

SELECTING THE PRIMARY ENDPOINT IN A RANDOMIZED CLINICAL TRIAL. THE ARE METHOD

OLEGUER PLANA-RIPOLL^{*1} AND GUADALUPE GÓMEZ²

ABSTRACT

The decision on the primary endpoint in a randomized clinical trial is of paramount importance and the combination of several endpoints might be a reasonable choice. Gómez and Lagakos (2013) have developed a method that quantifies how **much** more efficient **it** could be to use a composite instead of an individual relevant endpoint. From the information provided by the frequencies of observing the component endpoints in the control group and by the relative treatment effects on each individual endpoint, the Asymptotic Relative Efficiency (ARE) can be computed. This paper presents the applicability of the ARE method as a practical and objective tool to evaluate which components, among the plausible ones, are more efficient in the construction of the primary endpoint. The method is illustrated with two real cardiovascular clinical trials and is extended to allow for different dependence structures between the times to the individual endpoints. The influence of this choice on the recommendation on whether or not to use the composite endpoint as the primary endpoint for the investigation is studied. We conclude that the recommendation between using the composite or the relevant endpoint only depends on the frequencies of the endpoints and the relative effects of the treatment.

Key words: Asymptotic relative efficiency; Composite endpoint; Copulas; Logrank; Randomized clinical trial

ACKNOWLEDGEMENTS

This work is partially supported by grant MTM2012-38067-C02-01 from the Ministerio de Economía y Competitividad del Gobierno de España and 2014 SGR 464 from the Departament d'Economia i Coneixement de la Generalitat de Catalunya. The authors declare that they have no conflict of interest. We would like to thank Moisés Gómez-Mateu for programming a first version of the R code using Frank's copula and Dr. Martin Mächler for his help with the package `copula`. We are indebted to the GRASS group (<http://grass.upc.edu>) for fruitful discussions.

¹ Section of Epidemiology, Department of Public Health, Aarhus University, Aarhus, Denmark

² Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya, Barcelona, Spain.

* Corresponding author. E-mail: opr@ph.au.dk, Phone: +45 87167949

1 INTRODUCTION

When comparing two treatment groups by means of a randomized clinical trial (RCT), the choice of the primary endpoint is crucial. It is often the case in which several relevant events might be chosen as the primary endpoint for the analysis and the decision on which one to choose might be difficult. Sometimes, the event with the greatest clinical importance is the chosen one while the other events are assessed using secondary analysis. In other situations, two or more events are of comparable importance and, in those cases, it is common to use the union of them as the primary endpoint. In general, the decision on which endpoint to use is, among other criteria, based on the prior knowledge of the frequency of observing the “candidate” events as well as on the anticipated effect that the treatment could have on each event. However, in some occasions the choice of endpoints is controversial and the final decision is not taken on solid grounds.

Several authors have discussed the advantages and disadvantages of using a composite endpoint from a clinical perspective (Ferreira-González et al., 2007b; Freemantle et al., 2003; Montori et al., 2007; Ferreira-González et al., 2007a). One of the main clinical arguments is that the composite endpoint could capture the net benefit of the intervention, avoiding the need to choose a single main endpoint when different endpoints are of equal importance. From a statistical point of view, two main arguments are: i) a composite endpoint reduces the multiplicity problem that may occur if different endpoints are used separately (Huque et al., 2011); and ii) a larger number of events will be observed when using a composite endpoint than for any relevant event alone. However, as it is shown in Gómez and Lagakos (2013), a larger number of events would not necessarily imply a more powerful test for treatment efficacy. Moreover, the use of a composite endpoint could entail certain difficulties such as how to interpret the results of a trial when the primary endpoint is composed of events of different clinical importance (Song et al., 2008). It is also important to keep in mind that a significant treatment effect on the composite endpoint does not necessarily imply an effect on each of the components (Gómez, 2011). Song et al. (2008) clarify some of these issues and propose to use a recurrent composite endpoint.

Gómez and Lagakos (2013) frame this problem by considering two potential endpoints to answer the most interesting question for the investigators. We refer to these two endpoints as relevant and additional endpoints, corresponding, in many instances, to two feasible primary endpoints or a primary and a secondary endpoint. Gómez and Lagakos have provided a statistical methodology, we refer to it as ARE method, to derive efficiency guidelines for deciding whether to expand a study’s relevant endpoint to the composite endpoint defined as the event that has occurred if either the relevant or the additional endpoint is observed (Meinert, 2012). The method they propose is suitable for randomized clinical trials in which the two treatments are compared with respect to the time from randomization until the event occurs and when a logrank test is used for the comparison. The decision on whether to design the clinical trial using the relevant or the combination of the relevant and the additional is based on the Asymptotic Relative Efficiency (ARE). The ARE is to be interpreted as the ratio of the needed sample sizes to attain the same power for a given significance level when using either the relevant or the composite endpoint (Gómez and Gómez-Mateu, 2014). The ARE method provides a tool that allows to make a more informed decision on the choice of the primary endpoint of a RCT during its design and when the data is still not available. Keeping this in mind, the computation of the ARE is based on easily interpretable parameters, possibly obtained from relevant previous studies, such as the frequencies of observing the relevant and additional endpoints in the control group, the anticipated relative treatment effects on the relevant and additional endpoints given by the corresponding hazard ratios and to a lesser extent by the dependence degree between the relevant and the additional endpoints. Furthermore, the marginals and the joint law between the times from randomization until the relevant and additional events occur have to be defined and, in doing so, we face the challenge of modelling an empirical problem in such a way that is not too complex but still realistic. Gómez and Lagakos model the joint dependence structure by means of a copula function. A copula is best described, as in Joe (1997), as a multivariate distribution function that is used to bind each marginal distribution function to form the joint. The copula parameterises the dependence between the marginals, while the parameters of each marginal distribution function can be estimated separately.

Among the many clinical trials that use composite endpoints, we have chosen the studies by Packer et al. (2001) and Tardif et al. (2008) to illustrate the ARE method. The goal of Packer’s study is to test

the effect of carvedilol in patients who had symptoms of heart failure at rest or on minimal exertion. The primary endpoint of the study, denoted by \mathcal{E}_1 , was death by any cause. They use as a secondary endpoint, \mathcal{E}_* , the endpoint that combines death and hospitalization for any reason. We illustrate in Subsection 3.2 how the most efficient of these two endpoints could have been chosen ahead of time based on some parameters of the same nature of those needed for the computation of the sample size. Tardif's study evaluates the efficacy of the addition of succinobucol, a novel anti-oxidant and anti-inflammatory agent, to optimal medical therapy. They conduct a double-blind, placebo-controlled clinical trial to 6,144 high-risk patients with unstable angina or who had suffered heart attacks. The following six cardiovascular events are of interest: Cardiovascular death, resuscitated cardiac arrest, myocardial infarction, stroke, hospitalization due to unstable angina or hospitalization due to coronary revascularization. The study shows that succinobucol has no effect on the primary endpoint, \mathcal{E}_* , where all six events are considered, while it has a beneficial effect on the composite secondary endpoint defined as the union of the first 4 events. In this particular instance, the addition of the hospitalization events (355 (67%) in the succinobucol group and 318 (60%) in the placebo group) to the previous 4 events (207 versus 252) has yielded a non significant result for the primary \mathcal{E}_* from a beneficial effect that the treatment has on the composite secondary endpoint. **This study is a good example to show that had the ARE method been applied during the design of the trial, then a significant clinical benefit of the drug would have been found on the reduction of the composite endpoint of cardiovascular death, resuscitated cardiac arrest, myocardial infarction and stroke.**

The main aim of this paper is to present the applicability of the ARE method. This method can be used as a practical and objective tool to evaluate which components, among the plausible ones, are more efficient in the construction of the primary endpoint. This tool allows clinicians and statisticians to take a more informed decision for the final primary endpoint in a randomized clinical trial. We first present the methodology and show how to use it with the aid of the software R. As we will see, to use this method the investigators will have to provide values for the frequencies of observing the candidate endpoints in the control group as well as the range of values for the anticipated hazard ratios. Furthermore, although the choice of a copula to model the joint behavior between the two candidate endpoints is needed, we have run a study for several families of copulas, and concluded that the recommendations for the most efficient primary endpoint are independent of the copula chosen.

The remainder of this paper is organized as follows. In Section 2, we present the ARE method and show how to use it in terms of interpretable parameters. In Section 3, we introduce the above described clinical trials to motivate the problem and to show how to use this method. In Section 4 we extend the method to several families of copulas and study the sensitivity of the ARE values to the choice of the copula. Simulated scenarios have been set up and the ARE values have been compared using different copulas. Finally, we make the concluding remarks in Section 5.

2 THE ARE METHOD

2.1 The theoretical framework

Consider a two-arm randomized study with assignment to an active ($X = 1$) or control treatment ($X = 0$), for example new treatment versus standard of care or placebo. Assume that several individual endpoints are feasible to check the efficacy of the treatment and, although not all of them have to be of equal importance, neither all of them are expected to have the same benefit. For the sake of clarity denote by \mathcal{E}_1 and \mathcal{E}_2 two potential primary endpoints that could answer satisfactorily the study's primary clinical question and denote by \mathcal{E}_* the corresponding composite endpoint, which is considered to have occurred if either \mathcal{E}_1 or \mathcal{E}_2 is observed (Meinert, 2012). Let T_1 , T_2 and $T_* = \min\{T_1, T_2\}$ be the respective times from randomization to \mathcal{E}_1 , \mathcal{E}_2 and \mathcal{E}_* and assume that the end-of-study is the only non-informative censoring cause.

