Type-II Generalized Family-Wise Error Rate Formulas with Application to Sample Size Determination

Benoit Liquet[1,2], Pierre Lafaye de Micheaux[3] and Jérémie Riou[4]

[1] University de Pau et Pays de l'Adour, LMAP.
[2] ARC Centre of Excellence for Mathematical and Statistical Frontiers
Queensland University of Technology (QUT)

[3] CREST-ENSAI (France)

[4] Université d'Angers (France)

Contents

Clinical Context

Clinical endpoint: an *event or outcome* that can be measured objectively to determine whether the intervention being studied is *beneficial*. Some examples of endpoints are survival, improvements in quality of life, relief of symptoms, and disappearance of the tumor.

Clinical Context

Clinical endpoint: an *event or outcome* that can be measured objectively to determine whether the intervention being studied is *beneficial*. Some examples of endpoints are survival, improvements in quality of life, relief of symptoms, and disappearance of the tumor.

- 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 endpoints** or a **composite endpoint**.

Industrial Statistical Challenge in Nutrition

Effects of dairy products are often Multifactorial, Smaller than pharmaceutical products, with an Higher Variability

Industrial statistical challenge

1. **Sample Size Determination** in the context of Multiple Primary Endpoints;
2. **Data Analysis** in the context of Multiple Primary Endpoints.

Multiple Primary endpoints

The choice of the sample size computation procedure depends on strategy associated to primary endpoint definition [1].

- *"At least one win"*: The trial's main objective is met if one or more individual primary objectives are achieved ;
- *"All must win"*: The trial's main objective is met if all the m individual primary objectives are achieved ;
- *"At least r wins"*: The trial's main objective is met if r or more individual primary objectives are achieved ($1 \leqslant r \leqslant m$).

[1] Dmitrienko, A. et al.(2012), *Statistics in Medicine*.

Today Aims

1. *Brief description on* Sample Size Computation and Data Analysis in the context of "At least one win" primary continuous endpoints;

   Lafaye de Micheaux P., Liquet B., Marques S. and Riou J., Power and sample size determination in clinical trials with multiple primary continuous correlated end points. *Journal of Biopharmaceutical Statistics* 24:2, 378-97, (2014).

2. *More Details* on Sample Size Computation Methodology in the context of "'At least *r* wins" primary endpoints.

   Delorme P., Lafaye de Micheaux P., Liquet B. and Riou J., Type-II Generalized Family-Wise Error Rate Formulas with Application to Sample Size Determination. *Statistics in Medicine* (2016) In press.

Data

Indiv	Group	Primary Endpoints				
		1	…	j	…	m
1	0	X^0_{11}	…	X^0_{1j}	…	X^0_{1m}
⋮	⋮	⋮	⋱	⋮	⋱	⋮
i	0	X^0_{i1}	…	X^0_{ij}	…	X^0_{im}
⋮	⋮	⋮	⋱	⋮	⋱	⋮
n	0	X^0_{n1}	…	X^0_{nj}	…	X^0_{nm}
n+1	1	$X^1_{(n+1)\ 1}$	…	$X^1_{(n+1)j}$	…	$X^1_{(n+1)\ m}$
⋮	⋮	⋮	⋱	⋮	⋱	⋮
i	1	X^1_{i1}	…	X^1_{ij}	…	X^1_{im}
⋮	⋮	⋮	⋱	⋮	⋱	⋮
2n	1	$X^1_{(2n)1}$	…	$X^1_{(2n)j}$	…	$X^1_{(2n)m}$

Sample size for one endpoint: single hypothesis testing

		True state of Nature	
		$\mathcal{H}_0$ is true	$\mathcal{H}_1$ is true
Decision	We decide $\mathcal{H}_1$	Type I error	No error
	We decide $\mathcal{H}_0$	No error	Type II error

Sample size for one endpoint: single hypothesis testing

		True state of Nature	
		$\mathcal{H}_0$ is true	$\mathcal{H}_1$ is true
Decision	We decide $\mathcal{H}_1$	Type I error	No error
	We decide $\mathcal{H}_0$	No error	Type II error

The *Type I error* is when one decides $\mathcal{H}_1$ while it is $\mathcal{H}_0$ that is true.
The *Type II error* is when one decides $\mathcal{H}_0$ while it is $\mathcal{H}_1$ that is true.

$$\begin{aligned} \text{power function} &= P\left[\text{not decide } \mathcal{H}_0 \text{ when } \mathcal{H}_1 \text{ is true}\right] \\ &\equiv 1 - \beta. \end{aligned}$$

Sample size for one endpoint: single hypothesis testing

		True state of Nature	
		$\mathcal{H}_0$ is true	$\mathcal{H}_1$ is true
Decision	We decide $\mathcal{H}_1$	Type I error	No error
	We decide $\mathcal{H}_0$	No error	Type II error

The *Type I error* is when one decides $\mathcal{H}_1$ while it is $\mathcal{H}_0$ that is true.
The *Type II error* is when one decides $\mathcal{H}_0$ while it is $\mathcal{H}_1$ that is true.

$$\begin{aligned} \text{power function} &= P\left[\text{not decide } \mathcal{H}_0 \text{ when } \mathcal{H}_1 \text{ is true}\right] \\ &\equiv 1 - \beta. \end{aligned}$$

Generally easy to determine the necessary sample size n to use in order to control (with some given thresholds) both the maximal Type I error rate (under $\mathcal{H}_0$) and a Type II error rate (under $\mathcal{H}_1$).

Sample size for multiple primary endpoints ?

We want to evaluate the m following hypotheses:

$$\mathcal{H}_0^1 : \mu_1^E - \mu_1^C \leqslant d_1 \text{ versus } \mathcal{H}_1^1 : \mu_1^E - \mu_1^C > d_1$$

$$\mathcal{H}_0^2 : \mu_1^E - \mu_2^C \leqslant d_2 \text{ versus } \mathcal{H}_1^2 : \mu_2^E - \mu_2^C > d_2$$

$$\vdots$$

$$\mathcal{H}_0^m : \mu_m^E - \mu_m^C \leqslant d_m \text{ versus } \mathcal{H}_1^m : \mu_m^E - \mu_m^C > d_m$$

Each one of these *elementary hypotheses* will be tested using an associated test statistic. We thus have m test statistics $T_1, \ldots, T_m$.

