Clinical trial designs using CompARE. An on-line exploratory tool for investigators

Moisés Gómez-Mateu[1], Guadalupe Gómez Melis[1]

[1]Dept. of Statistics and Operations Research
Universitat Politècnica de Catalunya
Jordi Girona 1–3, 08034 Barcelona, Catalonia (Spain)

Report DR 2017/1
March 2017

Report

Clinical trial designs using CompARE. An on-line exploratory tool for investigators.

Moisés Gómez-Mateu, Guadalupe Gómez Melis

Background Conclusions from randomized clinical trials (RCT) rely primarily on the primary endpoint (PE) chosen at the design stage of the study. There should generally be only one PE which should be able to provide the most clinically relevant and scientific evidence regarding the potential efficacy of the new treatment. Therefore, it is of utmost importance to select it appropriately.

Composite endpoints, consisting of the union of several endpoints, are often used as PE in RCT. Gómez and Lagakos (2013) develop a statistical methodology to evaluate the convenience of using a CE as opposed to one of its components. Their strategy is based on the asymptotic relative efficiency (ARE), relating the efficiency of using the logrank test based on the CE versus the efficiency based on one of its components. This paper introduces the freeware online platform *CompARE* that facilitates the study of the performance of different candidate endpoints which could be used as PE at the design stage of a trial. *CompARE*, through an intuitive interface, implements the novel ARE method.

Results *CompARE* is an exploratory tool that can be used for a variety of purposes. First, to study how efficient might be a CE compared to one of its components in the sense of requiring a smaller sample size for the same significance level and power. Second, to compute the required sample size both for marginal components and for CE, even when the proportionality of the hazards' assumption for CE does not hold. Third, to visualize the shape of the hazard ratio over time allowing to check possible departures from constancy. Fourth, to provide other graphical outputs such as the survival functions for each endpoint. By means of a cardiovascular case study, we briefly illustrate the main capabilities of the platform.

Conclusion Our software helps trialists to make more informed decisions on the PE when encountered with several candidate endpoints in a study. Results for different parameter values are shown immediately through tables and plots with a user-friendly graphic interface. Conclusions and recommendations are also provided in written form. *CompARE* is accessible at: https://cinna.upc.edu/compare.

Keywords: Asymptotic relative efficiency, CompARE, composite endpoint, non-proportional hazards, sample size, web-based application.

Background

Composite outcomes consist of the union of several outcomes and are frequently used in randomized clinical trials (RCT). In time-to-event analysis, composite endpoints (CE) refer to the elapsed time from randomization until the earliest observation among its components. In oncology trials, for instance, progression-free survival is defined as the time to disease progression or death, whichever occurs first.

The decision on whether to use a CE versus a single component as the primary endpoint (PE) is controversial. The advantages and drawbacks regarding the use of CE have been extensively discussed in the literature (Ferreira-González, 2007; Freemantle, 2007; Tomlinson, 2010)[1, 2, 3]. A major controversy rely on the significance of the CE when it does not hold for the main component. Some authors have proposed approaches that combine the superiority of the CE with the non inferiority of the main relevant component (Rauch and Kieser, 2013)[4]. However, multiplicity adjustment procedures for multiple outcomes, and assumptions on the effect sizes and correlations between components must be taken into account (Song, 2009)[5].

Gómez and Lagakos (2013)[6] develop a statistical methodology in order to evaluate the convenience of using a relevant endpoint (RE), for instance cardiovascular death, versus a CE consisting of the union of the RE plus another additional endpoint AE, such as hospitalization. Their strategy is based on the value of the asymptotic relative efficiency (ARE), which relates the efficiency of using the logrank test based on the time to the CE versus the efficiency based on the time to the RE. This behaviour leads to the following criterion: choose the CE whenever ARE > 1 and keep the RE as the primary endpoint for the trial otherwise. The ARE can be interpreted as the reciprocal ratio of the sample sizes required for the logrank test based on RE and on CE to attain the same power at the same significance level (Gómez and Gómez-Mateu, 2014)[7].

The ARE can be expressed in terms of a short number of parameters that investigators can anticipate at the design stage of the trial as follows:

1. The event rates $p_R^{(0)}$, $p_A^{(0)}$ of observing the RE and the AE, respectively, in the control group, during the follow-up of the trial.

2. The relative treatment effects on the RE and the AE given by the hazard ratios HR_R and HR_A, respectively, which we assume that they are constant over time.

3. The Spearman correlation coefficient ρ between the times to both RE and AE components of the CE.

The ARE method is implemented for administrative censoring for a given follow-up period, and uses Weibull marginal times for the RE and the AE because of its flexibility, which allows decreasing, constant and increasing hazards. Furthermore, we consider a Frank's Archimedean survival copula relationship between the marginal times to derive the joint distribution. Further details and assumptions regarding the joint behaviour of the times to the RE and AE can be found in Gómez and Lagakos' (2013) paper[6].

To the best of our knowledge, there is not a software tool to study the efficiency of a composite endpoint versus its components, neither to compute sample sizes for non-constant hazard ratios.

In the current article, we present the new online tool *CompARE*. This software platform aids investigators the study of the behaviour of several endpoints considered as the PE at the design stage of a trial. This is accomplished by implementing the aforementioned ARE method. *CompARE* can be viewed as an exploratory tool since it allows to generate a variety of scenarios depending on different parameter values. It helps to evaluate how efficient might be a CE versus one of its components in terms of sample size for a given significance level and power. It is also a computational resource for trialists to calculate sample sizes, both for the CE and for its components, even when the assumption of the proportionality of the hazards for the CE does not hold. Graphical outputs permit to depict the distribution of each endpoint and the hazard ratios of the CE over time, providing a practical tool to visually check the proportionality of the hazards.