We could test the superiority of the new treatment either by means of the relevant endpoint \mathcal{E}_1 or using the composite endpoint \mathcal{E}_* . In either case, assume that the logrank test is used for the comparison.

The ARE method is based on the Asymptotic Relative Efficiency between the logrank test based on the observed data provided by the composite endpoint \mathcal{E}_* and the logrank test based on the observed

data provided by the relevant endpoint \mathcal{E}_1 . It can be shown (Gómez and Gómez-Mateu, 2014) that the ARE value corresponds to the ratio of the sample sizes that one might need if using either \mathcal{E}_* or \mathcal{E}_1 to evaluate the efficacy of the treatment, for the same significance level and the same power.

The expression of the ARE depends on whether \mathcal{E}_1 and \mathcal{E}_2 contain a terminating event (*i.e.* death) among its components. **In those cases in which death could be one of the components of \mathcal{E}_2 , then \mathcal{E}_1 will only be observed if occurs before \mathcal{E}_2 , and the computation of the ARE is a bit more complex because it depends on the cause-specific hazards instead of the marginal ones.** The discussion in this paper is restricted to those cases such that the components of the additional endpoint \mathcal{E}_2 do not include death. The reader is addressed to the paper by Gómez and Lagakos (2013) for the technical details and discussion of the assumptions for the 4 different censoring situations.

Assume that the relative treatment effects on \mathcal{E}_1 and on \mathcal{E}_2 given by their hazard ratios, HR_1 and HR_2 , are constant, let θ be a parameter measuring the dependence between T_1 and T_2 in either group (see Appendix A.1 for details) and denote by $f_*^{(0)}(t; \theta)$ the density function of T_* in group o, that is, $\int_0^1 f_*^{(0)}(t; \theta) dt = \text{Prob}(\text{observing } \mathcal{E}_* \text{ in group o})$. The expression of the ARE value is as follows:

$$\text{ARE}(\mathcal{E}_*, \mathcal{E}_1) = \frac{\left(\int_0^1 \log(HR_*(t; \theta)) f_*^{(0)}(t; \theta) dt \right)^2}{(\log HR_1)^2 \text{Prob}(\text{observing } \mathcal{E}_1 \text{ in group o}) \text{Prob}(\text{observing } \mathcal{E}_* \text{ in group o})}, \quad (1)$$

where $HR_*(t; \theta)$ stands for the relative treatment effect on \mathcal{E}_* , that is, the ratio between $\lambda_*^{(1)}(t; \theta)$ and $\lambda_*^{(0)}(t; \theta)$, the hazard functions at time t for the composite endpoint for groups 1 and o, respectively.

2.2 The ARE as ratio of needed sample sizes

Although $HR_*(t; \theta)$ does not have to be constant even if HR_1 and HR_2 are (Gómez, 2011), to have further insight into the ARE meaning, assume momentarily that $HR_*(t; \theta) = HR_*$ is approximately constant. Then, the ARE given in (1) would simplify to

$$\text{ARE}(\mathcal{E}_*, \mathcal{E}_1) = \frac{(\log HR_*)^2 \text{Prob}(\text{observing } \mathcal{E}_* \text{ in group o})}{(\log HR_1)^2 \text{Prob}(\text{observing } \mathcal{E}_1 \text{ in group o})} \quad (2)$$

If we recall sample size formulas (Collett, 1994) to distinguish between two treatment groups in a survival study, the required number e_1 of events \mathcal{E}_1 to declare that the observed hazard ratio is different from 1 with a probability $1 - \beta$ for a specified significance level α is given by

$$e_1 = \frac{4(z_{1-\alpha} + z_{1-\beta})^2}{(\log HR_1)^2}$$

where $\log HR_1$ reflects the magnitude of the treatment difference evaluated if endpoint \mathcal{E}_1 is used that is important to detect, and z_q is the q^{th} quantile of the standard normal distribution. If there were no right censoring, e_1 would be equal to the required sample size. Accounting for right censoring, the sample size would be

$$n_1 = \frac{e_1}{\text{Prob}(\text{observing } \mathcal{E}_1 \text{ in group o})}.$$

Analogous formulas apply for the total number of events \mathcal{E}_* needed to be observed, e_* , and for the corresponding sample size n_* to detect $\log HR_*$. Replacing accordingly $\log HR_1$ and $\log HR_*$ in (2), we have

$$\text{ARE}(\mathcal{E}_*, \mathcal{E}_1) = \frac{\text{Prob}(\text{observing } \mathcal{E}_* \text{ in group o})/e_*}{\text{Prob}(\text{observing } \mathcal{E}_1 \text{ in group o})/e_1},$$

hence $\text{ARE}(\mathcal{E}_*, \mathcal{E}_1) = \frac{n_1}{n_*}$, showing that ARE can be interpreted as the ratio of the required sample sizes for a given power and significance level. Therefore, a value of $\text{ARE}(\mathcal{E}_*, \mathcal{E}_1) > 1$ would favour the composite endpoint \mathcal{E}_* , while if $\text{ARE}(\mathcal{E}_*, \mathcal{E}_1) \leq 1$ the relevant endpoint should be preferred.

2.3 The ARE in terms of interpretable parameters

Since the main aim of the ARE methodology is to provide an anticipated value for the efficiency of using either primary endpoint (the composite \mathcal{E}_* or the relevant \mathcal{E}_1), on which to base a more informed decision to run the clinical trial, it is of uttermost importance to be able to do so in terms of interpretable and anticipatable quantities. That is, the method requires to setup values for certain parameters, which are often based on prior knowledge of the scientific study, of the same nature of what is commonly done for computing the sample size before starting to collect data for a given clinical trial.

Certain choices have to be made in order to compute the ARE. As we see in expression (1), the ARE value depends on the law of T_* , the time to the composite endpoint (through $f_*^{(0)}(t; \theta)$ and $HR_*(t; \theta)$). The behaviour of T_* depends on the joint law between the times T_1 and T_2 to \mathcal{E}_1 and \mathcal{E}_2 , respectively. Gómez and Lagakos did model the joint dependence structure by means of Frank's copula. Frank's copula binds the marginal laws of T_1 and T_2 through a function (see (6) in Appendix A.2) and the aid of a dependence parameter.

However, the election of the copula is debatable and the possibility that the ARE values depend too strongly on this choice has to be checked. In Section 4 we present a computational study for 11 different copula families and show how the recommendation based on the ARE values is, almost always, independent of the copula chosen. Appendix A.1 provides technical details for the general expressions of $f_*^{(0)}(t; \theta)$ and $HR_*(t; \theta)$ in terms of an arbitrary copula function, and Appendix A.2 exemplifies for three most common copulas: Frank, Gumbel and Clayton. Furthermore, since the chosen copula binds the marginals of the endpoints T_1 and T_2 , the marginal laws have to be characterized. We have chosen Weibull's laws because their flexibility allowing decreasing, constant and increasing hazard rates. Finally, the following quantities are required to compute the ARE given in (1):

- frequencies p_1 and p_2 of observing the endpoints \mathcal{E}_1 and \mathcal{E}_2 in treatment group 0,
- relative treatment effects on \mathcal{E}_1 and \mathcal{E}_2 given by the hazard ratios HR_1 and HR_2 ,
- decreasing, constant or increasing hazard rates of the marginal Weibull distributions of the endpoints T_1 and T_2 , and
- Spearman's rank correlation coefficient ρ between T_1 and T_2 (this coefficient is in one-to-one correspondence with the copula dependence parameter θ).

It is worth noting that the set of assumptions established in Gómez and Lagakos (2013) are often assumed when designing a clinical trial. Furthermore, the set of parameters required by the ARE method corresponds to those needed for the computation of the sample size in a trial. Henceforth, neither the assumptions nor the needed anticipated values could be considered as a limitation in the design of this type of studies.

From a pragmatic point of view, we could use Frank's copula and exponential (constant hazard rates) marginals by default, fix the anticipated values (p_1, p_2, HR_1), use the criterium $ARE \leq 1$ and $ARE > 1$ to base the decision between \mathcal{E}_1 and \mathcal{E}_* , respectively, together with plots as in Figure 1 to make decisions for large range of plausible values for HR_2 and ρ . How to use and interpret these plots is postponed to Section 3. Appendix A.3 provides details on the computation of the ARE using R.

3 USING THE ARE METHOD TO CHOOSE THE PRIMARY ENDPOINT

Two randomized clinical trials will serve to illustrate how to use the ARE method in a real setting, to compute the ARE values for several scenarios, to discuss which would have been our recommendation for the primary endpoint and to give insight into the selection of the primary endpoint. The reader is referred to Gómez et al. (2014), where a survey of the use of composite endpoints in the Cardiovascular Literature is described, other interesting cases of clinical trials are presented and a set of recommendations for future design choice between relevant and composite endpoints for cardiovascular clinical trials is discussed.

3.1 Case study 1: *Treating patients after an acute coronary syndrome with succinobucol*

Tardif et al. (2008) designed a randomized clinical trial to test the effect of the antioxidant succinobucol on cardiovascular outcomes in patients with recent acute coronary syndrome. The primary efficacy endpoint, denoted by \mathcal{E}_* , was the union of \mathcal{E}_1 and \mathcal{E}_2 , where \mathcal{E}_1 was the first occurrence between cardiovascular death, resuscitated cardiac arrest, non-fatal myocardial infarction and non-fatal stroke; and \mathcal{E}_2 was the first occurrence of hospitalization due to either unstable angina with objective evidence of ischaemia or coronary revascularisation. The trial did not show statistically significant differences in terms of the primary composite endpoint \mathcal{E}_* obtaining a hazard ratio $HR_* = 1.00$ (95% confidence interval (CI): 0.89 – 1.13, p-value (p)=0.96).