Multiple hypothesis testing, also called *multiple comparisons* or *multiple testing*, refers to the simultaneous testing of **more than one** individual hypothesis at a time.

Family of hypotheses

We have defined a *family of hypotheses* $\mathcal{H}_1, \ldots, \mathcal{H}_m$. We have m (individual) Type I errors, one for each of the individual hypotheses.

We now want to define some kind of **unique** *overall Type I error rate* for the **whole family**.

Note that, for a given family of hypotheses, an overall Type I error rate depends on which ones are assumed to be true and which ones are assumed false.

A (global) Type I error rate can thus be controlled in (at least) two ways:

- **Weak:** The overall Type I error rate $\leqslant \alpha$ when all null hypotheses are supposed to be true.
- **Strong:** All overall Type I error rates $\leqslant \alpha$, for any (sensible) given configuration of false and true null hypotheses.

FamilyWise Error Rate

The most widely used overall Type I error rate is probably the *Family Wise Error Rate* (FWER) defined as

$$FWER = P(\text{commit at least one Type I error}).$$

FamilyWise Error Rate

The most widely used overall Type I error rate is probably the *Family Wise Error Rate* (FWER) defined as

$$FWER = P(\text{commit at least one Type I error}).$$

Possible scenarii for m tests

		Decision		
	Null Hypotheses	Not Rejected	Rejected	Total
True state	True	U	V	m_0
	False	T	S	p
	Total	W	R	m

$$\text{Type-I FWER} = P(V \geqslant 1).$$

$$\text{Type-I q-gFWER} = P(V \geqslant q).$$

Power control

Possible scenarii for m tests

		Decision		
	Null Hypotheses	Not Rejected	Rejected	Total
True state	True	U	V	m_0
	False	T	S	p
	Total	W	R	m

$$\begin{aligned} \text{Disjunctive Power} &= P(S \geqslant 1), \\ \text{r-Power} &= P(S \geqslant r),\ 1 \leqslant r \leqslant p, \\ \text{Conjunctive Power} &= P(S = p). \end{aligned}$$

At least one win: Individual testing approach

- Let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)^T$, with $\delta_j = \mu_j^E - \mu_j^C$, $(1 \leqslant j \leqslant m)$, be the vector of the true differences between the test(E) and the control(C) products;
- **Individual Hypotheses:**

$$\mathcal{H}_0^j : \delta_j = 0 \textit{ versus } \mathcal{H}_1^j : \delta_j \neq 0;$$

- **Global Hypothesis:**

$$\mathcal{H}_0 = \cap_{j=1}^m \mathcal{H}_0^j \textit{ versus } \mathcal{H}_1 = \cup_{j=1}^m \mathcal{H}_1^j.$$

Statistics

- When σ_j^2 are **known**, the standardized test statistic is:

$$Z_j^{(n)} = \frac{\bar{X}_j^E - \bar{X}_j^C}{\sqrt{\frac{2}{n}}\sigma_j}, \text{ where } \bar{X}_j^k = \frac{1}{n}\sum_{i=1}^{n} X_{i,j}^k \text{ are the sample means for group k;}$$

- When σ_j^2 are **unknown**, they are estimated by the pooled variances:

$$T_j^{(n)} = \frac{\bar{X}_j^E - \bar{X}_j^C}{\sqrt{\frac{2}{n}}\widehat{\sigma}_j}, \text{ where } \widehat{\sigma}_j^2 = \frac{1}{2n-2}\sum_{i=1}^{n}\left[(X_{i,j}^E - \bar{X}_j^E)^2 + (X_{i,j}^C - \bar{X}_j^C)^2\right].$$

Simultaneous Control

- **FamilyWise Error Rate:**

$$\begin{aligned} FWER &= \text{pr}(\text{Reject at least one } \mathcal{H}_0^j, 1 \leqslant j \leqslant m | \mathcal{H}_0 \text{ is true}), \\ &= 1 - \text{pr}\left\{(|Z_1^n| \leqslant c_\alpha) \cap \ldots \cap (|Z_m^n| \leqslant c_\alpha) \,|\mathcal{H}_0 \text{ is true}\right\}, \end{aligned}$$

where c_α is chosen to satisfy $FWER = \alpha$, for a fixed significance level α.

- **Disjunctive Power:**

$$\begin{aligned} 1 - \beta &= \text{pr}(\text{Reject at least one } \mathcal{H}_0^j, 1 \leqslant j \leqslant m | \mathcal{H}_1 \text{ is true}), \\ &= 1 - \text{pr}\left\{(|Z_1^n| \leqslant c_\alpha) \cap \ldots \cap (|Z_m^n| \leqslant c_\alpha) \,|\mathcal{H}_1 \text{ is true}\right\}, \end{aligned}$$

Distribution

- **Normality assumption and known covariance matrix:**

$$\mathbf{Z}_n \overset{\mathcal{H}^0}{\sim} \mathcal{N}_m\left(\boldsymbol{0}_m, R\right) \quad \text{and} \quad \mathbf{Z}_n \overset{\mathcal{H}^1}{\sim} \mathcal{N}_m\left(\sqrt{\frac{n}{2}} P\boldsymbol{\delta}^*, R\right),$$

where $\boldsymbol{\delta}^* \neq \boldsymbol{0}_m$ is the value of $\boldsymbol{\delta}$ under $\mathcal{H}^1$ and where $R = P\Sigma P$ is the $m \times m$ correlation matrix associated with Σ, with P the diagonal matrix whose j^{th} element is $1/\sigma_j$.

