42. EXTENSIONS AND APPLICATIONS OF COMPARE: WEB PLATFORM TO SELECT THE PRIMARY ENDPOINT IN A RANDOMIZED CLINICAL TRIAL

M. Gómez-Mateu and G. Gómez Melis
Universitat Politècnica de Catalunya, Department of Statistics and Operations Research, Spain,
moises.gomez.mateu@upc.edu

ABSTRACT

We present CompARE, a web platform that computes how more efficient is to use a composite endpoint instead of one of its components. CompARE is based on the Asymptotic Relative Efficiency (ARE) method developed by Gómez and Lagakos. Users interact with CompARE through HTML form pages and no knowledge of R is needed. A list of candidate endpoints, anticipatable probabilities and hazard ratios are required to use CompARE. Results are shown immediately through plots, text and tables. We present advanced capabilities of CompARE such as assigning probabilities and hazard ratios to combinations of several endpoints (relevant or additional).

1. INTRODUCTION

Conclusions from randomized clinical trials (RCTs) rely on its primary endpoint chosen at the design stage of the study. In order to provide clinical evidence related to the primary objective of the trial the selection of the primary endpoint is of outmost importance.

Composite endpoints (CEs) consisting of the union of two or more outcomes are often used as the primary endpoint in RCTs. For instance, in time-to-event studies, time to MACE is a CE generally defined as the union of cardiovascular death,

reinfarction, target vessel revascularization and stroke. Pros and cons on the use of CEs have been extensively discussed in the literature. From a medical perspective, the use of CEs allows to combine multiple measurements, avoiding the need to choose a single outcome as primary endpoint. However, some cautions have to be considered when defining a composite endpoint. A CE should only be used if the individual components are clinically meaningful and of similar importance to the patient, the expected effects on each component are similar based on biological plausibility, and the clinically more important components do not affect negatively (Ferreira-González *et al.*, 2007). Otherwise, the use of the CE is questionable and may lead to misleading conclusions. From a statistical point of view, the rationale for considering a CE, other than multiplicity, is the likely reduction of the sample size needed to achieve a desired statistical power for a given significance level, as a result of increasing the event rate to attain a predefined treatment effect. Nevertheless, it has been discussed and shown in Gómez and Lagakos (2013), that adding inappropriate components to the relevant endpoint might actually lead to a loss of power to detect the true treatment differences.

2. THE ARE METHOD TO CHOOSE THE PRIMARY ENDPOINT

Gómez and Lagakos (2013) develop a statistical methodology that helps to decide between using a Relevant endpoint ε_R instead of a composite endpoint, consisting of the union of ε_R plus another additional endpoint ε_A, to evaluate the effect of a treatment. Their strategy is based on the value of the asymptotic relative efficiency (ARE).

The ARE relates the efficiency of using the logrank test Z_R based on ε_R versus the efficiency of the logrank test Z_* based on the CE. It can be shown that Z_R and Z_* are asymptotically $N(0, 1)$ under the null hypothesis of no treatment effect and asymptotically $N(\mu_R, 1)$ and $\mathrm{N}(\mu_*, 1)$, respectively, under a sequence of contiguous alternatives to the null hypothesis. Under these conditions, $\mathrm{ARE}(Z_*, Z_R) = (\mu / \mu)^2$. The composite endpoint will be advisable to use as primary endpoint whenever the $\mathrm{ARE}(Z_*, Z_R) > 1$. Otherwise, ε_R will be considered the best choice.

We assume that: i) the end-of-study censoring is the only non-informative censoring cause for both groups; ii) the hazard ratios HR_R and HR_A for ε_R and ε_A, respectively,

are constant; iii) marginal Weibull distributions for the time T_R and T_A to ε_R and ε_A, respectively; and iv) Frank's copula binding the marginals (T_R, T_A) through Speraman correlation ρ. Under these assumptions, when the additional endpoint ε_A does not include death, the expression of the ARE is given by:

$$\mathrm{ARE}(Z_*, Z_R) = \frac{\left(\int_0^1 \log\left\{\lambda_*^{(1)}(t)/\lambda_*^{(0)}(t)\right\} f_*^{(0)}(t)dt\right)^2}{\left(\log\left\{\mathrm{HR_R}\right\}\right)^2 \cdot\, p_{*} \cdot\, p_R},$$

where p_R and p_* are the probabilities of observing ε_R and ε_* in control group, respectively; and $\lambda_*^{(0)}(t)$ and $\lambda_*^{(1)}(t)$ are the hazard functions of the times $\mathrm{T}_*^{(0)}$ and $\mathrm{T}_*^{(1)}$ to the composite endpoint for each group, respectively. The ARE expression in either censoring case is expressed in terms of the following interpretable parameters:

- Probabilities p_R, p_A of observing ε_R, ε_A, respectively, in control group,
- relative treatment effects given by the hazard ratios HR_R, HR_A, and
- Spearman's correlation coefficient ρ between T_R and T_A.