Users introduce the information needed in *CompARE* through intuitive web-page forms, such as the list of candidate endpoints, together with the anticipated parameter values. Results for different parameter settings are shown immediately by means of tables, plots and in written form.

We remark that the development of a tool as *CompARE* has been imperative because of the complexity of the ARE expression. The reader is referred to the Appendix for the explicit expression for the ARE values. All the computations have been programmed in *R*[8]. Furthermore, users need no knowledge of *R*, nor do they need to install it on their computers, making *CompARE* a useful tool for trialists.

The LIFE study. A Motivating example

Throughout the next pages we will use The Losartan Intervention For Endpoint reduction in hypertension study (LIFE) trial (Dahlöf et al., 2002)[9] as an aid to present how *CompARE* works together with its capabilities.

The LIFE study was conducted to test the efficacy of an antihypertensive treatment (Losartan) in patients with hypertension. The primary composite endpoint was composed by the union of cardiovascular death, myocardial infarction and stroke. While cardiovascular death and myocardial infarction (CV death + MI) are considered the most clinically important components (Sankoh et al., 2014)[10], and we refer to them as the relevant endpoint RE, stroke acts as the additional endpoint AE (see Figure 1).

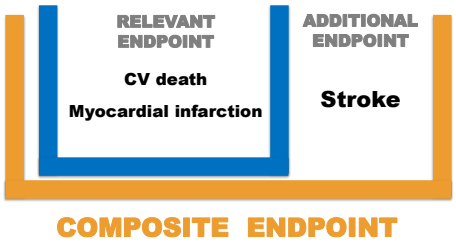

Figure 1: Pictorial representation of the composite endpoint and its components based on the LIFE trial. CV stands for cardiovascular.

In this study, significant results using the time to the CE were achieved, mainly due to the significance of the less important outcome, i.e. stroke. The main question that our approach tries to answer is under which circumstances the inclusion of only the most important outcomes CV death + MI would have been recommended as the primary endpoint. This illustration brings to light the need for a practical tool for trialists based on the previously described methodology in order to know in which specific cases the use of the RE alone or the use of CE is recommended in terms of efficiency, as we detail in the next sections.

Implementation

Tiki: software underneath *CompARE*

The web-based platform *CompARE* is built under the free software Tightly Integrated Knowledge Infrastructure (Tiki Wiki CMS/Goupware)[11]. Free and open-source software is widely used and is commonly developed with volunteer computer programmers, guaranteeing that every user has equal rights of access. It allows any programmer to study the source code, modify it, and share it (FSF)[12]. Moreover, online graphic interfaces leads to collaboration synergies between partners from different areas such as computing, biology or statistics.

We chose Tiki because the majority of other interface web programs that include the use of *R* routines present problems in the short or medium term (de Pedro and Sánchez, 2010)[13]. Moreover, Tiki is safe and updated periodically by their community members, who add new features, fix bugs and patch security holes. It is constantly maintained under the license LGPL (Lesser General Public License). Repositories are used for the version control system.

Other remarkable features of Tiki are the use of web-standard codes such as HTML, PHP and javascript. Due to its flexibility, different applications can be included through the use of *plugins* (Sapir, 2010)[14]. By means of the *pluginR*[15], developed by De Pedro and Sánchez (2010)[13], it is possible to execute the ARE method programmed with *R*[8] in *CompARE* (see the scheme in Figure 2). We highlight that it is not necessary to install *R* locally in a computer to run *CompARE*, nor is knowledge of *R* required.

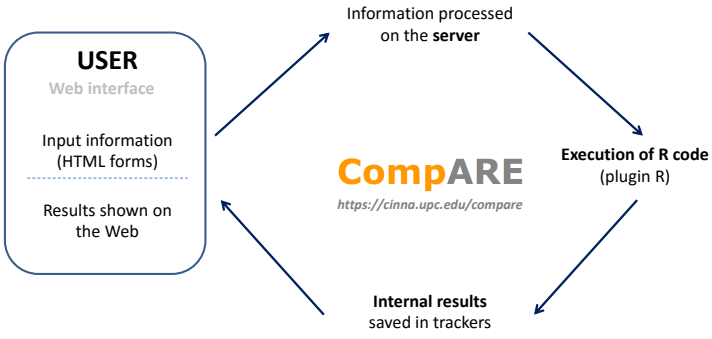

Figure 2: General scheme of *CompARE*.

CompARE step by step

Access and registration

You can access *CompARE* by means of any standard web browser such as Mozilla Firefox, Internet Explorer or Google Chrome by clicking on the following link: https://cinna.upc.edu/compare.

Only a quick registration is needed from the main web page. The system asks you for a username and a password, which will be used to enter the application under your own session. For security reasons, an e-mail is required. You will have to accept the registration from your own e-mail. In order to avoid spam registrations, you need to correctly introduce a captcha code (Completely Automated Public Turing test to tell Computers and Humans Apart). The web administrator accepts the registration as a final step.

Running *CompARE*. Example from the LIFE study

In this section we show the basic capabilities of *CompARE* by means of the LIFE study that has been introduced in the Background. Advanced options are described afterwards in another section. We assume that we have a set of candidate endpoints together with a range of anticipated values for the event probabilities and hazard ratios for the treatment effect.

Following the LIFE study data, we introduce the corresponding parameter values by means of an input grid, programed in Tiki as an entry form (see Figure 3):

1. In the first column, we provide the name of each endpoint. We indicate whether each endpoint is terminating (i.e., when the occurrence of it precludes the observation of other endpoints). In our example, it corresponds to CV death. Place the cursor over each header as an aid for getting a quick definition of each concept.

2. We specify the expected probabilities of observing the event in the control group during the follow-up period. By default, when the relevant or the additional endpoints consist of several components, *CompARE* will use their maximum probability to calculate the ARE values (see advanced options to choose other values). In our example, the RE is a combination of CV death + myocardial infarction, with a frequency of observation in the control group of 5%. The corresponding probability for stroke is 7%.