- **Asymptotic Context**:

$$\widehat{R}^{-1/2}\mathbf{T}_n \xrightarrow{L} \mathcal{N}_m\left(\boldsymbol{0}_m, I_m\right), \text{ under } \mathcal{H}^0,$$

$$\widehat{R}^{-1/2}\left(\mathbf{T}_n - \sqrt{n}\widehat{V}\boldsymbol{\delta}^*\right) \xrightarrow{L} \mathcal{N}_m\left(\boldsymbol{0}_m, I_m\right), \text{ under } \mathcal{H}^1 : \boldsymbol{\delta} = \boldsymbol{\delta}^* \neq \boldsymbol{0}_m,$$

where $\widehat{R} = \widehat{V}\,\widehat{\Sigma}\,\widehat{V}$ is a consistent estimator of R, the correlation matrix of $\mathbf{T}_n = \sqrt{n}\widehat{V}(\bar{\mathbf{X}}^E - \bar{\mathbf{X}}^C)$, $\widehat{V} = \text{diag}\left(1/\sqrt{\widehat{\sigma}^2_{j,E} + \widehat{\sigma}^2_{j,C}}\right)$ and $\widehat{\Sigma} = \widehat{\Sigma}^C + \widehat{\Sigma}^E$.

Application (1/2)

- **Objective**: Demonstrate the efficacy of the consumption of a dairy product on seric antibody titres for three strains of Influenza virus;
- The product will be considered as effective if at least one out of the three strains is statistically significant.
- Two pilot studies were planned to define the product effects and variability. Both were multicentric double blind randomized controlled trials conducted in France among elderly volunteers during the two vaccination seasons 2005 and 2006;
- The mean differences between the two groups are: $\widehat{\delta} = (0.35, 0.28, 0.46)^T$;
- The covariance matrix is: $\widehat{\Sigma} = \begin{pmatrix} 5.58 & 2.00 & 1.24 \\ 2.00 & 4.29 & 1.59 \\ 1.24 & 1.59 & 4.09 \end{pmatrix}$;
- Desired Disjuntive Power: 0.80 , and desired Type-I error rate: 0.05.
- **What is the required sample size ?**

Application (2/2)

Table 1: Sample size computation with Global method and Individual Procedure

Method	Type-I error	Sample size (n)
Global	0.05	359
Indiv	0.0178	336

Global: Global method based on multivariate model;

Indiv: Individual procedure for known covariance matrix.

At least r wins

Suppose we plan to collect some data from a true model.

Let us suppose a model P to be the true model for which p null hypotheses are false and $m - p$ are true.

For some $r \leqslant p$, our global type-II r-generalized family-wise error rate is:

$$\beta_{r,m}(P) = P(\text{make at least } p - r + 1 \text{ individual type-II errors among the } p \text{ false hypotheses}),$$

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

called *generalized disjunctive power* by Dmitrienko *et al.* (2015).

Motivation: Clinical trial in vaccination

ANRS 114 Pneumovac trial: measure the effect of two vaccine strategies against *Streptococcus pneumoniae* in adults infected by the HIV, which are more susceptible to infections caused by this bacterial pathogen.

Seven ($m = 7$) clinical endpoints: log-transformed (towards Gaussianity) measurements of serotype-specific antibody titer concentrations (continuous measurements in μg/ml).

Note: serotype refers to distinct variations within a species of bacteria or viruses or among immune cells of different individuals.

Motivation: Clinical trial in vaccination

ANRS 114 Pneumovac trial: measure the effect of two vaccine strategies against *Streptococcus pneumoniae* in adults infected by the HIV, which are more susceptible to infections caused by this bacterial pathogen.

Seven ($m = 7$) clinical endpoints: log-transformed (towards Gaussianity) measurements of serotype-specific antibody titer concentrations (continuous measurements in μg/ml).

Note: serotype refers to distinct variations within a species of bacteria or viruses or among immune cells of different individuals.

Pedrono *et al.* (2009) suggest that one vaccine strategy might be considered as superior to the other when at least 3, 5 or 7 serotypes are found significant.

Aim: compute the sample sizes necessary for a weak control of the r-power for $r = 3, 5, 7$ for different multiple procedure.

