MASTER'S DEGREE THESIS

Interuniversity Master in Statistics and Operations Research

Title: Using Gumbel copula to assess the efficiency of the main endpoint in a randomized clinical trial and comparison with Frank copula

Author: Oleguer Plana Ripoll

Advisor: Guadalupe Gómez Melis

Department: Statistics and Operations Research

University: *Universitat Politècnica de Catalunya*

Academic year: 2011-2012

Facultat de Matemàtiques i Estadística
UNIVERSITAT POLITÈCNICA DE CATALUNYA

UNIVERSITAT DE BARCELONA

USING GUMBEL COPULA TO ASSESS THE EFFICIENCY OF THE MAIN ENDPOINT IN A RANDOMIZED CLINICAL TRIAL AND COMPARISON WITH FRANK COPULA

Master Thesis in
Interuniversity Master in Statistics and Operations Research

Universitat Politècnica de Catalunya
Universitat de Barcelona

Oleguer Plana Ripoll

Supervised by Guadalupe Gómez Melis

Department of Statistics and Operations Research
Universitat Politècnica de Catalunya

June 11, 2012

Acknowledgements

I would like to express my gratitude to Professor Lupe Gómez because it has been an honor to work under her supervision in developing this master thesis. We have worked side by side and not only has she helped and guided me constantly during my work sharing her knowledge and expertise but she also showed me the meaning of belonging to a research group and encouraged me to keep on doing research in the future.

I would also like to thank her for inviting me to participate in GRASS group seminars and to share with them the problems that I found during my research. I thank the researchers of this group for their opinions about this thesis, especially Moises Gómez for his previous research and advice and Klaus Langohr for always being willing to help, either with doubts with R, when it has been necessary for applying for grants, etc.

I would like to thank people from *Institut Universitari d'Investigació en Atenció Primària (IDIAP) Jordi Gol* for being flexible with my work schedule these past few months and supporting me in completing this master's degree in statistics.

I would also like to thank Laia for reading my project again and again, helping me make it more understandable and being there for me; Isabelle for her corrections of my English; and the rest of my classmates in this master for all the moments we spent together.

Finally, I would like to thank my family and friends for supporting me in my decisions and being so understanding for my absence lately.

Summary

In time-to-event randomized clinical trials, it is common to use composite endpoints, defined as the union of two or more relevant events, as the main endpoint when comparing two treatment groups. A new statistical methodology has been recently developed in order to study the advantages and disadvantages of using a composite endpoint. Furthermore, this new methodology derives guidelines for deciding whether to expand the main endpoint from a relevant endpoint to a composite endpoint considering the inclusion of an additional endpoint.

This methodology, developed by Gómez and Lagakos [1], is based on the asymptotic relative efficiency (ARE) of a logrank test for comparing two treatment groups with respect to a relevant endpoint versus the composite endpoint. In order to compute the ARE, it is necessary to have the joint law of the time to the relevant and additional endpoints. A useful way of approaching it is specifying this law as a function of the marginal laws for each time and an association parameter θ via a copula model. A copula is defined as a function that joins or couples multivariate distribution functions to their one-dimensional marginal distribution functions.

Given the marginal laws for the time to the relevant and additional endpoints and the correlation between them, different copulas yield to different joint laws. Gómez and Lagakos [1] considered Frank Arquimedean copula and, hence, it is necessary to check whether the modelization with a different copula implies a change in the ARE recommendation. The main aim of this master thesis is to develop this methodology using Gumbel copula and to compare it with the results obtained using Frank copula.

The results of this project show that, using Gumbel copula, it is recommended to use the composite endpoint if the hazard ratio of the additional endpoint is smaller (higher beneficial effect) than the hazard ratio of the relevant endpoint. However, when these beneficial effects are about the same, the composite endpoint should be used if the probability of observe the relevant endpoint is lower than the probability of observe the additional endpoint.

These results are similar to the ones obtained using Frank copula. We conclude that the methodology based on the ARE is robust for both Frank and Gumbel copula. We observe that both copulas are highly correlated ($\rho = 0.999$) and they yield to the same recommendation of whether or not to use the composite endpoint in more than 98% of the simulated situations studied. More-

over, the difference between the ARE values in those cases in which there is not concordance is negligible. Therefore, we conclude that the ARE method is robust for the choice of the copula when restricted to Frank and Gumbel copulas.