3. We indicate the anticipated treatment effects in terms of the hazard ratios between groups. By default, when the relevant or the additional

endpoints consist of several components, *CompARE* will use the average hazard ratio to calculate the ARE values. We assume a hazard ratio of the RE and stroke of 0.825 and 0.75, respectively.

4. We select whether the endpoint is a component of the relevant or the additional endpoint.

5. In the last column, we select those candidates that form the composite endpoint.

6. At the bottom, one can click on "Remove executions history" to delete previous analyses that may have been done before.

7. The "Run" button executes the process.

By default, exponential distributions for the marginal times with moderate correlations are assumed.

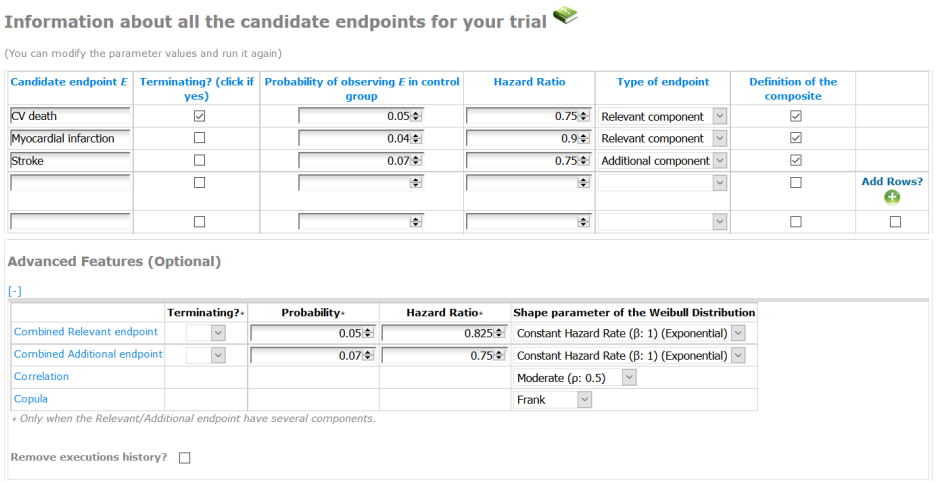

Figure 3: Input grid of information for each endpoint of the trial.

Results

Once the program executes the computations, an output screen is shown, which is divided into the following five tags: results, other scenarios, graphical outputs, recorded results and sample size:

- **Results**: In this tag, a table specifies the parameter information set by the user together with the exact value of ARE (see Figure 4). In our example, the ARE value is 3.49. Note that in this case, the use of the composite endpoint is clearly advisable, since the ARE is higher than 1. That is, we would need a more than 3 times larger sample size if we do not include stroke in the primary endpoint. A paragraph below the table shows a detailed recommendation written in text.

- **Other scenarios**: Several scenarios depending on different correlations and hazard ratios for stroke are detailed in a table (See Figure 5).

- **Graphical outputs**: A plot with survival distributions and hazard ratios is shown at the bottom and the results from the previous tag are shown graphically at the top (see Figure 6). Note that in the LIFE study, the decision remains the same irrespective of the correlation. However, if the expected hazard ratio for stroke would have been 0.9 instead of 0.75 (i.e., a smaller expected effect on the AE), the use of CV death + MI would have been recommended as primary endpoint if furthermore the two marginal times were strongly correlated (Gómez and Gómez-Mateu (2016)[16]).

- **Recorded results**: In this tag, the user can see a table which summarizes the results performed in previous analyses (See Figure 7).

- **Sample size**: Given a significance level of 5% and a statistical power of 80%, the required sample size using CV death + MI is 14,617 patients (see Figure 8). If we add stroke to the PE, the required sample size shrinks to 4,190. In the next subsection, we discuss the required sample size for a variety of scenarios based on different anticipated parameter values.

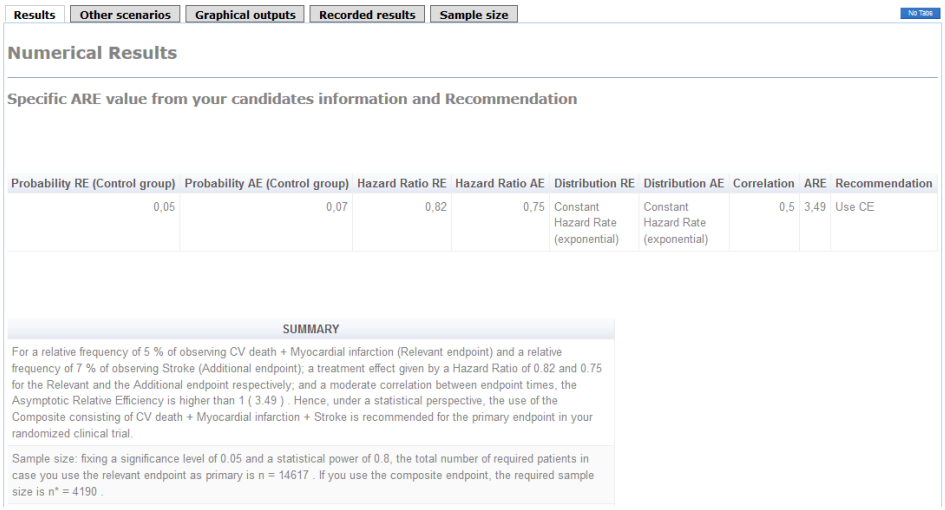

Figure 4: Summary of results in *CompARE*.