However, the effect of succinobucol in terms of \mathcal{E}_1 yielded a hazard ratio $HR_1 = 0.81$ (95% CI of 0.68 – 0.98, p =0.029). In this trial, there were many more hospitalizations (\mathcal{E}_2) than \mathcal{E}_1 events, but, since the number of hospitalizations were similar in both treatment groups, the effect on the composite endpoint \mathcal{E}_* was diluted. Conclusions from a randomized clinical trial should be only based on the primary endpoint and findings from secondary analyses must be taken cautiously, since they could be easily explained by chance, after inflating the type I error of the study. Although this study does not prove the efficacy of succinobucol on \mathcal{E}_1 , it serves the purpose of illustrating how adding an event \mathcal{E}_2 to a relevant one \mathcal{E}_1 could lead to the disappearance of the potential significant benefit of the active treatment that would have been found on the relevant outcome (\mathcal{E}_1) and exhibits the usefulness of the ARE method.

For the purpose of the following discussion, assume that the expected frequency of observing \mathcal{E}_1 (outcome given by either cardiovascular death, resuscitated cardiac arrest, non-fatal myocardial infarction or non-fatal stroke), in the placebo group, is $p_1 = 0.082$, while the frequency of observing a hospitalization, \mathcal{E}_2 , is $p_2 = 0.09$. If the expected risk reduction of \mathcal{E}_1 is about 20% ($HR_1 = 0.81$), the ARE values for risk reductions of \mathcal{E}_2 between 5% and 70% (HR_2 ranging from 0.3 to 0.95, actual value in the trial was 1.05) are computed for weak to strong correlations between T_1 and T_2 and for increasing, constant and decreasing hazard functions for T_1 and T_2 , yielding a total of 567 different scenarios (Table 1). Whenever the treatment reduces the risk of hospitalization by at most 10% ($HR_2 \geq 0.9$) (126 different scenarios) we observe $ARE < 1$ and in these cases the relevant endpoint \mathcal{E}_1 is always more efficient than adding hospitalisation for unstable angina or coronary revascularisation. However, if succinobucol would reduce the risk of hospitalization by at least 15% ($HR_2 \leq 0.85$) (441 different situations), then $ARE > 1$ and the composite endpoint combining the six individual endpoints \mathcal{E}_* should be preferred as primary efficacy endpoint.

Plots as the one shown in Figure 1 provide an extra tool to decide between \mathcal{E}_1 and \mathcal{E}_* , even when the values of HR_2 or ρ are only known within a range. The plot shows the ARE values (in logarithmic scale so that the significance of relative asymptotic efficiency is faithfully represented) for different Spearman's ρ correlation values and different effects of the treatment on the additional endpoint \mathcal{E}_2 . For those situations with $HR_2 \leq 0.85$, $ARE > 1$ for all ρ values while if $HR_2 \geq 0.9$, $ARE \leq 1$ for almost all ρ values. However, if the risk reduction on hospitalization is between 10% and 15% ($0.85 < HR_2 < 0.9$) then the plot shows $ARE > 1$ for weak to moderate correlations and $ARE \leq 1$ for moderate to strong correlations. Hence, considerations on the degree of association between \mathcal{E}_1 and \mathcal{E}_2 will lead to different recommendations for the definition of the primary endpoint.

We also observe that the ARE values decrease as the degree of association between \mathcal{E}_1 and \mathcal{E}_2 increases, and hence, the recommendation of sticking to the relevant endpoint \mathcal{E}_1 is stronger as \mathcal{E}_1 and \mathcal{E}_2 are more correlated. Figure 1 has been based on Frank's copula and assuming that the marginal hazards are both increasing. Other hazard combinations lead to similar results.

3.2 Case study 2: *Effect of carvedilol on survival in severe chronic heart failure*

Packer et al. (2001) designed a randomized clinical trial to test the effect of carvedilol in patients who had symptoms of heart failure at rest or on minimal exertion. The primary endpoint of the study (\mathcal{E}_1) was death from any cause. One of the secondary endpoints was the combined risk of death (\mathcal{E}_1) and hospitalization for any reason (\mathcal{E}_2) and we will denote it by \mathcal{E}_* . The trial showed statistically significant lower risk for patients treated with carvedilol compared to placebo on death from any cause, \mathcal{E}_1 , with 11.2% and 16.8% of deaths in the treatment and placebo groups, respectively. It was as well shown

that carvedilol reduced the combined number of deaths or hospitalisations, \mathcal{E}_* , with 36.8% and 44.7% in the treatment and placebo groups, respectively.

We use this example to illustrate how to use the ARE method to choose the more efficient endpoint during the design of the trial, based on the values that were assumed to compute the sample size. In [Packer et al. \(2001\)](#) the sample size was estimated anticipating a 28% ($p_1 = 0.28$) one-year mortality in the placebo group and an anticipated risk reduction as a result of the treatment of 20% ($HR_1 = 0.8$) (for a 90% power and two-sided $\alpha = 0.05$). In order to apply the ARE method we need, as well, anticipated values for the probability of one-year hospitalizations in the placebo group and values for the expected risk reductions as a result of the treatment in the number of hospitalizations. We will use three different frequency values for the expected frequency of hospitalization in the placebo group, $p_2 = 0.10, 0.30, 0.50$, and several expected risk reductions ($HR_2 = 0.3$ through 0.95) to cover a large number of scenarios. The ARE value is calculated for the above parameters, for weak to strong correlations between time to death from any cause, T_1 , and time to hospitalization for any reason, T_2 , and for increasing, constant and decreasing hazard functions for T_1 and T_2 , replicating the scenarios from Section 3.1. Table 1 reproduces the 567 different scenarios for each of the three different probabilities of hospitalisation p_2 .

Whenever the effect of the treatment reduces the risk of hospitalizations by at least 30% ($HR_2 \leq 0.7$) and the expected frequency of hospitalisation ranges from 10% to 50%, the ARE value is larger than 1 and, therefore, adding hospitalisation is efficient and the primary endpoint should be the combination of death and hospitalisation. However, if the effect that treatment has on hospitalisation only reduces the risk by as much as 10% ($HR_2 \geq 0.9$), then, sticking to death from any cause as primary endpoint would be a more efficient decision.

When the effect of treatment on the risk of hospitalisations \mathcal{E}_2 is similar to that on the risk of deaths, \mathcal{E}_1 , i.e. between 10% and 30% ($0.7 < HR_2 < 0.9$), then the recommendation on whether to use death as main primary endpoint or add hospitalisations depend on the probability of being hospitalised and the correlation between time to death, T_1 , and time to hospitalization, T_2 . As seen in the plots in Figure 2, the gain in efficiency by adding \mathcal{E}_2 and considering the composite of death and hospitalization increases as the probability of being hospitalised increases. At the same time, the effect of the degree of association between \mathcal{E}_1 and \mathcal{E}_2 becomes less relevant with the increase of p_2 . Figure 2 has been based on Frank's copula and assuming that the marginal hazards are both constant. Other hazard combinations lead to similar results.

4 HOW RELEVANT IS THE CHOICE OF THE COPULA? A COMPUTATIONAL STUDY

The ARE method is a powerful tool created to help in the choice of the primary endpoint during the design of a clinical trial. As it is explained in Subsection 2.3 the election of the copula as a joint model between T_1 and T_2 is a crucial step in the computation of the ARE value.

However, during the design stage, data is not available and the bivariate structure of T_1 and T_2 is unknown. **It is possible to obtain measures of association such as Pearson's correlation of T_1 and T_2 from previous studies. However, a pair of joint distributions with different bivariate structures - given by different copulas - could have similar values in the correlation of the two marginal distributions.** It is, hence, important to explore the impact that different copulas might have in the construction of the bivariate survival function of (T_1, T_2) since the recommendation on whether to use as primary endpoint the relevant endpoint \mathcal{E}_1 or the composite \mathcal{E}_* might heavily depend on this choice. The main purpose of this section is to expand the ARE method to copulas, other than Frank, and to check whether or not different copulas imply fundamental changes in the ARE recommendations.

4.1 Extension of the ARE method to other copulas

The law of the composite endpoint $T_*^{(j)} = \min\{T_1^{(j)}, T_2^{(j)}\}$ depends on the joint law between $T_1^{(j)}$ and $T_2^{(j)}$, where the superscript (j) stands for group $j = 0, 1$. This joint law can be characterized in many different ways, and a bivariate copula C (function from I^2 to $I = [0, 1]$) is a convenient choice. Copulas

can be used to define different dependence structures for pairs of random variables (Trivedi and Zimmer, 2007). In fact, multivariate distributions with similar degrees of dependence might exhibit substantially different dependence structures depending on the copula chosen.

If $S_1^{(j)}(t_1)$ and $S_2^{(j)}(t_2)$ are the marginal survival functions of $T_1^{(j)}$ and $T_2^{(j)}$, respectively, the joint survival function of $(T_1^{(j)}, T_2^{(j)})$ is given by

$$S_{(1,2)}^{(j)}(t_1, t_2; \theta) = S_1^{(j)}(t_1) + S_2^{(j)}(t_2) - 1 + C(1 - S_1^{(j)}(t_1), 1 - S_2^{(j)}(t_2); \theta),$$

where we are using the same copula C and the same association parameter θ for both groups. The survival function of $T_*^{(j)} = \min\{T_1^{(j)}, T_2^{(j)}\}$ is then given by

$$S_*^{(j)}(t; \theta) = \Pr\{T_1^{(j)} > t, T_2^{(j)} > t; \theta\} = S_{(1,2)}^{(j)}(t, t; \theta).$$

We have extended the ARE method to the following 11 families of copulas: Frank, Gumbel, Clayton, Farlie-Gumbel-Morgenstern (FGM), Normal, t , Plackett, Galambos, Hüsler-Reiss, Tawn and t -EV. Table 2 gives the expression for each of these copulas and Appendix A.1 provides further technical details. Since Frank, Gumbel and Clayton copulas are the copula families more often used, we present them with greater detail in Appendix A.2 and discuss the type of dependencies that these 3 families might exhibit.