Keywords: Clinical Trials, Combined outcomes, Competing risks, Composite endpoints, Logrank test, Copulas, R, Asymptotic relative efficiency
2010 Mathematical Subject Classification: Statistics / Survival and censored data (62N)

Contents

Chapter 1

Introduction

When using a randomized clinical trial to compare two treatment groups A and B, it is very important to choose correctly the main endpoint of the trial. One might have to choose among several relevant endpoints to define the main endpoint. A composite endpoint, defined as the union of two or more events, is often used in order to increase the power of the trial to detect differences between treatment groups.

There was little discussion in the literature on the advantages and disadvantages of this practice from a statistical point of view until the project promoted by Guadalupe Gómez, the director of this master thesis, and Stephen Lagakos, international leader in biostatistics and AIDS research from the *Harvard School of Public Health*. Despite the untimely death of Professor Lagakos, this work continues under the leadership of Professor Gómez. In this project, a statistical methodology is developed to choose a relevant endpoint or the addition of a additional event and, thus, consider the composite endpoint [1].

In this methodology, the asymptotic relative efficiency (ARE) of the logrank test to compare the treatment groups with respect to the relevant endpoint in comparison with the logrank test with respect to the composite endpoint is used to decide when it is better to use one or the other. Most of the times the two endpoints of interest are correlated and the joint law of the times to the event is needed. A useful way to approach the joint law is to specify this law as a function of the marginal densities and an association parameter via a copula model.

In this methodology presented by Gómez and Lagakos [1], it is assumed that the outcomes follow a Weibull distribution and the copula used for the joint law is the Frank Archimedean copula. This new methodology is very important to design time-to-event clinical trials which compare two treatment groups because it provides rigorous arguments to decide what should be the main outcome and, therefore, reduces economic resources and unnecessary efforts, on one hand, and the number of patients to be recruited, on the other, which is very important and ethically necessary in studies on health. For this reason, it is crucial to make this methodology

as extended and applicable possible. To achieve this objective, it is necessary to test this method with other laws for the marginal and joint distributions in order to prove the robustness of the methodology presented in [1]. Therefore, it becomes essential to study and test other families of copulas and marginal laws.

The main aim of this master thesis is to contribute in this research studying the robustness of the method based on ARE for different copulas in assessing the efficiency of the main endpoint in a randomized clinical trial.

To achieve this, the specific goals are the following:

1. To learn and understand how this methodology is developed.
   Chapter 2 summarizes the research done by Gómez and Lagakos, who developed this methodology [1] and Moises Gómez, whose master thesis gave recommendations in the cardiovascular area [2].

2. To study the definition of copula and their properties.
   Chapter 3 presents an introduction to copulas summarizing Nelsen (1999) [3] and Trivedi and Zimmer (2007) [4]. The definition and basic properties of copulas and different existing copula families are given in this chapter. It is important to know and understand the differences between each of theses families in order to study the behavior of the method depending on the copula used.

3. To develop the analytical expression of the ARE for a copula different than Frank copula.
   In the methodology presented by Gómez and Lagakos [1], the expression of ARE is given when Frank copula is used. In chapter 4, an expression for ARE in terms of Gumbel copula is found. In chapter 5, the differences in the behavior of ARE between these different copulas are studied checking whether the conclusions achieved with Frank copula are valid (robust). In Appendix D, the expression of ARE in terms of Clayton copula is found and it is left for future research the study of its results.

4. To improve my skills using the statistical software R.
   In the practical exercises of my master courses, I gained skills with software R. In this project, these skills are improved because algorithms to compute the ARE value are programmed for the different chosen copulas and the descriptive analysis of the results is carried out using this software.

5. To learn how to use LaTeX in scientific writing
   LaTeX is a very important tool for research in mathematics, statistics or other fields where math formulas and symbols are needed. Therefore, another of the objectives of this master thesis is to improve my writing using it and learn how to use Beamer for the project's oral presentation.