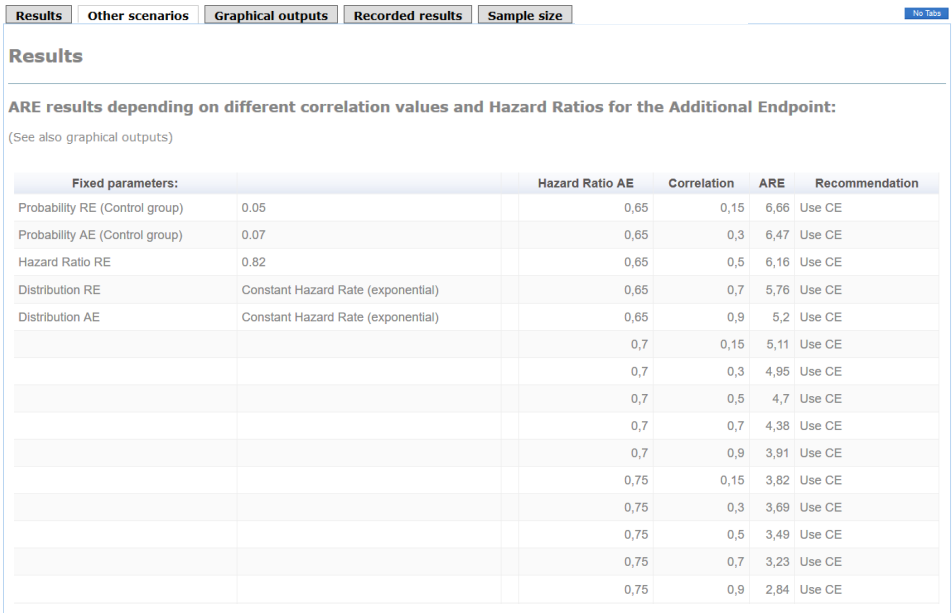

Figure 5: Numerical results for different scenarios.

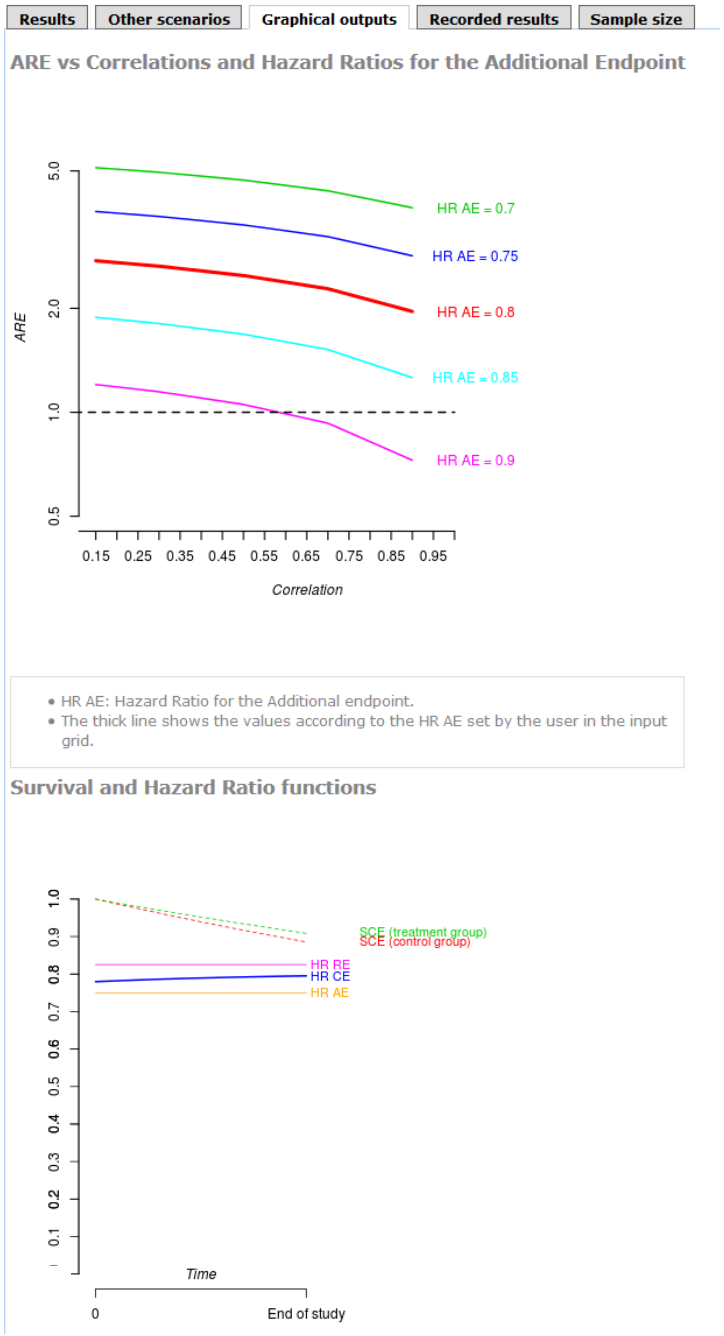

Figure 6: Graphical results in *CompARE* combining different parameter values. The "end of study" time point corresponds to the follow-up period of the clinical trial.

Execution	Relevant	endpoint	components	Additional	endpoint	components	Probability RE (Control group)	Probability AE (Control group)	Hazard Ratio RE	Hazard Ratio AE	Distribution RE	Distribution AE	Correlation	ARE	Recommendation
1	CV death	Myocardial infarction		Stroke			0,05	0,07	0,825	0,75	Constant Hazard Rate (exponential)	Constant Hazard Rate (exponential)	0,15	3,82	Use CE
2	CV death	Myocardial infarction		Stroke			0,05	0,07	0,825	0,75	Constant Hazard Rate (exponential)	Constant Hazard Rate (exponential)	0,5	3,49	Use CE
3	CV death	Myocardial infarction		Stroke			0,05	0,07	0,825	0,75	Constant Hazard Rate (exponential)	Constant Hazard Rate (exponential)	0,9	2,84	Use CE

Figure 7: History of previous computations.