4.2 Revisiting the succinobucol study for several copulas

In the randomized clinical trial introduced in Subsection 3.1, the first occurrence between cardiovascular death, resuscitated cardiac arrest, non-fatal myocardial infarction and non-fatal stroke (relevant endpoint \mathcal{E}_1) would have been recommended as primary endpoint if the risk for hospitalization of treated patients compared to placebo would have anticipated a reduction of less than 10% ($HR_2 > 0.9$) because, in these cases, as already discussed, the Asymptotic Relative Efficiency is, for all possible correlations and behaviours of the marginal hazards, less than 1.

We have computed the ARE for the remaining ten possible copulas and have found that all would have yield to the same recommendation. Figure 3 exemplifies the same ARE pattern for the eleven different copulas. Our study has considered 504 different scenarios, which are the ones described in Section 3.1 excluding the particular situation in which $HR_2 = 0.85$ – this relative risk value was only chosen to illustrate ARE values close to 1. In almost all the cases (492 scenarios, 97.6%), the same recommendation on whether to use \mathcal{E}_1 or \mathcal{E}_* would be given.

Our conclusions summarizing the findings are the following:

1. Use the composite endpoint \mathcal{E}_* if the treatment reduces by more than 10% the risk of hospitalization \mathcal{E}_2 ($HR_2 < 0.9$) and use the first occurrence between cardiovascular death, resuscitated cardiac arrest, non-fatal myocardial infarction and non-fatal stroke, \mathcal{E}_1 , if the reduction is less than 10% ($HR_2 > 0.9$);
2. If treatment reduces the risk of hospitalization by approximately 10% ($HR_2 = 0.9$) and T_1 and T_2 are moderate or strongly correlated ($\rho > 0.25$), it is more efficient to use the relevant endpoint \mathcal{E}_1 ;
3. If treatment reduces the risk of hospitalization by approximately 10% ($HR_2 = 0.9$) and T_1 and T_2 are weakly correlated ($\rho = 0.15$ or $\rho = 0.25$), then
 - It is more efficient to add hospitalization to the primary endpoint and use \mathcal{E}_* if Spearman's $\rho = 0.15$ and (i) the hazard function for the relevant endpoint \mathcal{E}_1 is decreasing while the hazard function for the additional endpoint \mathcal{E}_2 is constant or increasing; or (ii) the hazard function for \mathcal{E}_1 is constant while the hazard function for \mathcal{E}_2 is increasing;
 - It is more efficient to use the relevant endpoint, that is, the first occurrence between cardiovascular death, resuscitated cardiac arrest, non-fatal myocardial infarction and non-fatal stroke if Spearman's $\rho = 0.25$ and (i) the hazard function for the relevant endpoint \mathcal{E}_1 is increasing while the hazard function for the additional endpoint \mathcal{E}_2 is constant or decreasing; or (ii) the hazard function for \mathcal{E}_1 is constant while the hazard function for \mathcal{E}_2 is decreasing;

- We have encountered 12 situations (2.4%), corresponding to those situations listed in Table 3 in which $\rho = 0.15$ or $\rho = 0.25$ and the hazard functions for \mathcal{E}_1 and \mathcal{E}_2 do not fulfill the criteria described above, where there is not total concordance among the 12 copulas, meaning that ARE is less than 1 for some, while larger than 1 for the others. However the situation is not worrisome because the ARE values, in all these cases, are very close to 1 ($0.94 \leq \text{ARE} \leq 1.04$). Note that, in these situations, the difference in the sample size one might need if using either \mathcal{E}_* or \mathcal{E}_1 would be very small. In all these cases the benefits of using the composite endpoint over the relevant endpoint are marginal and the **most clinically relevant endpoint should be used as the primary endpoint**.

4.3 Computational study

We have shown with the succinobucol illustration that the recommendation on whether to use \mathcal{E}_1 or \mathcal{E}_* as the primary endpoint was, almost always, the same irrespective of the copula chosen. We explore via a large computational study how general is this pattern and whether or not the choice of the copula implies fundamental changes in the recommendation based on the ARE value.

The study reproduces 145,152 situations (shown in Table 4) corresponding to:

- Several frequency situations for events \mathcal{E}_1 and \mathcal{E}_2 by taking probabilities p_1 and p_2 equal to 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5.
- The relative treatment effect on the relevant endpoint \mathcal{E}_1 , given by the hazard ratio HR_1 , is set to 0.5, 0.6, 0.7 and 0.8, indicating that the effect of the treatment reduces the risk of \mathcal{E}_1 by 50%, 40%, 30% and 20%, respectively. Each hazard ratio is combined with eight different relative treatment effects on the additional endpoint \mathcal{E}_2 , namely HR_2 , and set to 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 0.95 (reduction of the risk of \mathcal{E}_1 ranging from 70% to 5%), reproducing situations where the effect on the additional endpoint is more beneficial than on the relevant endpoint, and viceversa.
- Marginal Weibull distribution with decreasing, constant and increasing hazards (shape parameters equal to 0.5, 1 and 2, respectively).
- Several degrees of association have been considered: from weak (Spearman's rank correlation $\rho = 0.15, 0.25$), through moderate ($\rho = 0.35, 0.45$) to strong ($\rho = 0.55, 0.65, 0.75$).
- Censoring case 1 where death is not among the components of \mathcal{E}_1 neither among \mathcal{E}_2 and Censoring case 3 where death is among the components of \mathcal{E}_1 but not of \mathcal{E}_2 .

Since results are very similar for both censoring cases, here we only present results for Case 1. A brief descriptive study of the ARE values is presented in Table 5 (72,576 simulated scenarios). ARE values for Frank, Gumbel and Clayton copulas range between 0.03 and 267.3, 0.03 and 272.7 and 0.02 and 301.3, respectively, with mean (standard deviation) equal to 4.95 (15.2), 5.08 (15.4) and 5.43 (16.9), respectively. These ranges are similar for other copulas.

We have compared the ARE values pairwise for any 2 copulas and, for each pair, the following measures of association have been used: Pearson's correlation coefficient ρ , Kendall's τ and Lin's concordance correlation coefficient (CCC) (Lin, 1989; 2000). Table 6 presents the association between the ARE values obtained when Frank copula is used versus the ARE values obtained for any of the other 10 copulas. We observe very large coefficients (larger than 92%) for all the copulas and all the association measures. In particular, Pearson's correlation coefficient and Lin's concordance correlation coefficient are always larger than 0.99 while Kendall's is a bit lower but above 0.92 in all the cases. Other comparisons for other pairs of copulas yielded similar results, for instance the pair Gumbel-Clayton has Pearson's $\rho=0.998$, Kendall's $\tau=0.932$ and Lin's CCC=0.994.

Scatter plots for the ARE values for any 2 copulas complement the above association measures. These plots have been restricted to ARE values within the range $[0, 2]$ because an ARE value = 1 draws the line between recommending the relevant endpoint \mathcal{E}_1 or the composite endpoint \mathcal{E}_* . Figure 4 shows three such plots for the comparisons Frank-Gumbel, Frank-Clayton and Gumbel-Clayton. We clearly observe that the values obtained using Frank and Gumbel copulas are highly correlated, although the values are, on average, slightly larger using Gumbel copula. The ARE values for the pair

Frank-Clayton are also correlated with larger variability than for the pair Frank-Gumbel. The comparison Gumbel-Clayton is also reproduced showing a similar behavior to Frank-Clayton. Comparisons to other copulas are similar to these ones and are omitted.

4.4 Results

When it comes to the real application of the ARE method, and given the clear difficulty in having to decide between one or other copula, it is of crucial importance to know how much agreement there will be in the recommendations as a function of the copula chosen. As it will be detailed explained below, our conclusion is that the recommendation for the choice of the primary endpoint does not differ if the copula used is changed. Therefore, we can use Frank's copula as we claim at the end of Subsection 2.3.

With this in mind, we define the degree of agreement between two copulas as the percentage of situations in which both copulas agree in either recommending the use of the relevant endpoint ($ARE \leq 1$) or recommending the composite endpoint ($ARE > 1$). Based on our reproduced settings, we find a very high degree of agreement (see Table 6). In all the comparisons, the degree of agreement is larger than 94%, for instance 98.0% for the pair Frank-Gumbel, 94.7% for Frank-Clayton and 96.3% for Gumbel-Clayton.

We then study the discordant situations, that is, those scenarios yielding an ARE value with copula $C_1 > 1$ while an ARE value with copula $C_2 \leq 1$, which would imply a different decision on which primary endpoint to use. We restrict the results presented here to the most popular copulas: Frank, Gumbel and Clayton (see Table 7) since the comparisons with the other copulas are similar to these ones. We observe that, mostly, the difference between the two ARE values is very small (medians ≤ 0.09). These situations are not worrisome because, when it comes to establishing statistical efficiency guidelines, the advantage of one endpoint over the other is very slight in the vicinity of one, and whenever $1 < ARE(Z_*, Z) < 1.1$, the benefits of using the composite endpoint over the relevant endpoint are marginal and, probably, too small to counteract the interpretational complications of a composite endpoint (Gómez and Lagakos, 2013). Furthermore, ARE values in the vicinity of 1 would imply, approximately, the same sample size whether ε_1 or ε_* is used. **The recommendation in these situations would be that the investigators choose the most convenient endpoint to answer the research question, taking into account practical issues such as regulatory guidances.** Finally, those discordant situations where the difference is large enough and the direction of the recommendation would be reversed is studied next. On one hand, the agreement between Frank and Gumbel is extremely good with absolute differences at most of 0.14. For the pairs Frank-Clayton and Gumbel-Clayton, there are respectively 1,138 (1.6%) and 617 (0.8%) discordant situations in which the distance between the two ARE values is larger than 0.15, which would lead to a difference in required sample sizes of more than 15%. Most of those situations (1,088 and 554, respectively) correspond to settings where the two relative treatment effects are very close ($HR_1 = HR_2$ or $HR_1 = HR_2 - 0.1$) and the relevant and the additional endpoints are moderate or highly correlated ($\rho \geq 0.35$).