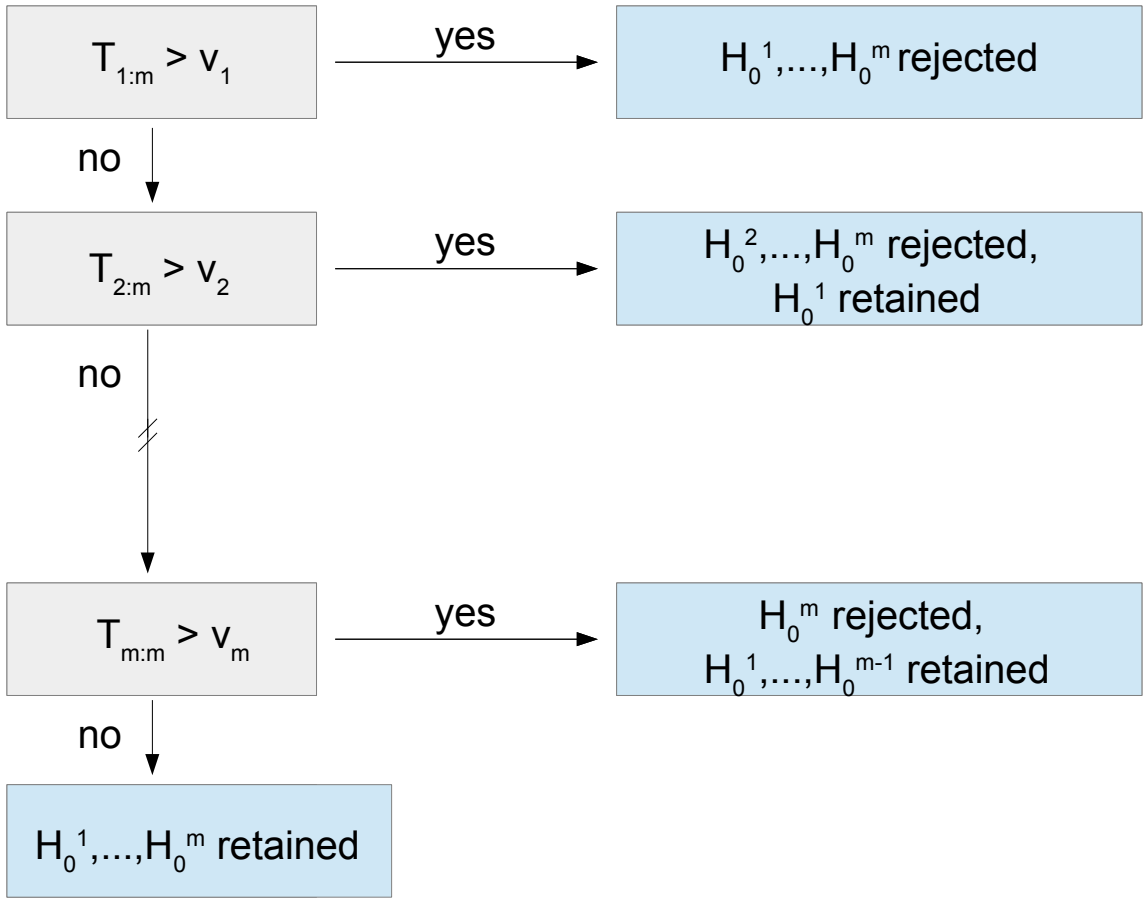
Multiple testing procedures

Many multiple testing procedures have been developped to control the FWER. They are usually categorized as *single-step* or *step-wise*:

- **One-step (or single-step)**: all p-values are compared to a pre-determined cut-off, usually only a function of α and m Equivalently, all test statistics T_k are compared to a common predetermined cut-off value c_{km}.
- **Step-down** (*e.g.* Holm);
- **Step-up** (*e.g.* Hochberg).

We note $p_{1:m} \leqslant \cdots \leqslant p_{m:m}$ the ordered p-values, and we note $\mathcal{H}_{0:1}, \ldots, \mathcal{H}_{0:m}$ the *ordered hypotheses* corresponding to the order statistics $T_{1:m} \leqslant \cdots \leqslant T_{m:m}$.

Step-up procedure



Hochberg procedure

The Hochberg's algorithm proceeds as follows:

- *Step 1:* If $p_{m:m} < \alpha$ or $T_{1:m} > u_1 = c_{1-\alpha}$, reject $\mathcal{H}_{0:i}, i = 1, \ldots, m$ and stop; otherwise go to *Step 2*.
- *Step 2:* If $p_{(m-1):m} < \alpha/2$ or $T_{2:m} > u_2 = c_{1-\alpha/2}$, reject $\mathcal{H}_{0:i}, i = 2, \ldots, m$ and stop; otherwise go to *Step 3*.
- ...
- *Step m:* If $p_{1:m} < \alpha/m$ or $T_{m:m} > u_m = c_{1-\alpha/m}$, reject $\mathcal{H}_{0:m}$ and stop.

Control of the *q*-generalized-FWER

- Bonferroni's single-step approach. Lehmann and Romano (2005) states that a simple modification of the usual Bonferroni's procedure:
  comparing marginal *p*-values to $q\alpha/m$ instead of α/m leads to a control of the *q*-generalized family-wise error rate.
- Modified Hochberg's step-up approach. Romano and Shaikh (2006) proposed a modification of the usual Hochberg's procedure which leads to a control of the *q*-generalized family-wise error rate for any structure of dependence of the *p*-values.

Derivation of the r-Power: Step-up setting

For simplication we consider all null hypotheses are false: $p = m$.

The "r-Power" or multiple must-win power is:

$$\begin{aligned}\Pi_{r,m} &= P(\text{reject at least } r \text{ false null hypotheses among } m) \\ &= \sum_{j=0}^{m-r} P\left(\text{reject exactly } m-j \text{ false hypotheses among } m\right).\end{aligned}$$

For *Step-Up* methods, we have:

$$\begin{aligned}\{\text{reject exactly } m-j \text{ hypotheses}\} &= \\ \left\{\text{reject } \mathcal{H}_{0:(j+1)}, \ldots, \mathcal{H}_{0:m}\right\} \cap \left\{\text{not reject } \mathcal{H}_{0:1}, \ldots, \mathcal{H}_{0:j}\right\} &= \\ \left\{\mathsf{T}_{(j+1):m} > u_{j+1}\right\} \cap \bigcap_{k=1}^{j} (\mathsf{T}_{k:m} \leqslant u_k).\end{aligned}$$

Derivation of the r-Power: Step-up setting

The r-Power can be written as:

$$\begin{aligned}
\Pi^{u}_{r,m} &= \sum_{j=0}^{m-r} P\left[\left(\bigcap_{k=1}^{j}(\mathrm{T}_{k:m} \leqslant u_k)\right) \cap (\mathrm{T}_{(j+1):m} > u_{j+1}) \,\middle|\, \cap_{j=1}^{m} \mathcal{H}_1^j\right] \\
&= \sum_{j=0}^{m-r}\left(P\left[\bigcap_{k=1}^{j}(\mathrm{T}_{k:m} \leqslant u_k)\right] - P\left[\bigcap_{k=1}^{j+1}(\mathrm{T}_{k:m} \leqslant u_k)\right]\right) \\
&= 1 - P\left[\bigcap_{k=1}^{m-r+1}(\mathrm{T}_{k:m} \leqslant u_k)\right] = 1 - \text{“a Type II gFWER”}.
\end{aligned}$$

The objective is now to obtain a computable expression, namely one **not involving order statistics**.

For this purpose, we will need some theorems giving the joint CDF of order statistics.