The ARE as ratio of sample sizes

If we were to test H_0 versus H_a with two different test statistics S_n and T_m , the asymptotic relative efficiency would be defined as the ratio m/n, where n and m are the required sample sizes for S_n and T_m, respectively, to attain the same power for a given significance level.

In our setup we have two different set of hypotheses: H_0 versus H_a testing the treatment effect on ε_R and H_0^* versus H_a^* testing the treatment effect on 0 a ε_R. Gómez and Lagakos prove that the ARE can be interpreted as n_R/n_*, where n_R and n_* stand for the required sample sizes needed when using the Relevant endpoint or the composite endpoint, respectively.

3. CompARE: A TOOL TO CHOOSE THE PRIMARY ENDPOINT IN A RCT

With the aim of making the ARE method widely applicable to the scientific community, a free web-based platform called *CompARE* has been created.

CompARE provides a tool to compute the ARE among several subsets of relevant components. It is of great help when planning a clinical trial since it quantifies how efficient is a relevant subset of outcomes respect to a larger subset. CompARE is a friendly tool, based on the free software Tiki Wiki CMS/Goupware. Although it is internally programmed in R, users do not need knowledge of R, neither to have R installed in their computer. By means of web-page forms, users can easily introduce the required information, step by step. This information is saved and executed by the system, returning ad-hoc results depending on each case.

ARE values are calculated from the anticipated information of the probabilities and hazard ratios p_R, p_A, HR_R and HR_A by means of an input grid. Moderate correlations ρ and exponential distributions are considered by default. Other advanced options are available such as Weibull distributions with decreasing or increasing hazard rates or different correlations.

Results from CompARE are shown by means of summary tables and plots. Outputs combining several HR_A together with different correlation values are of great help to understand the role that the Additional endpoint plays (see Figure 1). Additionally, survivals and hazard ratios for ε_R, ε_A and CE are depicted. A history table saves each result to compare previous analyses. Moreover, conclusions and recommendations are given in written form as an aid. CompARE is currently accessible as a beta version on the following website (http://composite.upc.edu/CompARE).

4. APPLICATIONS TO THE CARDIOVASCULAR RESEARCH AREA

The ARE method has been applied to the cardiovascular research area (Gómez, Gómez-Mateu and Dafni, 2014), where a set of general recommendations are reported from RCTs published in 2008. General guidelines are provided based on the frequencies of observation and treatment effects of the Relevant and the Additional

endpoints. In the following subsection we illustrate some of the CompARE capabilities by means of a real RCT.