6. To gain research skills and improve my writing in English
   As a first experience in research, this project gives me a good chance to gain research skills for my future. On the other hand, English is the most used language in investigation and, hence, it is essential for me to get used to it if I want to keep on doing research. Therefore, I have decided to write my master thesis in English, even if it means quite a challenge to me.

Chapter 2

State of the art

2.1 Clinical Trials

A clinical trial is a research study designed to provide extensive data that will allow for statistically valid evaluation of treatment or interventions on a group of individuals [5]. A medical clinical trial compares an endpoint between two or more groups. It is typically used to test a treatment versus placebo or standard of care. The most optimal trial is a randomized clinical trial, that allocates patients to one of the arms randomly and it is useful to avoid bias in this process.

2.2 Composite Endpoints

When using a randomized clinical trial to compare two treatment groups A and B, it is very important to choose correctly the main endpoint of the trial. One might have to choose among several relevant endpoints to define the main endpoint. A composite endpoint, defined as the union of 2 or more different events, is often used in order to increase the number of events expected to observe. In this case, subjects are followed until one of the events of interest occurs, whichever happens first, or until the follow-up period of the study ends.

The main clinical arguments for using a composite endpoint [6] are that the composite endpoint (1) captures the net benefit of the intervention and avoids the need to choose a single main endpoint when the different endpoints are of equal importance and (2) avoids interpretational problems associated with competing risks in the sense of preventing an apparent benefit being attributed to a given event.

The statistical arguments given for the use of a composite endpoint are (1) to increase the statistical efficiency and, therefore, have a more powerful test of the treatment efficacy and (2) to reduce the multiplicity problems that may occur when the different endpoints are used separately.

On the other hand, the main disadvantages of using a composite endpoint are the difficulty of interpret the results when the different single endpoints have different clinical importance and the fact that treatment effect on the composite endpoint does not necessarily imply an effect on each component.

More details about advantages and disadvantages of using a composite endpoint can be found in [6, 7, 8, 9, 10].

2.3 Statistical Considerations

It is common to use composite endpoints in time-to-event analysis, where the focus is the time from randomization until the first of a set of clinical outcomes occurs. Little research appears to have been reported on the question of increased statistical efficiency to be gained from using a composite endpoint in this context until the project promoted by Gómez and Lagakos [1], in which a methodology designed to decide when it is better to use the relevant endpoint or the composite endpoint was presented in a context of time-to-event analysis. The following sections summarize this research project.

2.3.1 Definition of the endpoint

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. We assume that we only have two relevant endpoints, that we denote by E_1 or relevant endpoint and E_2 or additional endpoint. Individuals are followed until the event of interest, or until the end of the study, whatever occurs first.

The composite endpoint is defined as $E_* = E_1 \cup E_2$ and the time to E_* is defined as $T_* = \min\{T_1, T_2\}$ where T_1 and T_2 denote the times to E_1 and E_2, respectively, and are assumed to be absolutely continuous so that ties cannot occur. C represents the time from randomization to the end of the study and it is the only noninformative censoring cause.

Observation of endpoints E_1 and E_2 depends on whether or not they include a terminating event. For example, E_1 might not be observed if E_2 includes death. We can consider four different censoring situations depending on theses premises:

Case 1: This first case corresponds to clinical trials where neither of the two endpoints (E_1 and E_2) includes a terminating or fatal event. In this case, E_i ($i = 1, 2$) will be observed if it occurs before the end of the study, i.e., whenever $T_1 < C$ and $T_2 < C$.

Case 2: In a case-2 censoring situation, the relevant endpoint E_1 does not include mortality whereas the additional endpoint E_2 includes a terminating event. In this case, E_1 is only observed if

$T_1 < \min\{T_2, C\}$ while E_2 is observed if $T_2 < C$. That is E_1 is observed if it occurs before E_2 and the end of the study and E_2 is always observed if it occurs before the end of the study.

Case 3: This situation is analogue to case 2 but, in this case, E_1 includes a terminating event and E_2 does not include it. E_1 will be observed if the time to the relevant endpoint, T_1, occurs whitin the study time ($T_1 < C$) while the additional endpoint E_2 will only be observed if it occurs within the time of the study and before the relevant endpoint, i.e., $T_2 < \min\{T_1, C\}$.