Theorem of Maurer and Margolin (1976):

Let $\underline{\ell} = (\ell_1, \ldots, \ell_q)$ such that $1 \leqslant \ell_1 \leqslant \ldots \leqslant \ell_q \leqslant m$ and $u_{\ell_1} \leqslant \ldots \leqslant u_{\ell_q}$. We obtain the **joint distribution of order statistics**:

$$P\left[\bigcap_{h=1}^{q}(\mathsf{T}_{\ell_h:m} \leqslant u_{\ell_h})\right] = (-1)^{\ell_+} \sum_{\underline{a}=\underline{\ell}}^{\underline{a}^*}(-1)^{a_+} P_{\underline{a}} \prod_{i=1}^{q}\binom{(\Delta a_i) - 1}{a_i - \ell_i}$$

with $\ell_+ = \sum_{h=1}^{q} \ell_h$, $\Delta a_i = a_i - a_{i-1}$ and

$$P_{\underline{a}} = \sum_{\underline{j} \in \mathscr{J}(\underline{a},m)} P\left[\bigcap_{i=0}^{q-1}\left(\bigcap_{k=a_i+1}^{a_{i+1}} \mathsf{T}_{j_k} \leqslant u_{\ell_{i+1}}\right)\right].$$

$\Rightarrow$ We can now replace **ordered** statistics with **unordered** ones!

Sample Size Computation

Our 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.

We considered at this stage only **continuous endpoints**. This is done using the following test statistics:

$$T_k = \left(\widehat{\mathrm{Var}}\left(\bar{X}_k^E - \bar{X}_k^C - d_k\right)\right)^{-1/2} (\bar{X}_k^E - \bar{X}_k^C - d_k),$$

where $\bar{X}_k^g = n_g^{-1} \sum_{i=1}^{n_g} X_{i,k}^g$.

Sample Size Computation

Our 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.

We considered at this stage only **continuous endpoints**. This is done using the following test statistics:

$$T_k = \left(\widehat{\mathrm{Var}}\left(\bar{X}_k^E - \bar{X}_k^C - d_k\right)\right)^{-1/2} (\bar{X}_k^E - \bar{X}_k^C - d_k),$$

where $\bar{X}_k^g = n_g^{-1} \sum_{i=1}^{n_g} X_{i,k}^g$.

Different estimators of the variance of the difference between the means have been implemented in our R package (function `indiv.analysis()`) depending on the structure of Σ^g.

Joint Distribution of Test Statistics for Continuous Multiple Endpoints

We investigate the case of a multivariate Gaussian distribution

$$\left(\boldsymbol{X}_1^g, \ldots, \boldsymbol{X}_{n_g}^g\right)^\top \sim \mathcal{N}_m^{n_g}\left((\boldsymbol{\mu}^g, \ldots, \boldsymbol{\mu}^g)^\top, \mathcal{I}_{n_g} \otimes \Sigma^g\right),$$

Various classical scenarios on the structure of the covariance matrices Σ^g:

- Unstructured covariance matrix
  - When $\Sigma^E = \Sigma^C$
  - When $\Sigma^E \neq \Sigma^C$
- Multisample compound symmetry covariance matrix:

$$K_\varrho = (1-\varrho) I_m + \varrho J \text{ with } J = \begin{pmatrix} 1 & \ldots & 1 \\ \vdots & 1 & \vdots \\ 1 & \ldots & 1 \end{pmatrix}.$$

  - $\Sigma^g = \sigma^{2,g} K_\varrho$
  - $\Sigma^E = \Sigma^C = \sigma^2 K_\varrho$

Joint Distribution of Test Statistics for Continuous Multiple Endpoints

- Unstructured covariance matrix

  - When $\Sigma^E = \Sigma^C$, we get a multivariate type-II Student distribution.
  - When $\Sigma^E \neq \Sigma^C$, we get a a non-asymptotic approximation to a multivariate type-II Student distribution.
  - Asymptotic distribution of $\boldsymbol{T} = (T_1, \ldots, T_m)$ to a multivariate Gaussian distribution.

Joint Distribution of Test Statistics for Continuous Multiple Endpoints

- Unstructured covariance matrix

  - When $\Sigma^E = \Sigma^C$, we get a multivariate type-II Student distribution.
  - When $\Sigma^E \neq \Sigma^C$, we get a a non-asymptotic approximation to a multivariate type-II Student distribution.
  - Asymptotic distribution of $\boldsymbol{T} = (T_1, \ldots, T_m)$ to a multivariate Gaussian distribution.