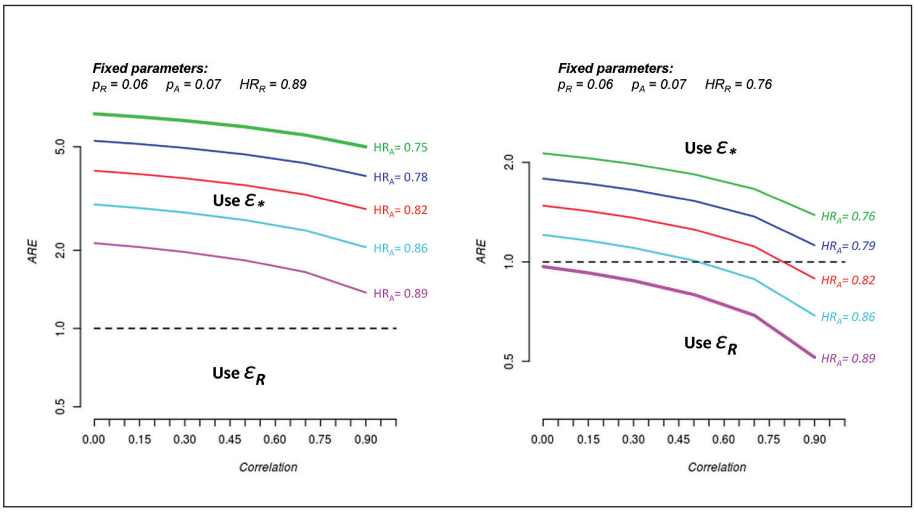

Figure 1. ARE values combining different values of the hazard ratio HR_A of the Additional endpoint ε_A and correlation between times to the Relevant endpoint ε_R and ε_A. Fixed parameters: probabilities p_R and pA of observing ε_R and ε_A, respectively, in control group; and hazard ratio HR_R of ε_R.

A case study: The LIFE trial

The Losartan Intervention For Endpoint reduction in hypertension study (LIFE) trial (Dahlöf *et al.*, 2002) was performed to test the efficacy of Losartan-based antihypertensive treatment in patients with hypertension. The primary composite endpoint (CE) was composed by cardiovascular death, myocardial infarction and stroke. Cardiovascular death + myocardial infarction are considered the most clinically important components (Sankoh *et al.*, 2014), and hence we refer to them as the Relevant endpoint ε_R, and stroke as the Additional endpoint ε_A. Significant results were achieved using the CE, consequence of the significant effect on stroke (see Table 1).

Endpoint	Type	Control treatment Probability (n)	Hazard ratio (CI, 95%)	p-value
Cardiovasc. mortality	ε_{r1}	0.05 (234)	0.89 (0.73 – 1.07)	0.206
Myocardial infarction	ε_{r2}	0.04 (188)	1.07 (0.88 – 1.31)	0.491
Stroke	ε_A	0.07 (309)	0.75 (0.63 – 0.89)	0.001
ε_{r1} U ε_{r1} U ε_A	Composite	0.13 (588)	0.87 (0.77 – 0.98)	0.021

Table 1. Summary of the results from the LIFE trial. ε_{r1} and ε_{r2} stand for Relevant endpoint component 1 and 2 respectively; ε_A stands for Additional endpoint; and CI stands for the confidence interval.

We present an ARE study based on two different situations. In both we will assume a probability of observing ε_R and ε_A in control group of p_R = 0.06 and p_A = 0.07, respectively, as feasible anticipated values of this study.

If we anticipate relative treatment effects as reported in the LIFE study, that is HR_R = 0.89 for the relevant endpoint and HR_A = 0.75 for the additional endpoint, we would have recommended to use the composite endpoint, as they did. Indeed, as Figure 1 (left) shows, the ARE value is always greater than 1 for every possible correlation between ε_R and ε_A.

However, if the expected relative treatment effect on the relevant endpoint would have been stronger, for instance HR_R = 0.76, then the CE would have not been always advised. As shown in Figure 1 (right), we would conclude that using ε_R as primary endpoint would be the best recommendation whenever the relative treatment effect of the Losartan intervention on stroke was larger than 0.9. On the other hand, adding stroke to the primary endpoint would be advisable if the relative treatment effect was at most 0.8. For the situations where the relative treatment effect on stroke were expected to be between 0.8 and 0.9 special attention should be paid at the correlation between cardiovascular death + myocardial infarction and stroke.

5. PRACTICAL ISSUES TO ASSIGN ANTICIPATABLE COMBINED PROBABILITIES AND HAZARD RATIOS

The ARE method is based on the assumption that, even when the Relevant endpoint consist of several components, both combined probability and hazard ratio can be

anticipated by researchers (analogously, for the Additional endpoint). In some occasions trialists might anticipate these probabilities and hazard ratios for each one of the components. In what follows, we describe how to compute the combined probability and hazard ratio and discuss how it is implemented in CompARE.

Following Bahadur (1961), the combined probability p_R can be computed in terms of the probability components p_{R1} and p_{R2} of observing the relevant components ε_{R1} and ε_{R2} in the control group, respectively, for a given correlation δr between endpoints (more than two components can also be considered). Concerning the combined HR_R, the exact value is in terms not only of the hazard ratios HR_{r1} and HR_{r2} of ε_{R1} and ε_{R2}, respectively, but also depends on the probabilities p_{r1} and p_{r2}, the marginal laws of each component, the correlation and the joint distribution between components. Furthermore, HR_R does not have to be constant even though HR_{r1} and HR_{r1} are constant.

