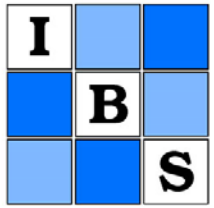


CompARE: Sample Size Calculation for Time-To-Event Composite Endpoints

Jordi Cortés, Marta Bofill, Lupe Gómez

IBC
July 2018

Outline

- Non proportional hazards with composite endpoints
- Methodologies
  - Naïve
  - ARE
  - Simulation
- CompARE: Naïve vs simulation & ARE vs. simulation

Introduction to the problem

- Often clinical trials (CT) have several options for the primary endpoint.
- Researchers must decide about using one or more than one of these endpoints.
- One of the biggest concerns in using composite endpoint (CE) in time-to-event studies arises from the lack of proportional hazards[1].
- Sample size computation may become a great challenge in the design phase of a CT.

[1] Gómez G. and Lagakos SW. (2013). *Statistical considerations when using a composite endpoint for comparing treatment groups*. Statistics in Medicine.

Introduction to the problem

- We deal with the situation of **two** endpoints:

$$\boldsymbol{\varepsilon_1}\ (relevant) \quad \& \quad \boldsymbol{\varepsilon_2}\ (additional)$$

- The composite endpoint (ε^*) is defined as follows:

$$\varepsilon_* = \varepsilon_1 \cup \varepsilon_2$$

- Common CE in literature:
  - **Progression Free Survival** (PFS): Death and Progression of disease
  - **Major Adverse Cardiac Events** (MACE): cardiovascular death, myocardial infarction, stroke, or non-coronary artery bypass graft-related major bleeding

Non-proportional hazards in CE

- A possible measure of non-proportionality of the hazards might be the difference between the maximum and the minimum $HR_*(t)$ of the composite event (ε_*) over time:

$$r = \max\{HR_*(t)\} - \min\{HR_*(t)\} \qquad t \epsilon [0, \tau]$$

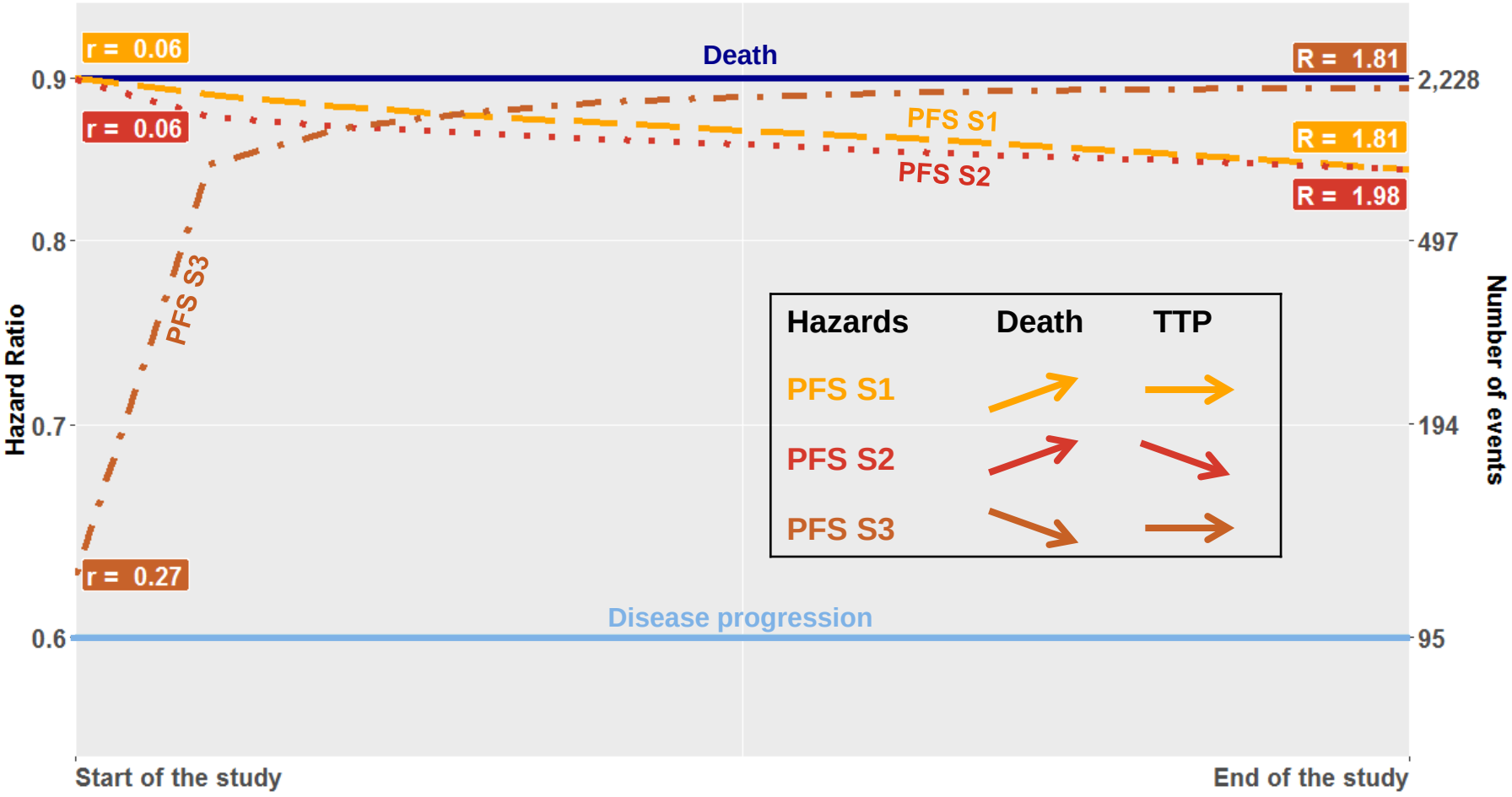
- An alternative measure of non-proportionality is:

$$R = \left(\frac{\log(averaged\{HR_*(t)\})}{\log(\max\{HR_*(t)\})}\right)^2 \qquad t \epsilon [0, \tau]$$

This measure represents the ratio of the samples sizes considering the minimum detectable effect (MHR_*) and the averaged effect (aHR_*):

$$R = \frac{n_{MHR_*}}{n_{aHR_*}}$$

Non-proportional hazards in CE

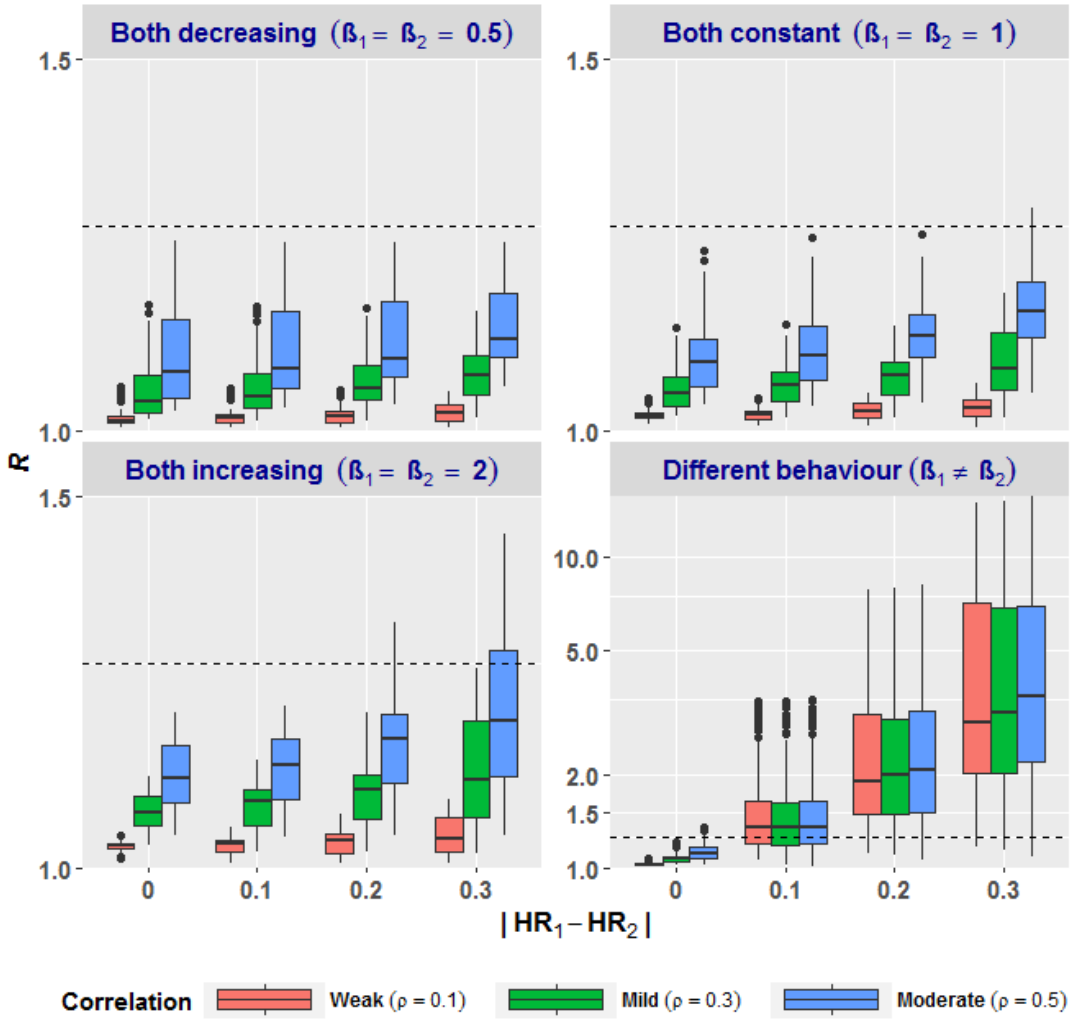


$$r = max\{HR_*(t)\} - min\{HR_*(t)\} \rightarrow \text{Effect size}$$

$$R = \frac{n_{MHR_*}}{n_{aHR_*}} \rightarrow \text{Sample size}$$

Non-proportional hazards in CE

- *R* measure depends especially on the behaviour of the marginal distributions.
- Non-proportionality increases if:
  - Hazards in both components go in opposite directions
  - Difference between the effects (HRs) is large
  - Correlation is large



Needed information for Sample Size

- **Probability** of observing events
- **Effect size**
- If **death** is one of the components
- Information about the **marginal distributions**
- **Correlation** between endpoints
- Specified probabilities of Type I (α) and Type II (β) **errors**

Naïve method

- (Some) clinicians use the averaged HR for determining the SS:

$$HR_*(t) = HR' = \frac{HR_1 + HR_2}{2}$$

- Formulas for calculate SS with a single endpoint:

$$\left.\begin{array}{ll} (Schoendfeld) & E = \frac{4 \cdot \left(z_{1-\alpha} + z_{1-\beta}\right)^2}{(\ln(HR'))^2} \\ (Freedman) & E = \frac{\left(HR'+1\right)^2 \cdot \left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{(HR'-1)^2} \end{array}\right\} \rightarrow N = \frac{2E}{p_{10} + p_{11}}$$

ARE method

- The asymptotic relative efficiency[1] (ARE) is a measure of how much efficient could be a design based on the ε^* respect to one based on the relevant endpoint (ε_1)

- We want to know if *ARE* is a good approximation for the SS ratio between designs[2] using ε_* & ε_1

$$ARE = \frac{N_1}{N_*} \rightarrow N_* = \frac{N_1}{ARE}$$

This method might not guarantee the right computation of the SS.

[1] Gómez G. and Lagakos SW. (2013). *Statistical considerations when using a composite endpoint for comparing treatment groups*. Statistics in Medicine.
[2] Gómez G, and Gómez-Mateu M. (2014). The Asymptotic Relative Efficiency and the ratio of sample sizes when testing two different null hypotheses.