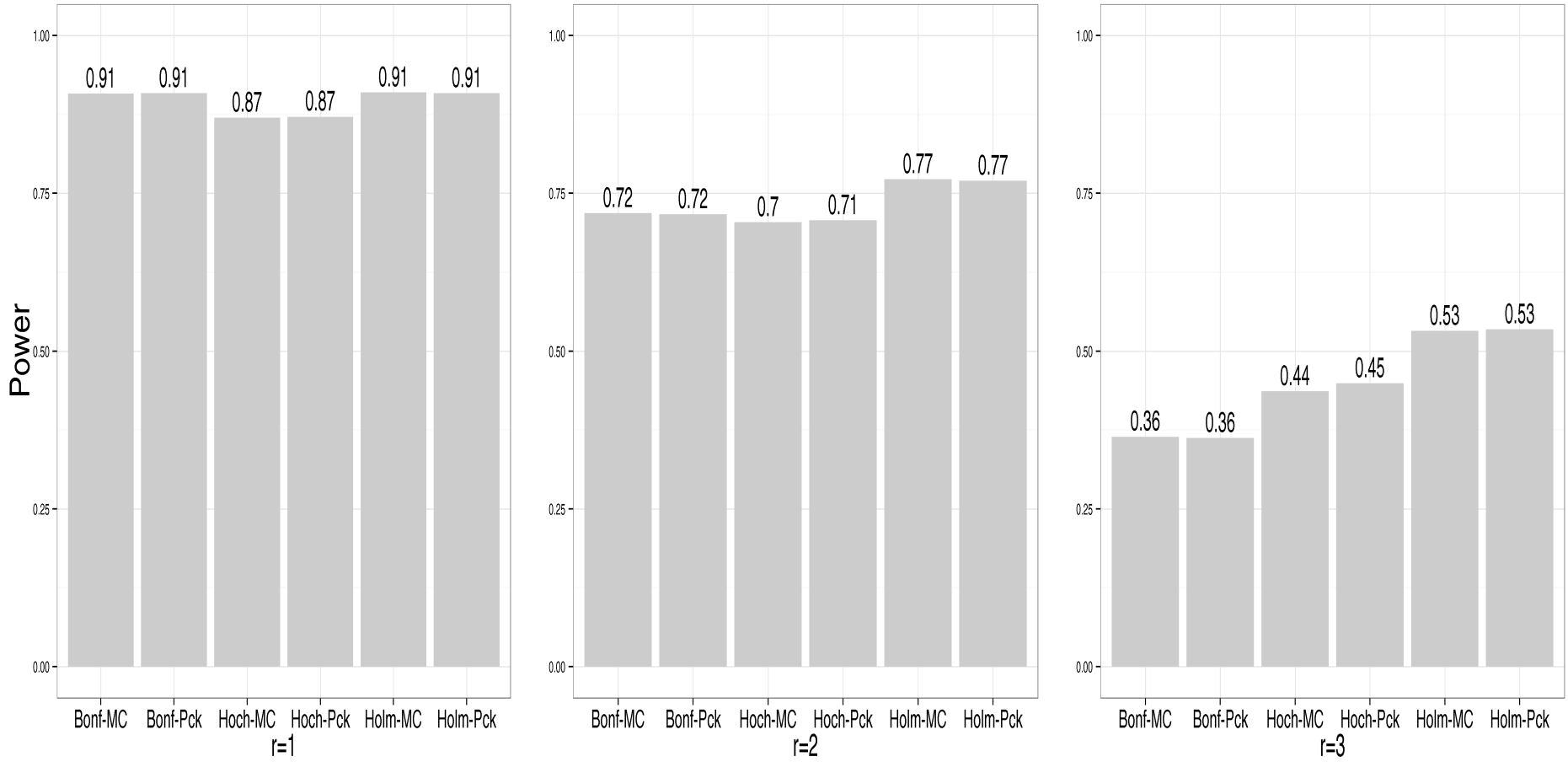
- Multisample compound symmetry covariance matrix:

  - $\Sigma^g = \sigma^{2,g} K_\varrho$, we get $\boldsymbol{T} \overset{approx}{\sim}$ Kshirsagar distribution
  - $\Sigma^E = \Sigma^C = \sigma^2 K_\varrho$, we get a Kshirsagar distribution

Simulation Study

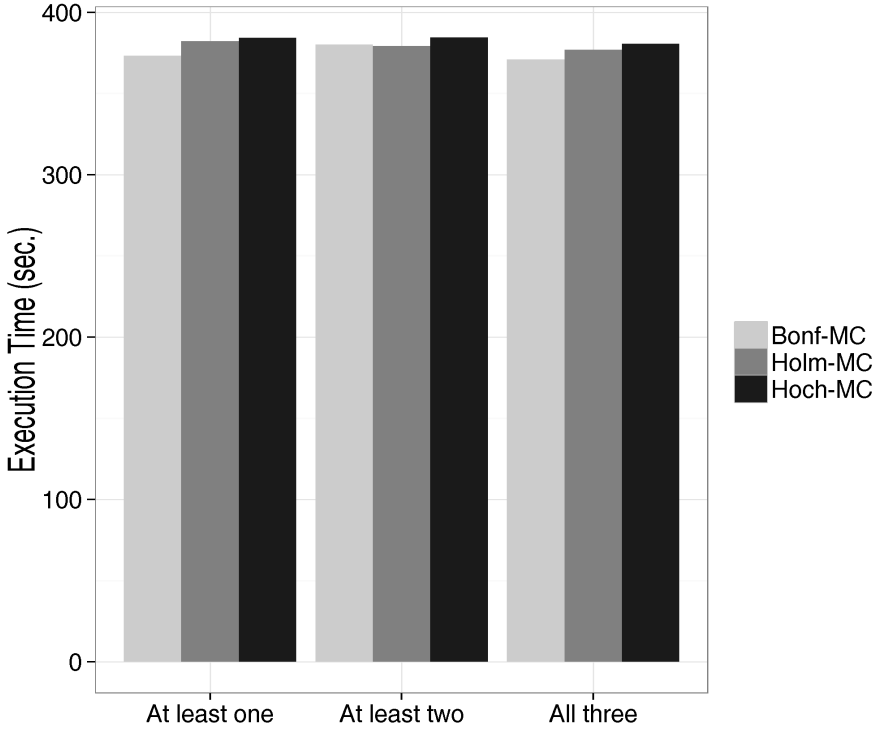
Recently, authors have used a **Monte-Carlo simulation** in order to compute the **r-power** of a procedure in a clinical trial.

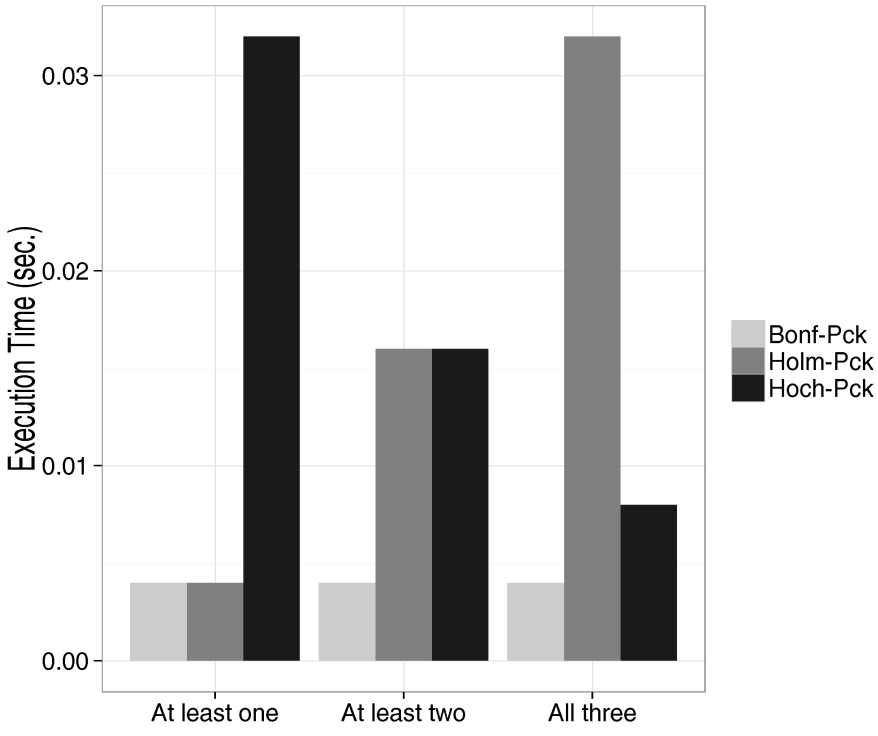
- New treatment against schizophrenia with a primary endpoint based on change from baseline for **three dosing groups**;
- **Continuous endpoints**, true mean changes are expected to be given by vector $\boldsymbol{\delta} = (5.0, 5.0, 3.5)^T$;
- We considered $\alpha = 0.025$, $n = 260$, the same standard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