Summarizing, the recommendation for the choice between a relevant and a composite endpoint does not substantially differ when using different copulas and we can safely use Frank's copula for the computation of the ARE value. **In situations in which the ARE value is close to 1, the investigators should consider other copulas to check the robustness of the choice of the most efficient endpoint.**

5 CONCLUDING REMARKS

The decision on which endpoint to choose as the primary endpoint for the analysis of the efficacy of two treatments is of uttermost importance for an effective clinical trial. In this paper we have presented the applicability of the ARE method in two cardiovascular trials and have shown how the information that the ARE would yield at the design stage of the trial, to choose between a relevant endpoint of interest or the combination of this endpoint with an additional one, could help towards a more informed choice of the primary endpoint. The method is based on the value of the Asymptotic Relative Efficiency (ARE) between the logrank test for the null hypothesis of no treatment effect observed in the

composite endpoint versus the logrank test of no treatment effect observed in the relevant component of the composite endpoint.

The ARE method will only be of practical use if it can be computed from a very small number of anticipatable values. We have shown that all what is needed is the expected proportion of subjects experiencing both the relevant and the additional endpoints in the control group and the hazard ratios as a measure of the expected relative effect of the treatment to these endpoints. This set of quantities is part of the information required for the computation of the sample size in any trial and they could not therefore be considered as a limitation during the design phase. Furthermore, plots like those shown in Figure 1 are proposed to derive recommendations without the need of anticipating the degree of association between the relevant and the additional endpoints and for several different relative effects of the treatment on the additional endpoint.

Since the ARE value depends on the joint law between the times to the relevant endpoint and to the additional endpoint (by means of a copula function), it is of great concern the possible influence that this choice might have. To this end, the paper has extended the ARE method to other families of copulas and has shown that the ARE values for any two different copulas provides, almost always, recommendations on the same direction.

Summarizing, this methodology is widely applicable and the computation of the ARE with the purpose of choosing the primary endpoint can be, in general, restricted to Frank's copula. To make the method widely applicable, we are developing a friendly interactive website, CompARE. The design of this free tool allows the user to enter their own values when designing a clinical trial and to show ARE values in plots, as the ones reproduced in Figures 1 and 2. Beta versions can be found at <http://composite.upc.edu/CompARE> or are available from the second author under request.

REFERENCES

- D. Collett. *Modelling survival data in medical research*. Chapman and Hall, 1994.
- I. Ferreira-González, J. Busse, D. Heels-Ansdell, V. Montori, E. Akl, D. Bryant, and others. Problems with use of composite endpoints in cardiovascular trials: systematic review of randomized controlled trials. *British Medical Journal*, 334:786–792, 2007a.
- I. Ferreira-González, G. Permanyer-Miralda, J. Busse, D. Bryant, V. Montori, P. Alonso-Coello, S. Walter, and G. Guyatt. Methodological discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *Journal of Clinical Epidemiology*, 60:651–657, 2007b.
- N. Freemantle, M. Calvert, J. Wood, J. Eastaugh, and C. Griffin. Composite outcomes in Randomized Trials. Greater precision but with greater uncertainty? *Journal of the American Medical Association*, 289: 2554–2559, 2003.
- G. Gómez. Some theoretical thoughts when using a composite endpoint to prove the efficacy of a treatment. In *Proceedings of the 26th International Workshop on Statistical Modelling*, Edited by D. Conesa, A. Forte, A. López-Quílez and F. Muñoz. *International Workshop on Statistical Modelling*. <http://hdl.handle.net/2117/22571>, pages 14–21, Valencia, 2011. ISBN 978-84-694-5129-8.
- G. Gómez and M. Gómez-Mateu. The asymptotic relative efficiency and the ratio of sample sizes when testing two different null hypotheses. *SORT*, 38:73–88, 2014.
- G. Gómez and S. Lagakos. Statistical considerations when using a composite endpoint for comparing treatment groups. *Statistics in Medicine*, 32:719–738, 2013.
- G. Gómez, M. Gómez-Mateu, and U. Dafni. Informed choice of composite endpoints in cardiovascular trials. *Circulation: Cardiovascular quality and outcomes*, 7:170–178, 2014.
- M. Hofert and M. Maechler. Nested Archimedean Copulas Meet R: The nacopula Package. *Journal of Statistical Software*, 39(9):1–20, 2011.
- M. Hofert, I. Kojadinovic, M. Maechler, and J. Yan. *copula: Multivariate Dependence with Copulas.*, 2014. R package version 0.999-8.

- M. F. Huque, M. Alosch, and R. Bhore. Addressing multiplicity issues of a composite endpoint and its components in clinical trials. *Journal of Biopharmaceutical Statistics*, 21:610–634, 2011.
- H. Joe. *Multivariate models and multivariate dependence concepts*. Monographs on Statistics and Applied Probability. Chapman and Hall, London, 1997.
- I. Kojadinovic and J. Yan. Modeling Multivariate Distributions with Continuous Margins Using the copula R Package. *Journal of Statistical Software*, 34(9):1–20, 2010.
- L. I. Lin. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45:255–268, 1989.
- L. I. Lin. A note on the concordance correlation coefficient. *Biometrics*, 56:324–325, 2000.
- C. L. Meinert. *Clinical Trials Dictionary*. John Wiley and Sons, Inc., 2012.
- V. Montori, G. Permanyer-Miralda, I. Ferreira-González, J. Busse, V. Pachecho-Huergo, D. Bryant, J. Alonso, E. Akl, A. Domingo-Salvany, E. Mills, P. Wu, H. Schunemann, R. Jaeschke, and G. Guyatt. Validity of composite endpoints in clinical trials. *British Medical Journal*, 330:594–596, 2007.
- R. B. Nelsen. *An Introduction to Copulas. Lecture Notes in Statistics*. Springer-Verlag, 1999.
- M. Packer, A. Coats, M. Fowler, H. Katus, H. Krum, P. Mohacsi, J. Rouleau, M. Tendera, A. Castaigne, E. Roecker, M. Schultz, and D. DeMets. Effect of carvedilol on survival in severe chronic heart failure. *The New England Journal of Medicine*, 344:1651–1658, 2001.
- R. Song, T. D. Cook, and M. R. Kosorok. What we want versus what we can get: a closer look at failure time endpoints for cardiovascular studies. *Journal of Biopharmaceutical Statistics*, 18:370–381, 2008.
- J. Tardif, J. McMurray, E. Klug, R. Small, J. Schumi, J. Cooper, R. Scott, E. Lewis, P. L’Allier, and Pfeffer, MA. for the Aggressive Reduction of Inflammation Stops Events (ARISE) Trial Investigators. Effects of succinobucol (AGI-1067) after an acute coronary syndrome: a randomised, double-blind, placebo-controlled trial. *The Lancet*, 371:1761–1768, 2008.
- P. Trivedi and D. Zimmer. *Copula Modelling: An Introduction for Practitioners*. Foundation and Trends in Econometrics, 2007.
- J. Yan. Enjoy the Joy of Copulas: With a Package copula. *Journal of Statistical Software*, 21(4):1–21, 2007.

A APPENDIX

A.1 Joint law of T_1 and T_2 by means of a copula

Given a bivariate copula C , marginal survival functions $S_1^{(j)}(t_1)$ and $S_2^{(j)}(t_2)$ of T_1 and T_2 , respectively, in treatment group j ($j = 0, 1$), and a dependence parameter θ between T_1 and T_2 , the joint survival function of (T_1, T_2) in group j is given by

$$S_{(1,2)}^{(j)}(t_1, t_2; \theta) = S_1^{(j)}(t_1) + S_2^{(j)}(t_2) - 1 + C(1 - S_1^{(j)}(t_1), 1 - S_2^{(j)}(t_2); \theta), \quad (3)$$

where C , function from I^2 to $I = [0, 1]$, is a two-dimensional copula that binds together $1 - S_1^{(j)}(t_1)$ and $1 - S_2^{(j)}(t_2)$ as follows:

$$\Pr\{T_1 \leq t_1, T_2 \leq t_2 | \text{group } j\} = C(1 - S_1^{(j)}(t_1), 1 - S_2^{(j)}(t_2); \theta).$$

Whereas each copula parametrises the dependence between two marginals, the dependence parameter θ quantifies the magnitude of this dependence and there is a one-to-one relationship with Spearman's rank correlation ρ between T_1 and T_2 , as follows (Nelsen, 1999):

$$\rho(\theta) = 12 \int_0^1 \int_0^1 [C(u, v; \theta) - uv] du dv$$

Note here that the dependence parameter θ has been taken equal for both treatment groups as in Gómez and Lagakos (2013).