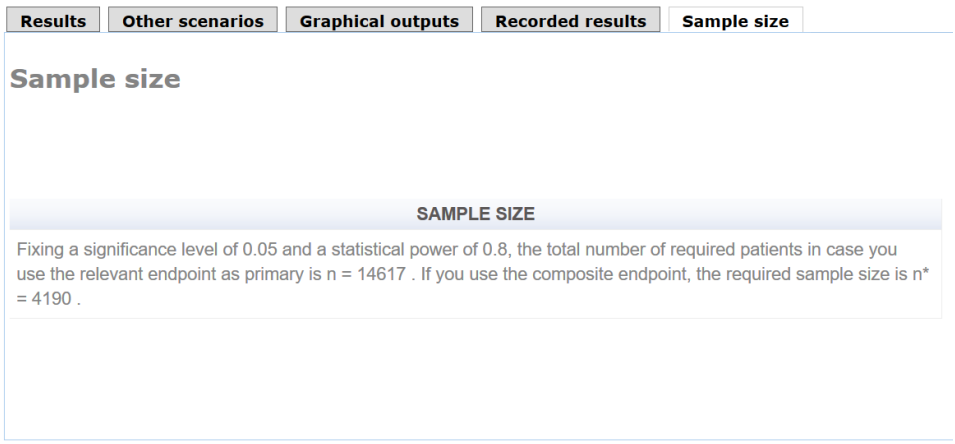

Figure 8: Required sample size computations.

Sample size

We briefly introduce the sample size formulae implemented in *CompARE* and use the LIFE trial to illustrate it. The main original feature is that the sample size of the CE, even when the hazards are not proportional, can be calculated based on the ARE values.

Sample size for the relevant endpoint

For time-to-event endpoints, the power of the test depends on the number of events rather than on the sample size. Based on the asymptotic behaviour of the logrank statistic, Schoenfeld's formula (1981)[17] for the required number of events $e_{_R}$ (one-tailed test) is as follows:

$$e_{_R} = \frac{(z_\alpha + z_\beta)^2}{(\ln(HR_R))^2 \Pi(1-\Pi)},$$

where z_α and z_β are the standard normal quantiles corresponding to the left tail probability for an α-significance level and a β-type II error, respectively. Π is the proportion of patients allocated to the control group. The sample size n_R is as follows[18]:

$$n_R = \frac{e_R}{[p_R^{(0)}\Pi + p_R^{(1)}(1-\Pi)](1-W)},$$

where $p_R^{(0)}$, $p_R^{(1)}$ are the probability of observing the relevant endpoint in the control and treatment group, respectively, for a fixed follow-up period. W is the anticipated withdrawal proportion due to loss to follow-up. For equal allocation between groups $\Pi = 1/2$, and without loss to follow-up $W = 0$, the total number of required patients n_R is given by:

$$n_R = \frac{2e_{_R}}{p_R^{(0)} + p_R^{(1)}}. \tag{1}$$

Sample size for the composite endpoint

In Gómez and Gómez-Mateu [7], it is shown that the asymptotic relative efficiency ARE of the logrank statistic for the CE versus the logrank statistic for the RE can be interpreted, for a given significance level α and power $(1-\beta)$, as the ratio of the required sample sizes for the relevant endpoint (n_R) and for the composite endpoint (n_*); that is, $ARE = n_R/n_*$. It follows that for marginal and constant hazard ratio HR_R and under the other anticipated parameters needed to compute the ARE and described in the Background section, the required number of patients if using the CE is given by:

$$n_* = \frac{n_R}{ARE}. \tag{2}$$

Note that n_R can be computed from (1). We address the reader to the Appendix for the specific expression of the ARE.

Whenever the hazard ratio of the CE, $HR_*(t)$, is approximately constant, one could straightforwardly use Schoenfeld's formulas to calculate the sample size n_*. If the $HR_*(t)$ is not constant, which usually occurs in practice (Gómez-Mateu, 2016)[19] (see Figure 9), *CompARE* computes the sample size depending on different parameter values, based on the ARE approach by means of the aforementioned formulas.

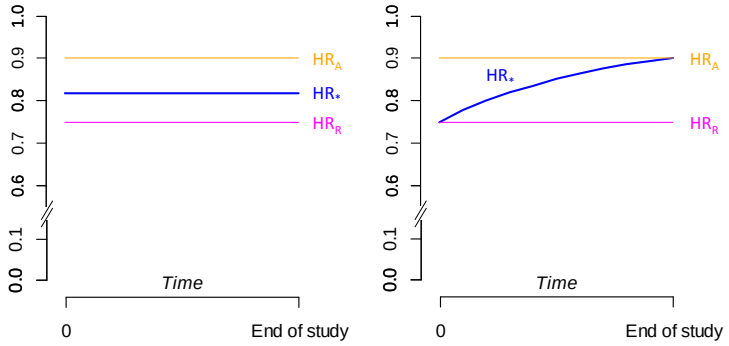

Figure 9: Marginal hazard ratios $HR_R = 0.75$, $HR_A = 0.9$ for the relevant (RE) and additional endpoint (AE), respectively, and hazard ratio HR_* for the CE over time. We fix the probabilities $p_R^{(0)} = 0.05$, $p_A^{(0)} = 0.04$ of observing the RE and the AE in the control group, respectively, with constant hazards for the time to the RE and the AE, with null correlation ($\rho = 0$) (left), and increasing hazards for the time to the AE with hight correlation ($\rho = 0.9$) (right).

Sample size example from the LIFE study

We illustrate the required sample size for the CE (CV death + myocardial infarction + stroke) in the LIFE trial depending on several parameter values. We are using the parameter values of this study, namely the probability $p_R^{(0)} = 0.05$ of observing the relevant endpoint RE (CV death + myocardial infarction) in the control group (Atenolol), and the hazard ratio $HR_R = 0.82$ for the RE. The probability $p_A^{(0)}$ of observing the additional endpoint AE (stroke) in the control group and the hazard ratio HR_A for the AE are 0.07 and 0.75, respectively. All the computations assume a significance level $\alpha = 0.05$ and a statistical power $(1 - \beta)$ equal to 80%.