As suggested by Dmitrienko *et al.* (2013), "the information presented in the central panel may be used to improve the sponsor's ability to characterize the dose-response relationship. If the sponsor was interested in identifying two or three doses with a desirable efficacy profile, the sample size could be adjusted to achieve a higher value for the probability to detect at least two significant doses."

Computation time



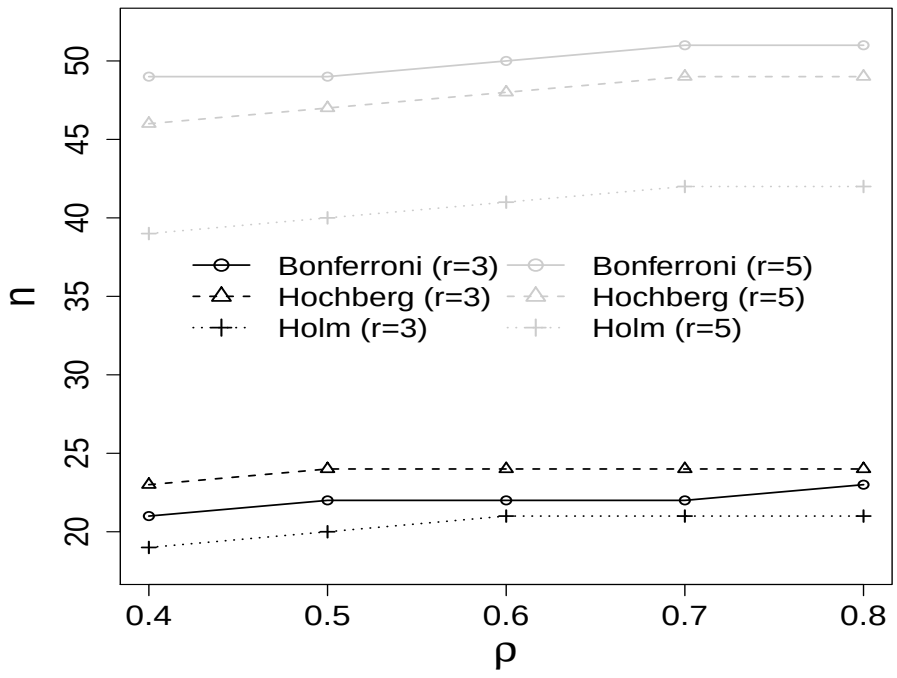


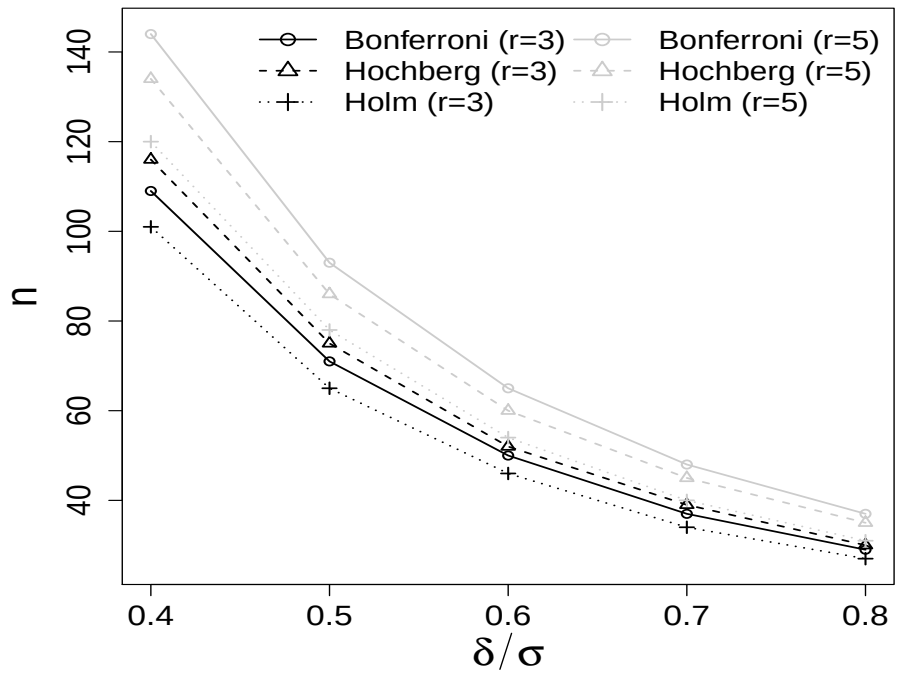
Application to he Pneumovac trial

- Endpoints used for the evaluation of immunogenicity in the Vaccine trials are means of antibody concentrations for each serotype;
- Data comes from ANRS 114 Pneumovac Trial, where the multivalent vaccine yields a response on 7 serotypes;
- Effect size and correlation were taken in Pedrono *et al.* (2009).
- We assume a common unstructured covariance matrix for both vaccinal strategies

