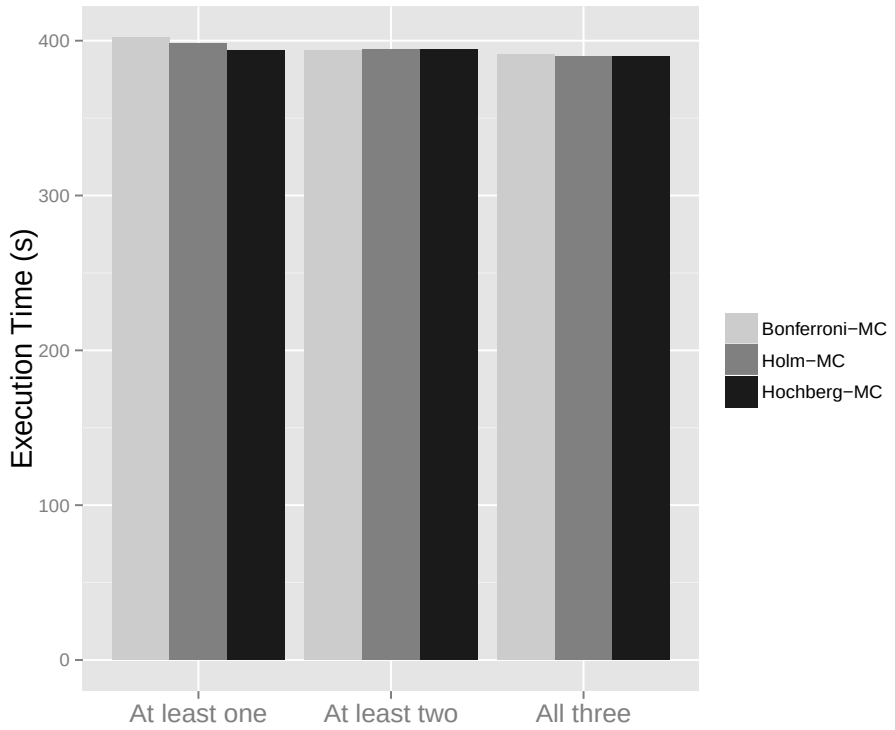
[3] Dmitrienko, A. and D'Agostino, R.(2013), *Statistics in Medicine.*

Performance (2/3) - Comparison of Monte Carlo and `rPowerSampleSize`

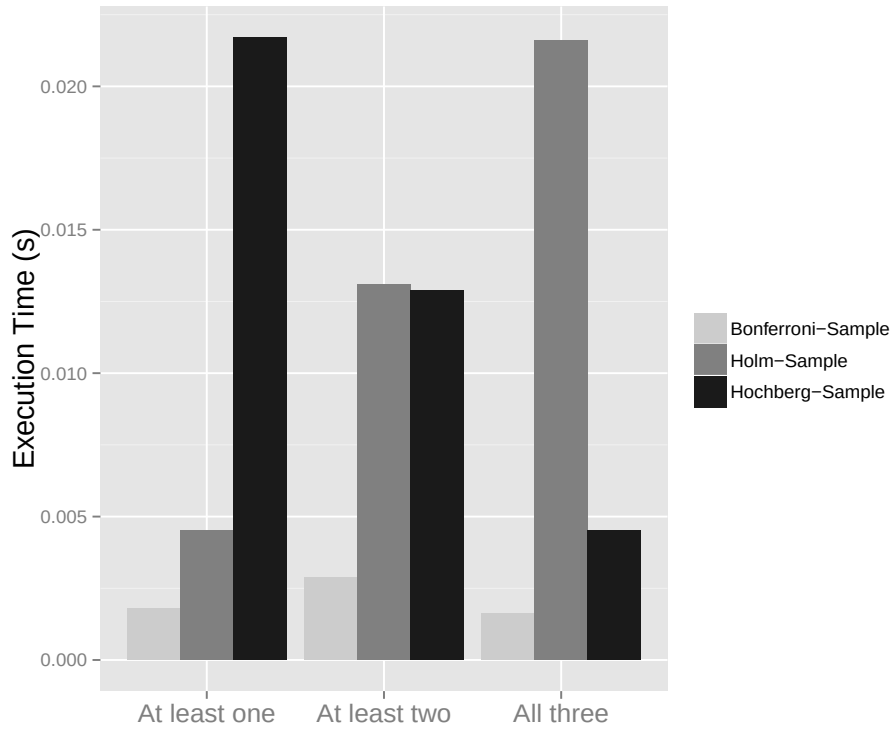


Performance (3/3) - Comparison of Monte Carlo and `rPowerSampleSize`

Computation Time (MC) (rPowerSampleSize)



Computation Time



Conclusion

- In this example, power results are similar for both procedures, and the computation time of rPowerSampleSize is 20,000 times faster than MC,
- All developed methods are **completely new** and should be **fully integrated into current clinical practice**.
- They allow many statisticians to have a methodology for **sample size computation in line with their clinical aims**, and to obtain more accurate sample sizes.
- The  **Package** rPowerSampleSize is **available soon on the CRAN**.
- This work was **submitted** for publication.