The required sample size, n_R, for CV death + MI using formula (1) is equal to 14,617 patients. The required sample size n_* following equation (2) would be 5150, 4190 and 3831 for high ($\rho = 0.9$), moderate ($\rho = 0.5$) or low correlation ($\rho = 0.15$) between the times to (CV death + MI) and stroke, respectively (see Table 1). However, if we approximate $HR_*(t)$ by the constant average between HR_R and HR_A, that is $HR_* = 0.79$, the sample sizes n'_* according to the above correlations are 4572, 4213 and 4119, respectively. This shows the relevance of taking into account the degree of association as well as the hazard ratio values for the computation of the sample size.

Note that when the relationship between CV death + MI and stroke is weak ($\rho = 0.15$), the required number of patients n_* using the ARE method is 3831, while if we assume constant hazard ratio (HR) it would be 4119 patients. In contrast, whenever we have strong correlations ($\rho = 0.9$), the sample size n_* using the ARE method is 5150, while the constant HR would imply a smaller sample size ($n'_* = 4572 < n_* = 5150$).

Table 1: Sample sizes n_R, n_* and n'_* for the relevant endpoint RE, composite endpoint CE with ARE, and for CE approximating the hazard ratio $HR_* = 0.79$ by the average between the marginal hazard ratios $HR_R = 0.82$ and $HR_A = 0.75$, respectively. The probabilities of observing the RE and AE in treatment group $p_R^{(1)} = 0.04$ and $p_A^{(1)}$ (which varies with ρ), respectively. $p_*^{(0)}$ and $p_*^{(1)}$ stand for the probability of observing the CE in the control and treatment group, respectively.

n_R	ARE	n_*	n'_*	Correlation ρ	$p_*^{(0)}$	$p_*^{(1)}$
14617	3.82	3831	4119	0.15	0.12	0.09
14617	3.49	4190	4213	0.5	0.11	0.09
14617	2.84	5150	4572	0.9	0.10	0.08

Advanced options in *CompARE*

CompARE is extended to accommodate the comparison of different scenarios coming from a combination of different parameter values, distinct than the given by default, that could represent other realistic situations for the design of a clinical trial. First, the user can choose different marginal laws for the time to each endpoint as well as different degrees of correlation and different copulas. Second, *CompARE* is also extended to quantify specific values for the combined probability and hazard ratios whenever it comes from a combination of several components. When the user cannot anticipate some of the needed parameters, *CompARE* provides a range of plausible values.

Marginal distributions

We have extended *CompARE* to allow Weibull distributions. The density and survival functions for each group j ($j = 0, 1$) used in *CompARE* are parametrized as follows:

$$\begin{aligned} f^{(j)}(t) &= \frac{\beta}{(b^{(j)})^{\beta}} t^{\beta-1} e^{(-(t/b^{(j)})^{\beta})} \\ S^{(j)}(t) &= e^{(-(t/b^{(j)})^{\beta})}, \end{aligned}$$

where the shape parameter β for each group are assumed equal so that constant hazard ratios' assumption holds, and the scale parameters $b^{(j)}$ are derived automatically taking into account all the anticipated parameter values.

From the drop-down menu in the advanced features box, the user can choose between the following options (see Figure 3, bottom):

- Weibull distribution with decreasing hazard rate ($\beta = 0.5$),
- Weibull with constant hazard rate ($\beta = 1$) (exponential distribution),
- Weibull distribution with increasing hazard rate ($\beta = 2$).

Correlation

As we have seen previously in our example, the correlation might play a crucial role. Trialist might need to compare the efficiency of one endpoint over the other in terms of the strength of the association between the times to the relevant endpoint and the additional endpoint. Therefore, *CompARE* incorporates the possibility to change the Spearman's correlation values (see Figure 3, bottom).

Copulas

An additional feature of *CompARE* is the possibility to change the copula used to build the bivariate distribution between the marginal times (see Figure 10). In this sense, although the values of ARE might not be exactly the same depending on the chosen copula, it has been proved that the concordance in the decision on the primary endpoint is markedly high in the majority of cases (Plana-Ripoll and Gómez, 2015)[20].

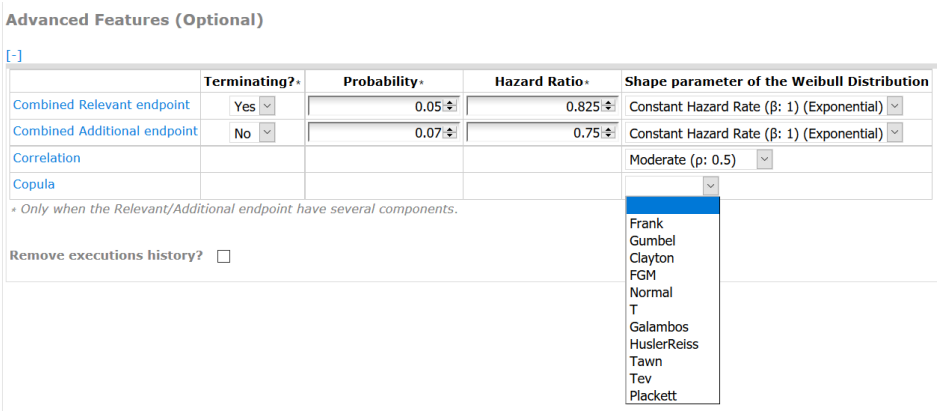

Figure 10: Menu indicating different choices of copulas.