The survival function of $T_* = \min\{T_1, T_2\}$ for group j , namely $S_*^{(j)}(t; \theta) = \Pr\{T_* > t; \theta | X = j\}$, is given by

$$S_*^{(j)}(t; \theta) = \Pr\{T_1 > t, T_2 > t; \theta\} = S_{(1,2)}^{(j)}(t, t; \theta).$$

It follows that the survival, density and hazard function for the composite endpoint T_* in terms of the copulas are given by

$$S_*^{(j)}(t; \theta) = S_1^{(j)}(t) + S_2^{(j)}(t) - 1 + C(1 - S_1^{(j)}(t), 1 - S_2^{(j)}(t); \theta) \quad (4)$$

$$f_*^{(j)}(t; \theta) = f_1^{(j)}(t) + f_2^{(j)}(t) - \frac{\partial}{\partial t} C(1 - S_1^{(j)}(t), 1 - S_2^{(j)}(t); \theta) \quad (5)$$

$$\lambda_*^{(j)}(t; \theta) = \frac{f_*^{(j)}(t; \theta)}{S_*^{(j)}(t; \theta)}.$$

A.2 Frank, Gumbel and Clayton copulas

The expression of Frank's copula is given by

$$C_F(u, v; \theta) = \frac{-1}{\theta} \log \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right) \quad (6)$$

where the dependence parameter may assume any real value, ($\theta \in (-\infty, \infty) \setminus \{0\}$), and the limiting case $\theta \rightarrow 0$ corresponds to the independence between T_1 and T_2 . Frank copula is a symmetric copula that allows both negative and positive dependence between variables and exhibits the same dependence in the left and in the right tail. A two-way scatter diagram of realizations from simulated drawings from copulas illustrates tail dependences in a bivariate framework (see Figure 5). Frank copula is best suited for applications in which tail dependence is relatively weak, as it is shown in the plot.

Gumbel copula is given by

$$C_G(u, v; \theta) = \exp \left(-[(-\log(u))^\theta + (-\log(v))^\theta]^{1/\theta} \right) \quad (7)$$

where the dependence parameter may assume any real number greater or equal than 1, ($\theta \in [1, \infty)$). Gumbel copula only allows positive dependence and exhibits relatively weak left tail dependence and

strong right tail dependence (Trivedi and Zimmer, 2007) (see Figure 5). Gumbel is an appropriate choice for the joint behavior when the two endpoints are likely to simultaneously realize upper tail values.

Clayton copula is given by

$$C_C(u, v; \theta) = \max\{[u^{-\theta} + v^{-\theta} - 1]^{-1/\theta}, 0\} \quad (8)$$

where the dependence parameter may assume any real number larger or equal than -1 ($\theta \in [-1, \infty) \setminus \{0\}$). Clayton copula only accounts for positive dependence and exhibits strong left tail dependence and relatively weak right tail dependence (Trivedi and Zimmer, 2007) (see Figure 5). Clayton copula is best suited for applications in which the two outcomes are likely to experience low values together.

A.2.1 Survival and density functions of T_*

1. Frank copula

If Frank copula C_F presented in (6) is used, the joint survival function for (T_1, T_2) in group j ($j = 0, 1$) is given by

$$S_{(1,2)}^{(j)}(t_1, t_2; \theta) = \frac{-1}{\theta} \log \left(1 + \frac{(e^{-\theta S_1^{(j)}(t_1)} - 1)(e^{-\theta S_2^{(j)}(t_2)} - 1)}{e^{-\theta} - 1} \right).$$

It follows from (5) that the density of $T_* = \min\{T_1, T_2\}$ is as follows:

$$f_*^{(j)}(t; \theta) = \frac{1}{e^{-\theta} - 1} \left[\frac{e^{-\theta S_1^{(j)}(t)}(e^{-\theta S_2^{(j)}(t)} - 1)}{e^{-\theta S_{(1,2)}^{(j)}(t,t;\theta)}} f_1^{(j)}(t) + \frac{e^{-\theta S_2^{(j)}(t)}(e^{-\theta S_1^{(j)}(t)} - 1)}{e^{-\theta S_{(1,2)}^{(j)}(t,t;\theta)}} f_2^{(j)}(t) \right].$$

2. Gumbel copula

Using Gumbel copula C_G given in (7), it follows from (4) that the survival function of $T_* = \min\{T_1, T_2\}$ in group j ($j = 0, 1$) is given by

$$S_*^{(j)}(t; \theta) = S_1^{(j)}(t) + S_2^{(j)}(t) - 1 + \exp \left(- \left[(-\log(1 - S_1^{(j)}(t)))^\theta + (-\log(1 - S_2^{(j)}(t)))^\theta \right]^{1/\theta} \right),$$

and from (5) that the density of T_* is as follows:

$$\begin{aligned} f_*^{(j)}(t; \theta) &= f_1^{(j)}(t) + f_2^{(j)}(t) - \exp \left(- \left[(-\log(1 - S_1^{(j)}(t)))^\theta + (-\log(1 - S_2^{(j)}(t)))^\theta \right]^{1/\theta} \right) \\ &\quad \left[(-\log(1 - S_1^{(j)}(t)))^\theta + (-\log(1 - S_2^{(j)}(t)))^\theta \right]^{\frac{1-\theta}{\theta}} \\ &\quad \left((-\log(1 - S_1^{(j)}(t)))^{\theta-1} \frac{f_1^{(j)}(t)}{1 - S_1^{(j)}(t)} + (-\log(1 - S_2^{(j)}(t)))^{\theta-1} \frac{f_2^{(j)}(t)}{1 - S_2^{(j)}(t)} \right) \end{aligned}$$

3. Clayton copula

Using Clayton copula C_C given in (8), it follows from (4) that the survival function for $T_* = \min\{T_1, T_2\}$ in group j ($j = 0, 1$) is given by

$$S_*^{(j)}(t; \theta) = S_1^{(j)}(t) + S_2^{(j)}(t) - 1 + [(1 - S_1^{(j)}(t))^{-\theta} + (1 - S_2^{(j)}(t))^{-\theta} - 1]^{-1/\theta}$$

and from (5) that the density of T_* is as follows:

$$\begin{aligned} f_*^{(j)}(t; \theta) &= f_1^{(j)}(t) + f_2^{(j)}(t) - [(1 - S_1^{(j)}(t))^{-\theta} + (1 - S_2^{(j)}(t))^{-\theta} - 1]^{-\frac{1+\theta}{\theta}} \\ &\quad \left((1 - S_1^{(j)}(t))^{-(1+\theta)} f_1^{(j)}(t) + (1 - S_2^{(j)}(t))^{-(1+\theta)} f_2^{(j)}(t) \right) \end{aligned}$$

A.3 Computation of the ARE values using R

As it is shown in Subection 2.3, the expression of the ARE value relates to the following parameters: $(p_1, p_2, HR_1, HR_2, \rho, \beta_1, \beta_2)$, representing, respectively, the frequencies of observing the endpoints \mathcal{E}_1 and \mathcal{E}_2 in treatment group 0, the relative treatment effects on \mathcal{E}_1 and \mathcal{E}_2 , Spearman's rank correlation between T_1 and T_2 in either group and the shape parameters of the Weibull marginal survivals.

Using the R-package `copula` (Hofert et al., 2014; Yan, 2007; Kojadinovic and Yan, 2010; Hofert and Maechler, 2011), we have written a program to compute the ARE values for a given copula chosen among the 11 families introduced in Subsection 4.1, taking into account whether \mathcal{E}_1 contains a terminating event (case) and as a function of the above parameters:

$$\text{ARE}(\text{rho}, \text{beta1}, \text{beta2}, p1, p2, \text{HR1}, \text{HR2}, \text{case}, \text{copula})$$

This function is easy to use and allows the computation of the ARE value for a large amount of simulated situations. The body of this function can be divided into three parts:

1. the first one computes the dependence parameter θ for a given Spearman's correlation ρ ;
2. the second part of the algorithm builds the marginal distribution functions from the anticipated values $(p_1, p_2, HR_1, HR_2, \beta_1, \beta_2)$; and
3. the third part of the program computes the value of the ARE for the dependence parameter and the marginal distributions set in the first two parts of the function.

Details on the code can be requested to either author.

TABLES AND FIGURES

Table 1: Values of the parameters used to compute the ARE value for the succinobucol and carvedilol studies given in Section 3. We are considering: increasing ($\beta_i = 2$), constant ($\beta_i = 1$) and decreasing ($\beta_i = 0.5$) marginal hazard functions for T_1 and T_2 ; several relative effects of the treatment on the additional endpoint (HR_2) and several Spearman's correlation values (ρ) between the relevant and the additional endpoints.

β_1	0.5	1	2						
β_2	0.5	1	2						
HR_2	0.3	0.4	0.5	0.6	0.7	0.8	0.85	0.9	0.95
ρ	0.15	0.25	0.35	0.45	0.55	0.65	0.75		

Table 2: Copula expression, $C(u, v; \theta)$, and domain of the association parameter θ for the following copulas: Frank, Gumbel, Clayton, Farlie-Gumbel-Morgenstern (FGM), Normal, t , Plackett, Galambos, Hüsler-Reiss, Tawn and t -EV.