Simulation. Procedure

1. Range of **initial values**
   - Calculate SS for N_* based on ARE method
   - Looking for values (N_{SIM}^i) into the following interval until reaching a target power (e.g. 80%):

$$[\,0.8 \cdot N_*, 1.2 \cdot N_*]$$

[If the power is not reached, the interval is extended]

2. We perform **1,000 iteration** for each N_{SIM}^i:
   - **Generate N_{SIM}^i values** from the specific distributions and correlation through a pre-specified copula.
   - **Censore** these values according to the follow-up time.
   - Perform **log-rank test** & obtain the **p-value**.
3. **Power** is the proportion of significant results into the 1,000 iterations

Simulation. Scenarios

Parameter	Endpoint	Values	Scenarios
Probability of event	*Relevant*	$\pi = 0.01, 0.05, 0.10$	3
	Additional	$\pi = 0.05, 0.10, 0.20$	3
Distribution	*Relevant*	Exponential ($\beta_1 = 1$)	1
	Additional	Exponential ($\beta_2 = 1$)	1
HR	*Relevant*	HR = 0.6, 0.7, 0.8	3
	Additional	HR = 0.6, 0.7, 0.8	3
Death	*Relevant*	Yes	1
	Additional	No	1
Correlation	-	$\rho = 0$, 0.2, 0.5, 0.8	4
Copula	-	Frank	1
Type Error I	-	$\alpha = 0.05$ (one sided)	1
Power	-	$1 - \beta = 0.8$	1
		no. scenarios	**324**
		Time by scenario (min)	**2**
		Total time (hours)	**9**

CompARE: tool for CE in RCTs

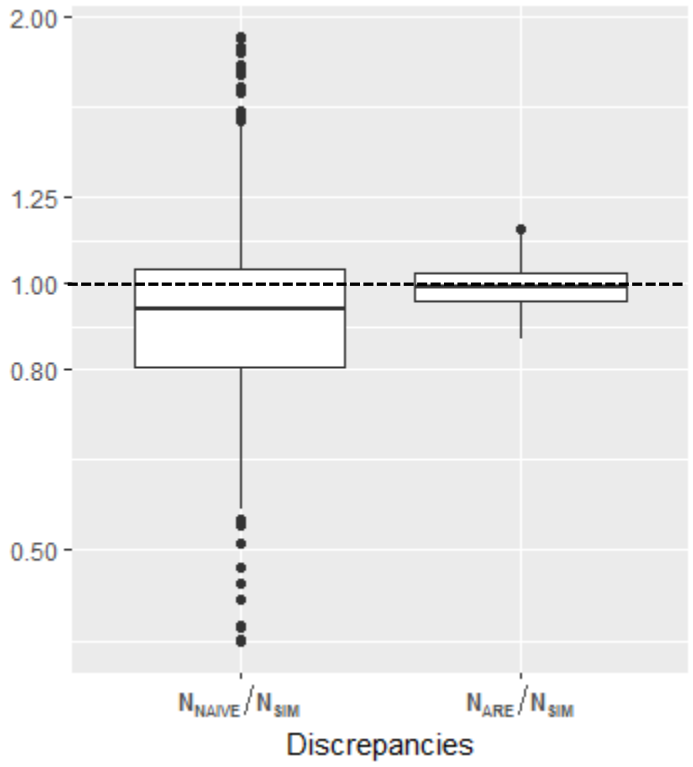
CompARE[3] is a web app built with *shiny*[4]. Intended to help researchers in the design and analysis of clinical trials with CE for binary and time-to-event endpoints

[3] Gómez-Mateu M and Gómez G. Clinical trial designs using CompARE. An on-line exploratory tool for investigators. Report DR 2017/1

[4] RStudio, Inc (2017) shiny: Web Application Framework for R. URL http://CRAN.R-project.org/package=shiny. R package version 1.0.5.