	Normal			Kshirsagar		
	$r = 3$	$r = 5$	$r = 7$	$r = 3$	$r = 5$	$r = 7$
Bonferroni	21	50	201	22	52	202
Hochberg (modified)	23	48	147	24	49	148
Holm	20	41	116	21	42	116

Sensitivity analysis





rPowerSampleSize Package

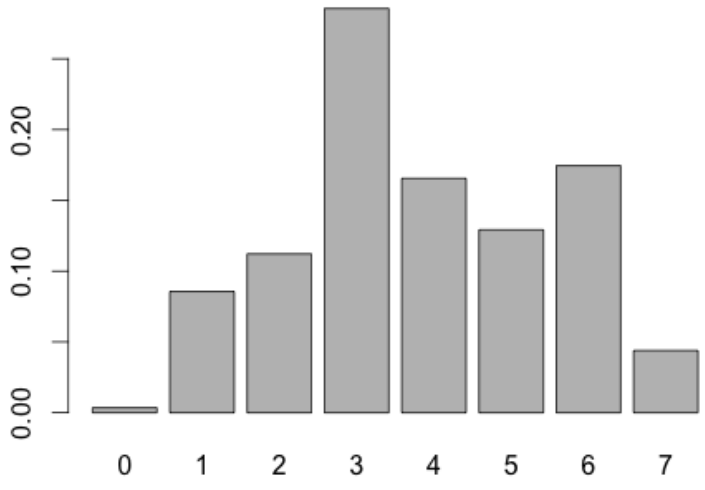
- `rPowerSampleSize` package is available on `http://www.r-project.org`
- First designed for the case $r = 1$ (see Lafaye *et al* (2014))
- The new version of the package can tackle any value of $r \leqslant m$.
- It includes functions related to power computation (`Psirmd()`, `Psirms()`, `Psirmu()`)
- The main function is `indiv.rm.ssc()` related to sample size determination controlling the q-gFWER, for a given value of r-power.

R code related to The Pneumovac trial

```
> nCovernE <- 1
> m <- 7
> r <- 3
> alpha <- 0.05
> pow <- 0.8
> q <- 1
> asympt <- FALSE # corresponding to Kshirsagar distribution
> delta <- c(0.55, 0.34, 0.38, 0.20, 0.70, 0.38, 0.86)
> sigma <- c(0.3520, 0.6219, 0.5427, 0.6075, 0.6277, 0.5527, 0.8066)
> var <- sigma ^ 2
> SigmaE <- SigmaC <- cov
> maxpts <- 2500000
> abseps <- 0.001
> result <- indiv.rm.ssc(method = "Bonferroni", asympt = asympt, r = r, m = m,
+    p=m, nCovernE = 1, muC = NULL, muE = NULL, d = NULL, delta = delta,
+    SigmaC = cov, SigmaE = cov, power = pow, alpha = alpha,
+    interval = c(2, 100), q = q, maxpts = maxpts, abseps = abseps)
 > result
[1] 22
```

From this finding ($n = 22$) the user could visualise the distribution of the number of significant results (i.e, the realized values r) by using the `plot.rPower()` function

```
> nbcores <- parallel::detectCores() - 1
> set.seed(10)
> res.MC <- montecarlo(method = "Bonferroni", M = 10 ^ 4, nE = 22, r = 3, m = 7,
+    nCovernE = 1, muC = rep(0 , 7), muE = delta, d = rep(0.0, 7),
+    SigmaE = cov, SigmaC = cov, alpha = 0.05, q = 1, nbcores = nbcores)
> res.MC$rpowBonf
[1] 0.7987
> plot.rPower(res.MC) # To produce plot in Figure 4.
```



Concluding Remarks

- General power formulas has been derived when one wants at least r among m statistical tests to be significant.
- Formulas have been used to compute the necessary sample size to control weakly or strongly the type-II r-generalized family-wise error rate, for procedures that already control any type-I global error rate.
  - Weak control at level β of the type-II r-generalized family-wise error rate is reached when $\beta_{r,m}(P) \leqslant \beta$ for a potential choice P of the true model under which all null hypotheses tested are false.
  - Strong control at level β occurs when $\beta_{r,m}(P) \leqslant \beta$ for all potential choices P of the true model such that $p \geqslant r$ null hypotheses are false.
- Available through rPowerSampleSize R package
  - "At least one win": Global and Individual methods;
  - "At least r wins": Single step and Step-Wise methods (Bonferroni, Holm and Hocberg)
- A parallel implementation is available using the argument `nbcores`.
- Focus on continuous multiple endpoints $\rightarrow$ Extend our work to categorical, and mixed primary endpoints ...

References

- Chen J, Luo J, Liu K, Mehrotra D. On power and sample size computation for multiple testing procedures. Computational Statistics & Data Analysis 2011; 55(1):110-122.
- Delorme P., Lafaye de Micheaux P., Liquet B. and Riou J., Type-II Generalized Family-Wise Error Rate Formulas with Application to Sample Size Determination. *Statistics in Medicine* (2016) In press.
- Dmitrienko A, D'Agostino Sr R. Traditional multiplicity adjustment methods in clinical trials. Statistics in Medicine 2013; 32(29):5172-5218.
- Lafaye de Micheaux P, Liquet B, Marque S, Riou J. Power and sample size determination in clinical trials whith multiple primary correlated endpoints. Journal of Biopharmaceutical Statistics 2014; 24(2):378-397.
- Lehmann E, Romano J. Generalizations of the familywise error rate. The Annals of Statistics 2005; 33(3):1138-1154.
- Maurer W, Margolin B. The multivariate inclusion-exclusion formula and order statistics from dependent variates. The Annals of Statistics 1976.
- Pédrono G, Thiébaut R, Alioum A, Lesprit P, Fritzell B, Lévy Y, Ch?ene G. A new endpoint definition improved clinical relevance and statistical power in a vaccine trial. Journal of Clinical Epidemiology 2009; 62(10):1054-1061.
- Romano J, Shaikh A. Stepup procedures for control of generalizations of the familywise error rate. The Annals of Statistics 2006; 34(4):1850-1873.

ANY QUESTIONS ?