Copula type	$C(u, v; \theta)$	θ -domain
Frank	$\frac{-1}{\theta} \log \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right)$	$(-\infty, \infty)$
Gumbel	$\exp \left(-[(-\log(u))^\theta + (-\log(v))^\theta]^{1/\theta} \right)$	$[1, \infty)$
Clayton	$\max\{[u^{-\theta} + v^{-\theta} - 1]^{-1/\theta}, 0\}$	$[-1, \infty) \setminus \{0\}$
FGM	$uv + \theta uv(1-u)(1-v)$	$[-1, 1]$
Normal	$\Phi_\theta(\Phi^{-1}(u), \Phi^{-1}(v))$	$[-1, 1]$
t	$t_{\theta, \nu}(t_\nu^{-1}(u), t_\nu^{-1}(v))$	$[-1, 1]$
Plackett	$\frac{[1+(\theta-1)(u+v)] - \sqrt{[1+(\theta-1)(u+v)]^2 - 4uv\theta(\theta-1)}}{2(\theta-1)}$	$(0, \infty)$
Galambos	$uv \exp\{[(\log(u))^{-\theta} + (\log(v))^{-\theta}]^{-1/\theta}\}$	$[0, \infty)$
Hüsler-Reiss	$\exp\{-\log(u)\Phi[\frac{1}{\theta} + \frac{1}{2}\theta \log(\frac{\log(u)}{\log(v)})] - \log(v)\Phi[\frac{1}{\theta} + \frac{1}{2}\theta \log(\frac{\log(v)}{\log(u)})]\}$	$[0, \infty)$
Tawn	$uv \exp\left(-\theta \frac{\log(u)\log(v)}{\log(uv)}\right)$	$[0, 1]$
t -EV	$\exp\left(\log(uv)A_{\theta, \nu} \frac{\log(u)}{\log(v)}\right)$	$[0, 1]$

Φ_θ and $t_{\theta, \nu}$ with ν degrees of freedom denote the standard bivariate normal and Student's t joint distribution function with correlation coefficient θ , respectively. Φ and t_ν with ν degrees of freedom denote the standard normal and Student's t distribution function, respectively. $A_{\theta, \nu}$ is the Pickands dependence function based in the bivariate Student's $t_{\theta, \nu}$. In this study, the degrees of freedom have been set to $\nu = 4$.

Table 3: ARE values for the discordant cases of succinobucol study given in Section 3.1. Reproduction of the 12 scenarios in which the ARE value is > 1 for some copulas and < 1 for the others. In these cases the probability of observing \mathcal{E}_1 is $p_1 = 0.082$, the probability of observing \mathcal{E}_2 is $p_2 = 0.09$, the relative effect of the treatment are $HR_1 = 0.81$ and $HR_2 = 0.9$ and the Spearman’s correlation ρ between T_1 and T_2 is $\rho \leq 0.25$.

			ARE values depending on the copula chosen										
ρ	β_1	β_2	Frank	Gumbel	Clayton	Normal	t	Plackett	Galambos	Hüsler-Reiss	t -EV	Tawn	FGM
0.15	0.5	0.5	1.02	1.04	0.99	1.02	1.00	1.02	1.04	1.04	1.04	1.03	1.03
0.15	1	0.5	1.01	1.02	0.98	1.01	0.99	1.01	1.03	1.03	1.02	1.02	1.02
0.15	1	1	1.02	1.04	0.99	1.02	1.00	1.02	1.04	1.04	1.04	1.03	1.03
0.15	2	0.5	1.01	1.02	0.97	1.00	0.99	1.00	1.02	1.02	1.02	1.01	1.01
0.15	2	1	1.01	1.02	0.98	1.01	0.99	1.01	1.03	1.03	1.02	1.02	1.02
0.15	2	2	1.02	1.04	0.99	1.02	1.00	1.02	1.04	1.04	1.04	1.03	1.03
0.25	0.5	0.5	0.98	1.00	0.94	0.97	0.97	0.98	1.00	1.00	1.00	1.00	0.99
0.25	0.5	1	1.00	1.02	0.95	0.99	0.98	0.99	1.02	1.02	1.02	1.01	1.01
0.25	0.5	2	1.02	1.04	0.97	1.00	1.00	1.01	1.03	1.04	1.03	1.03	1.02
0.25	1	1	0.98	1.00	0.94	0.97	0.97	0.98	1.00	1.00	1.00	1.00	0.99
0.25	1	2	0.99	1.01	0.95	0.99	0.98	0.99	1.02	1.02	1.02	1.01	1.01
0.25	2	2	0.98	1.00	0.94	0.97	0.97	0.98	1.00	1.00	1.00	1.00	0.99

Table 4: Values of the parameters used to compute the ARE values for the computational study in Section 4.2. We are considering: increasing ($\beta_i = 2$), constant ($\beta_i = 1$) and decreasing ($\beta_i = 0.5$) marginal hazard functions for T_1 and T_2 ; several frequencies p_1 and p_2 for the relevant and additional endpoints; several relative effects of the treatment on the relevant endpoint (HR_1) and on the additional endpoint (HR_2) and several Spearman’s correlation values (ρ) between the relevant and the additional endpoint.

β_1	0.5	1	2						
β_2	0.5	1	2						
p_1	0.05	0.1	0.2	0.3	0.4	0.5			
p_2	0.05	0.1	0.2	0.3	0.4	0.5			
HR_1	0.5	0.6	0.7	0.8					
HR_2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	
ρ	0.15	0.25	0.35	0.45	0.55	0.65	0.75		

Table 5: Descriptive analysis of the ARE values obtained in the computational study in Section 4.2 using several copulas. Results for Tawn and FGM copulas are restricted to those scenarios in which the Spearman’s correlation ρ between T_1 and T_2 is within the range for these copulas ($\rho < 0.587$ ($n=51,840$) and $\rho < 1/3$ ($n=20,736$), respectively).

ARE using	mean (SD)	min	Q ₁	median	Q ₃	max
Frank copula	4.95 (15.2)	0.03	0.76	1.18	2.93	267.3
Gumbel copula	5.08 (15.4)	0.03	0.79	1.22	3.06	272.7
Clayton copula	5.43 (16.9)	0.02	0.86	1.21	3.12	301.3
Normal copula	5.13 (15.7)	0.03	0.80	1.22	3.06	280.4
t copula	5.33 (16.4)	0.03	0.84	1.24	3.13	283.2
Plackett copula	5.03 (15.5)	0.03	0.78	1.19	2.95	275.7
Galambos copula	5.08 (15.4)	0.03	0.79	1.23	3.07	272.4
Hüsler-Reiss copula	5.07 (15.4)	0.03	0.79	1.23	3.08	271.8
t -EV copula	5.08 (15.5)	0.03	0.79	1.22	3.06	273.1
Tawn copula	5.16 (15.4)	0.06	0.81	1.29	3.21	264.1
FGM copula	5.24 (15.2)	0.08	0.82	1.35	3.42	261.7

Table 6: Pearson's ρ , Kendall's τ , Lin's concordance correlation coefficient (CCC) and degree of agreement in recommending the main primary endpoint between the ARE value computed using Frank's copula and the ARE value computed using each of the other 10 copulas for the computational study in Section 4.2. Results for Tawn and FGM copulas are restricted to those scenarios in which the Spearman's correlation ρ between T_1 and T_2 is within the range for these copulas ($\rho < 0.587$ ($n=51,840$) and $\rho < 1/3$ ($n=20,736$), respectively)

Copula	Pearson's ρ	Kendall's τ	Lins's CCC	Agreement
Gumbel	> 0.999	0.982	> 0.999	98.0%
Clayton	0.997	0.927	0.991	94.7%
Normal	> 0.999	0.972	0.999	97.8%
t	> 0.999	0.964	0.996	95.5%
Plackett	> 0.999	0.984	> 0.999	98.8%
Galambos	> 0.999	0.982	> 0.999	98.1%
Hüsler-Reiss	> 0.999	0.981	> 0.999	98.1%
t -EV	> 0.999	0.982	> 0.999	98.0%
Tawn	> 0.999	0.971	> 0.999	97.4%
FGM	> 0.999	0.997	> 0.999	99.9%

Table 7: Summary of the absolute differences between each pair of ARE values for the pairs of copulas: Frank-Gumbel, Frank-Clayton and Gumbel-Clayton when there is no agreement between copulas C_1 and C_2 ($ARE(C_1) > 1$ while $ARE(C_2) \leq 1$), for the computational study in Section 4.2.

Discordant cases	n (%)	mean (SD)	min	Q ₁	median	Q ₃	P ₉₅	max
$ ARE_F - ARE_G $	1426 (2.0)	0.04 (0.03)	0.004	0.02	0.05	0.06	0.11	0.14
$ ARE_F - ARE_C $	3812 (5.3)	0.11 (0.08)	0.001	0.04	0.09	0.17	0.27	0.36
$ ARE_G - ARE_C $	2696 (3.7)	0.09 (0.07)	0.001	0.03	0.07	0.14	0.23	0.38

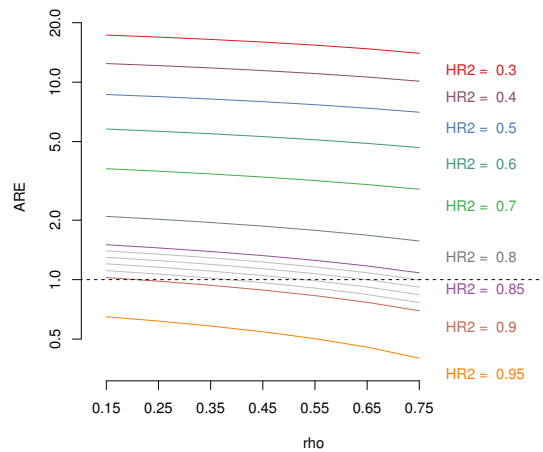


Figure 1: ARE values for the succinobucol case study given in Section 3.1. The plots are drawn for the values given in the study: probability of observing a cardiovascular death, resuscitated cardiac arrest, non-fatal myocardial infarction or non-fatal stroke, $p_1 = 0.082$; probability of observing hospitalization, $p_2 = 0.09$; relative effect of the treatment on cardiovascular death, resuscitated cardiac arrest, non-fatal myocardial infarction or non-fatal stroke, $HR_1 = 0.81$, for several relative effects of the treatment on hospitalization HR_2 and for increasing marginal hazard functions for the times T_1 and T_2 to either event. We plot the Spearman's correlation ρ between T_1 and T_2 in the x-axis and the ARE values (in logarithmic scale) on the y-axis