Combined HR and probabilities

CompARE is also extended to accommodate the computation of combined probabilities and combined hazard ratios based on the marginal components. By default, *CompARE* uses the maximum probability and the average hazard ratio to calculate the ARE values. Alternatively, *CompARE* proposes a range of plausible values of the corresponding combined probability and combined hazard ratio (for the latter case, only implemented for two components so far). Based on this range of values, the user can introduce a specific value for the combined parameters by means of the advanced features box (see Figure 3, bottom). In order to visually evaluate the departure from constancy that the combined hazard ratio might have, *CompARE* depicts the shape of the hazard ratio over time.

Discussion

The development of computational tools in biomedical research and clinical trial studies is important to design more efficient trials. There exist several software tools like EAST[21], nQuery[22], and PASS[23] that allows to calculate sample sizes for specific endpoints, and therefore to study the most efficient choice, but none includes the study of the efficiency of a composite endpoint neither the calculation of the corresponding sample sizes whenever the hazards are not proportional.

Our new web-based tool *CompARE* implements the ARE method and allows to study the efficiency of each endpoint for different parameter value combinations, even when the hazard ratio for the composite outcome is time dependent. This tool performs complex computations based on the ARE expression and it is crucial to facilitate the results to trialists without programming skills.

CompARE is extended with new functionalities and features, and we are constantly improving it thanks to the feedback of colleagues from different universities, institutions and companies. For example, we are currently implementing sample size calculations based on Freedman's approach[24], considering different accrual rates and including withdrawals. We are as well working on extending *CompARE* to binary outcomes and to observational studies (Gómez et al., 2016)[25].

Conclusion

The free web-based platform *CompARE* is an exploratory tool that helps trialists to make informed decisions on the primary endpoint of a clinical trial. It allows to study how efficient might be a single endpoint versus a combination of several outcomes. This decision would entail the option with fewer numbers of patients in each treatment arm at the design stage of the study. We remark that most of the cases, the hazard ratio of the composite endpoint is not constant. With this computational resource, under some assumptions, researchers can calculate sample sizes even when the proportionality of the hazards' assumption does not hold, and visually study its departure over time.

CompARE performs the computations by means of a user-friendly graphic interface, and no programming skills are required. Results are shown in written form, detailed numerically with tables and depicted in plots for multiple scenarios.

Abbreviations

AE: Additional endpoint; ARE: Asymptotic relative efficiency; CE: Composite endpoint; CV: Cardiovascular; LGPL: Lesser general public license; MI: Myocardial infarction; PE: Primary endpoint; RE: Relevant endpoint; RCT: Randomized clinical trial

Funding

This work is partially supported by grants MTM2015-64465-C2-1-R (MINECO/ FEDER) from the Ministerio de Economía y Competitividad (Spain), and 2014 SGR 464 (GRBIO) from the Departament d'Economia i Coneixement de la Generalitat de Catalunya (Spain).

Availability of data and material

CompARE is available at: https://cinna.upc.edu/compare

The R code to compute the Asymptotic Relative Efficiency (ARE) values is available at: https://github.com/moisesgomezmateu/compare

References

[1] Ferreira-González, I., Busse, J., Heels-Ansdell, D., Montori, V., Akl, E., Bryant, D., Alonso-Coello, P., Alonso, J., Worster, A., Upadhye, S., Jaeschke, R., Schünemann, H., Permanyer-Miralda, G., Pacheco-Huergo, V., Domingo-Salvany, A., Wu, P., Mills, E., Guyatt, G.: Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. BMJ, 334–7597786 (2007). doi:10.1136/bmj.39136.682083.AE

[2] Freemantle, N., Calvert, M.: Weighing the pros and cos for composite outcomes in clinical trials. J Clin Epidemiol **60**, 658–659 (2007). doi:10.1016/j.jclinepi.2006.10.024

[3] Tomlinson, G., Detsky, A.: Composite end points in randomized trials: there is no free lunch. JAMA **303**, 267–8 (2010). doi:10.1001/jama.2009.2017

[4] Rauch, G., Kieser, M.: An expected power approach for the assessment of composite endpoints and their components. Comput Statist Data Anal **60**, 111–122 (2013). doi:10.1016/j.csda.2012.11.001

[5] Song, J.: Sample size for simultaneous testing of rate differences in non-inferiority trials with multiple endpoints. Comput Statist Data Anal **53**, 1201–1207 (2009). doi:10.1016/j.csda.2008.10.028

[6] Gómez, G., Lagakos, S.: Statistical considerations when using a composite endpoint for comparing treatment groups. Stat Med **32**, 719–738 (2013). doi:10.1002/sim.5547

[7] Gómez, G., Gómez-Mateu, M.: The asymptotic relative efficiency and the ratio of sample sizes when testing two different null hypotheses. SORT **38**, 73–88 (2014)

[8] The R project for statistical computing. https://www.r-project.org

[9] Dahlöf, B., Devereux, R., Kjeldsen, S., Julius, S., Beevers, G., de Faire, U., Fyhrquist, F., Ibsen, H., Kristiansson, K., Lederballe-Pedersen, O., Lindholm, L., Nieminen, M., Omvik, P., Oparil, S., Wedel, H.: Cardiovascular morbidity and mortality in the losartan intervention for endpoint reduction in hypertension study (life): a randomised trial against atenolol. Lancet **359**, 995–1003 (2002). doi:10.1016/S0140-6736(02)08089-3

[10] Sankoh, A., Li, H., D'Agostino, R.: Use of composite endpoints in clinical trials. Stat Med **33**, 4709–4714 (2014). doi:10.1002/sim.6205

[11] Tiki wiki cms groupware. http://info.tiki.org

[12] Free software foundation (FSF). https://www.fsf.org

[13] de Pedro, X., Sánchez, A.: Usando de forma segura r vía web con tiki. II Jornadas de Usuarios de R en Castellano. Mieres, Oviedo (2010)

[14] Sapir, R.: Tiki essentials. what every smarty needs to know about tiki wiki cms groupware. Under a Creative Commons Attribution-Share Alike 3.0 License