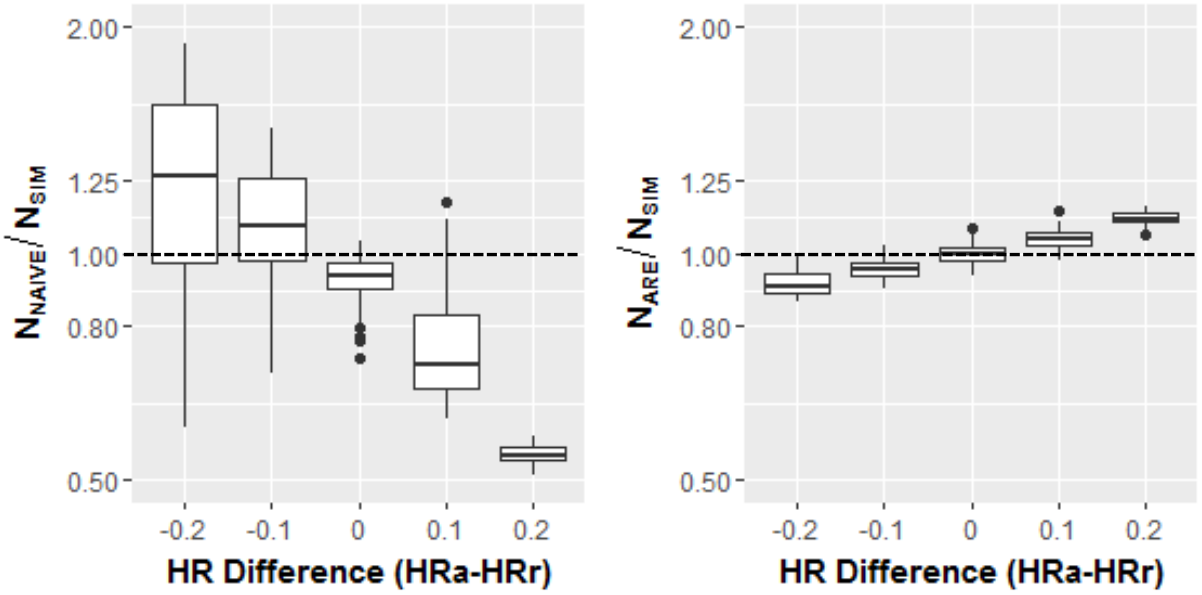
Comparison **Naïve versus ARE (I)**

- We compare Naïve and ARE methodology in respect to simulations results



	N_{NAIVE}/N_{SIM}	N_{ARE}/N_{SIM}
Min.	0.39	0.86
Q1	0.81	0.95
Median	0.94	0.99
Mean	0.96	0.99
Q3	1.04	1.02
Max.	1.90	1.15

Comparison **Naïve versus ARE (II)**



- SS discrepancies only depend on **HRs**, don't depend on probability of event nor correlation.
- ARE method **fits better** the actual SS than Naïve method
- ARE method works better when **HRs match**

Conclusions

- Computing SS based on **ARE is a better option** than to average the single HRs unless we have similar effect sizes for the different endpoints

- To assure a power of 80% in at least 90% of designs, **we recommend an increase of a 10% in the SS** obtained with ARE.

- We would recommend to **perform simulations and/or use CompARE** in order to:
  - Estimate SS with greater accuracy
  - Asses its robustness according to the assumptions about the parameters.

References

1. Gómez G and Lagakos SW. (2013). *Statistical considerations when using a composite endpoint for comparing treatment groups*. Statistics in Medicine, 32, 19-738.

2. Gómez G, and Gómez-Mateu M. (2014). *The Asymptotic Relative Efficiency and the ratio of sample sizes when testing two different null hypotheses*. SORT, 38, 73-88.

3. Gómez-Mateu M and Gómez G. Clinical trial designs using CompARE. An on-line exploratory tool for investigators. Report DR 2017/1

4. RStudio, Inc (2017) *shiny: Web Application Framework for R*. URL http://CRAN.R-project.org/package=shiny. R package version 1.0.5.

THANKS FOR YOUR ATTENTION

Jordi Cortés, Marta Bofill, Lupe Gómez

IBC
July 2018