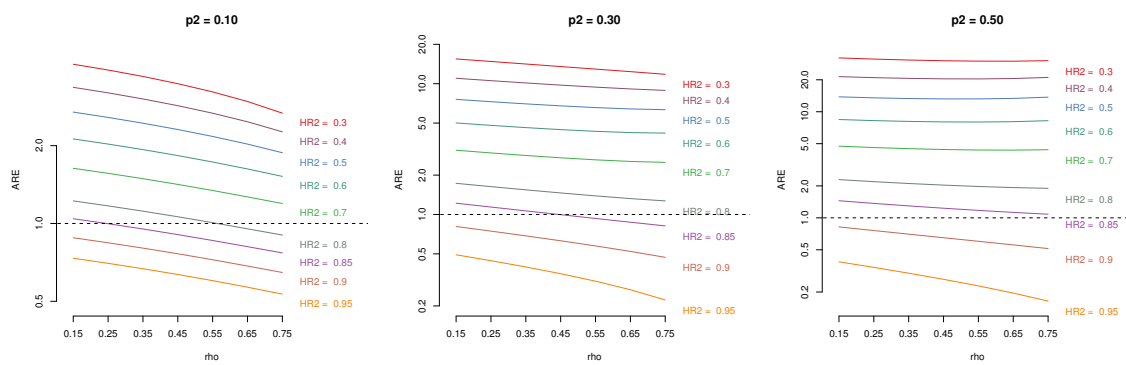


Figure 2: ARE values for the carvedilol case study given in Section 3.2. The plots are drawn for the values used in the study to compute the sample size: probability of observing a death, $p_1 = 0.28$ and relative effect of the treatment on death $HR_1 = 0.8$, for different probabilities p_2 of observing hospitalization (0.10, 0.30 and 0.50), for several relative effects of the treatment on hospitalization HR_2 and for constant marginal hazard functions for time to death T_1 and time to hospitalization T_2 . We plot the Spearman's correlation ρ between T_1 and T_2 in the x-axis and the ARE values (in logarithmic scale) on the y-axis

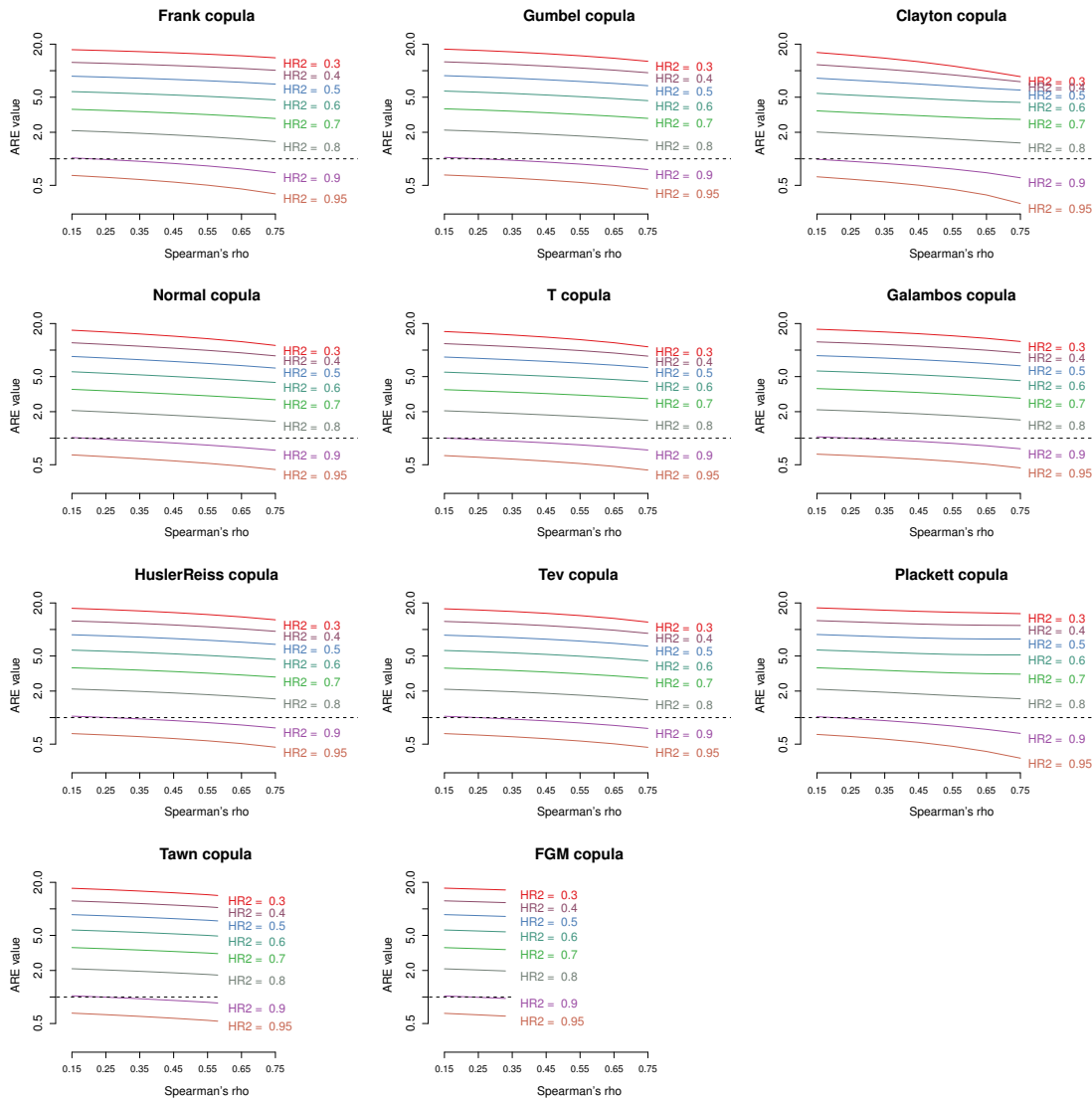


Figure 3: ARE values for the succinobucol case study given in Section 3.1. The plots are drawn for the values given in the study: probability of observing a cardiovascular death, resuscitated cardiac arrest, non-fatal myocardial infarction or non-fatal stroke, $p_1 = 0.082$; probability of observing hospitalization, $p_2 = 0.09$; relative effect of the treatment on cardiovascular death, resuscitated cardiac arrest, non-fatal myocardial infarction or non-fatal stroke, $HR_1 = 0.81$, for several relative effects of the treatment on hospitalization HR_2 and for increasing marginal hazard functions for the times T_1 and T_2 to either event. We plot the Spearman's correlation ρ between T_1 and T_2 in the x-axis and the ARE values (in logarithmic scale) on the y-axis. Each plot corresponds to one of the 11 different copulas. FGM copula only allows weak dependence ($\rho < 1/3$) and Tawn copula weak-moderate dependence ($\rho < 0.587$)

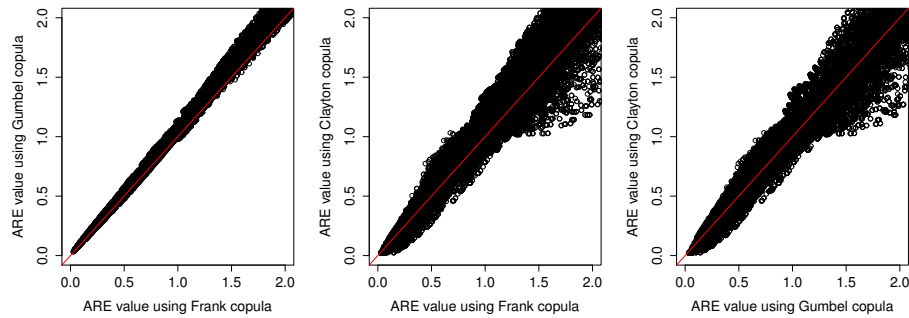


Figure 4: Scatter plot of the pairs of ARE values in the computational study in Section 4.2 using Frank and Gumbel copulas, Frank and Clayton copulas and Gumbel and Clayton copulas within the range $[0, 2]$. The diagonal line represents the situation in which both values are the same.

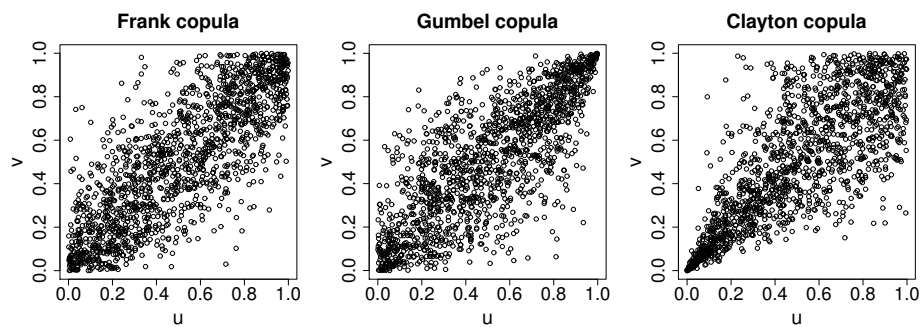


Figure 5: Simulation of 1500 samples of the function $C(u, v; \theta)$ corresponding to Frank, Gumbel and Clayton copulas. The association parameter θ has been set such that Spearman's correlation $\rho = 0.8$ for the three copulas. The conditional distribution method (Nelsen, 1999) has been used to generate these samples.