[15] Tiki documentation: PluginR. https://doc.tiki.org/pluginr

[16] Gómez, G., Gómez-Mateu, M.: Comments on "Use of composite endpoints in clinical trials" by Abdul J. Sankoh, Haihong Li and Ralph B. D'Agostino, Sr. Stat Med **35**, 317–318 (2016). doi:10.1002/sim.6483

[17] Schoenfeld, D.: The asymptotic properties of nonparametric tests for comparing survival distributions. Biometrika **68**, 316–319 (1981). doi:10.1093/biomet/68.1.316

[18] Machin, D., Campbell, M., Fayers, P., Pinol, A.: Sample size tables for clinical studies. Blackwell Science (1997) (1997)

[19] Gómez-Mateu, M.: Composite endpoints in clinical trials: Computational tools, practical guidelines and methodological extensions. PhD Thesis, Universitat Politècnica de Catalunya, Spain (2016). Advisors: G. Gómez G, U. Dafni. http://www.tdx.cat/handle/10803/396263

[20] Plana-Ripoll, O., Gómez, G.: Selecting the primary endpoint in a randomized clinical trial. J Biopharm Stat **23**, 1–19 (2015). doi:10.1080/10543406.2015.1094808

[21] East 6.3 (Cytel). http://www.cytel.com/software-solutions/east/base

[22] nQuery (Statsols). http://www.statsols.com/nquery-sample-size-calculator

[23] PASS 14 (NCSS). https://www.ncss.com/software/pass

[24] Freedman, L.: Tables of the number of patients required in clinical trials using the logrank test. Stat Med **1**, 121–129 (1982)

[25] Gómez, G., Plana-Ripoll, O., Dafni, U.: Selection of the primary end point in an observational cohort study. J Epidemiol Community Health **70**, 950–953 (2016). doi:10.1136/jech-2015-206656

Appendix

Computation of the Asymptotic Relative Efficiency (ARE)

Consider that the effect of treatment is to be evaluated on the time $T_R^{(j)}$ to a relevant event $\mathcal{E}_R$, where the superscript j indicates the treatment group ($j = 0$ for the control group and $j = 1$ for the treatment group). Assume now that an additional endpoint $\mathcal{E}_A$ is considered as a component of the primary endpoint and that the composite endpoint $\mathcal{E}_* = \mathcal{E}_R \cup \mathcal{E}_A$ is to be used instead in order to prove the efficacy of the new treatment. The effect of treatment would then be evaluated on the time $T_*^{(j)}$ to $\mathcal{E}_*$, where $T_*^{(j)} = \min\{T_R^{(j)}, T_A^{(j)}\}$, and $T_A^{(j)}$ stands for the time to $\mathcal{E}_A$ for each group.

Let $\lambda_k^{(j)}(t)(k = R, A)$ denote the marginal hazard for the relevant endpoint ($k = R$) or the additional endpoint ($k = A$) when $\mathcal{E}_A$ does not include a terminating event. Let $\lambda_{Ck}^{(j)}(t)(k = R, A)$ denote the cause-specific hazard for the relevant endpoint ($k = R$) or the additional endpoint ($k = A$) when $\mathcal{E}_A$ includes a terminating event.

The relative treatment effects on $\mathcal{E}_R$ and on $\mathcal{E}_A$ are given by the hazard ratios $HR_R(t) = \widetilde{\lambda}_R^{(1)}(t)/\widetilde{\lambda}_R^{(0)}(t)$ and $HR_A(t) = \widetilde{\lambda}_A^{(1)}(t)/\widetilde{\lambda}_A^{(0)}(t)$, respectively, where $\widetilde{\lambda}$ stands indistinctly for λ or λ_{Ck}, which we assume that they are constant over time. That is, $HR_R(t) = HR_R$ and $HR_A(t) = HR_A$. Let $HR_*(t)$ denote the hazard ratio for the composite endpoint.

Whenever the additional endpoint does not include a terminating event, the asymptotic relative efficiency $ARE(Z_*, Z_R)$ of the logrank test (Z_*) based on $\mathcal{E}_*$ versus the logrank test (Z_R) based on $\mathcal{E}_R$ is given by:

$$ARE(Z_*, Z_R) = \frac{\left(\int_0^1 \log\left\{HR_*(t) f_*^{(0)}(t)dt\right\}\right)^2}{\left(\log\left\{HR_R\right\}\right)^2 \left(\int_0^1 f_*^{(0)}(t)dt\right)\left(\int_0^1 f_R^{(0)}(t)dt\right)},$$

and when the additional endpoint includes a terminating event,

$$ARE(Z_*, Z_R) = \frac{\left(\int_0^1 \log\left\{\frac{HR_R\lambda_{CR}^{(0)}(t) + HR_A\lambda_{CA}^{(0)}(t)}{\lambda_{CR}^{(0)}(t) + \lambda_{CA}^{(0)}(t)}\right\} f_*^{(0)}(t)dt\right)^2}{(\log\{HR_R\})^2 \left(\int_0^1 f_*^{(0)}(t)dt\right) V},$$

being

$$V = \int_0^1 \frac{e^{-HR_A \int_0^t \lambda_{CA}^{(0)}(u)du} S_*^{(0)}(t)\lambda_{CR}^{(0)}(t)}{e^{-\int_0^t \lambda_{CA}^{(0)}(u)du}\pi + e^{-HR_A \int_0^t \lambda_{CA}^{(0)}(u)du}(1-\pi)} dt,$$

where $f_R^{(0)}(t)$ and $f_*^{(0)}(t)$ are the density functions of $T_R^{(0)}$ and $T_*^{(0)}$, respectively; $S_*^{(0)}(t)$ stands for the survival function of $T_*^{(0)}$; and π stands for the null probability of being in group 1. We refer to the original paper ([6]) for further details.