CompARE implements, as advanced options, the possibility of assigning a value for p_R and for HR_R. In particular, whenever the correlation between ε_{R1} and ε_{R2} is unknown, the user can choose between any value within the following bounds: max(p_{r1}, p_{r2}) and $p_{R1} + p_{R2}$ (see Figure 2). For the combined HR_R a set of possible values is as well proposed.

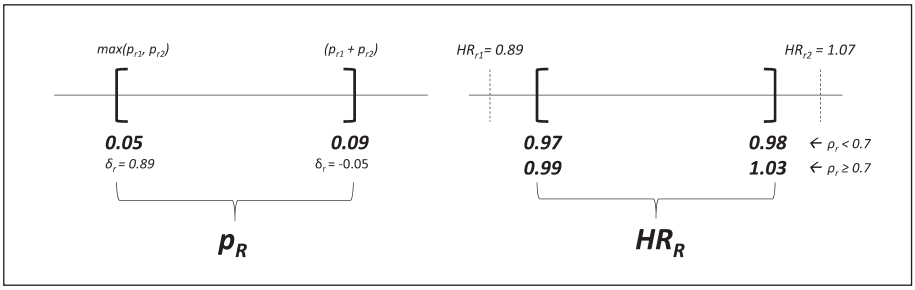

Figure 2. Boundaries for the combined probability p_R and hazard ratio HR_R in terms of Pearson's correlation δr and Spearman's correlation ρ_r, respectively. p_{r1} and p_{r2} stand for the marginal probability of observing the Relevant component ε_{r1} and ε_{r2} in control group, respectively. HR_{r1} and HR_{r2} stands for the marginal hazard ratio of each component. Component parameter values are taken from the LIFE study.

6. CONCLUDING REMARKS

This paper addresses the ARE method as an intuitive and interpretable way of comparing the efficiency of several endpoints, candidates for the primary endpoint of a RCT. We present as well CompARE, a web platform that performs all the ARE computations, which can be freely accessed and is friendly to use. Ongoing research include several extensions of CompARE :

1. Sample size computation based on the anticipatable parameters.

2. Combined probabilities and hazard ratios.

3. Computations when both $\mathcal{E}_R$ and $\mathcal{E}_A$ include death.

4. Different copulas other than Frank's.

ACKNOWLEDGMENTS

This research is partially supported by Grant MTM2012-38067-C02-01 from the Ministerio de Economía y Competitividad and by Grant 2014 SGR 464 from the Departament d'Economia i Coneixement de la Generalitat de Catalunya.

REFERENCES

Bahadur, R.R. (1961). "A Representation of the Joint Distribution of Responses to n Dichotomous Items". *Stanford University Press*, 158-168.

Dahlöf, B., Devereux, R.B., Kjeldsen, S.E., Julius, S., Beevers, G., de Faire, U., Fyhrquist, F., Ibsen, H., Kristiansson, K., Lederballe-Pedersen, O., Lindholm, L.H., Nieminen, M.S., Omvik, P., Oparil, S., Wedel, H., and LIFE Study Group (2002). "Cardiovascular Morbidity and Mortality in the Losartan Intervention For Endpoint Reduction in Hypertension Study (LIFE): A Randomised Trial Against Atenolol". *Lancet*, 359: 995-1003.

Ferreira-González, I., Permanyer-Miraldaa G., Busse J.W., Bryant D.M., Montori V.M, Alonso-Coello P., Walter S.D, and Guyatt G.H. (2007). "Methodologic Discussions for Using and Interpreting Composite Endpoints are Limited, but still Identify Major Concerns". *J Clin Epidemiol*, 60, 651-657.

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.

Gómez, G., Gómez-Mateu, M., and Dafni, U. (2014). "Informed Choice of Composite End Points in Cardiovascular Trials". *Circulation*, 7: 170-178.

Gómez, G., and Lagakos, S.W. (2013). "Statistical Considerations when Using a Composite Endpoint for Comparing Treatment Groups". *Statistics in Medicine*, 32, 719-738.

Gómez-Mateu, M., and Gómez, G. (Last date of access: January 13, 2015). http://composite.upc.edu/CompARE.

Sankoh, A.J., Lia, H., D'Agostino, R.B. (2014). "Use of Composite Endpoints in Clinical Trials". *Statistics in Medicine,* 33: 4709-4714.