Case 4: In a case-4 censoring situation both events of interest, E_1 and E_2 include a terminating event. In this situation, E_1 is observed if $T_1 < \min\{T_2, C\}$ and E_2 is observed if $T_2 < \min\{T_1, C\}$.

It is important to remark that censoring cases 2 and 3 are not symmetrical. The time to E_1, T_1, in cases 1 and 3, is only right-censored by the end of the study C because the additional endpoint E_2 does not include a terminating event. Note that, in cases 2 and 4, E_2 does include mortality and it is competing with the censoring random variable C in the observation of E_1. Thus, in cases 2 and 4, we have a competing risk situation with dependent censoring on T_1.

The observed outcome will be denoted by U in every case, defining $U = \min\{T_1, C\}$ in cases 1 and 3 and $U = \min\{T_1, T_2, C\}$ in cases 2 and 4. On the other hand, E_* will be always observed if $T_* < C$. Thus, the observed outcome will be denoted by U_* and defined by $U_* = \min\{T_*, C\}$.

In this master thesis, we focus on in-depth study of cases 1 and 3, leaving the extension of these to cases 2 and 4 for future research (Figure 2.1).

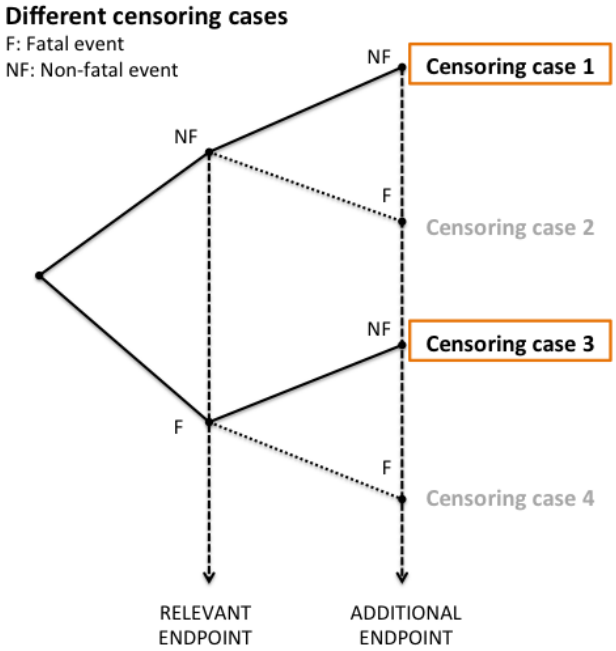


Figure 2.1: Different censoring cases depending on the fatality of the endpoints.

2.3.2 Logrank test and asymptotic relative efficiency

The methodology developed in [1] considers the asymptotic relative efficiency (ARE) of the logrank test to compare treatment groups with respect to E_1 in comparison with E_* in order to recommend the use of E_1 or E_* in terms of initial parameters provided by expert researchers in the field, such as the proportion of events expected in the control group and the hazard ratio between two treatments.

The logrank test is a nonparametric hypothesis test to compare the curves S_1 and S_2 from two samples of survival data given, for example, from a randomized clinical trial with two different treatment groups [11]. Formally, the following test is defined

$$H_0 : S_1(t) = S_2(t) \text{ for all } t \in [0, \tau] \text{ versus } H_1 : S_1(t) \neq S_2(t) \text{ for, at least, one } t \in [0, \tau]$$

where S_1 and S_2 are the survival functions from each of the two populations and they are supposed to be defined in $[0, \tau]$. The hypothesis test defined above is equivalent to the following one

$$H_0 : \lambda_1(t) = \lambda_2(t) \text{ for all } t \in [0, \tau] \text{ versus } H_1 : \lambda_1(t) \neq \lambda_2(t) \text{ for, at least, one } t \in [0, \tau]$$

where λ_1 and λ_2 are the hazard functions for each of the two samples.

The logrank test statistic is defined as

$$Z(\tau) = \frac{\sum_{i=1}^{D} \left(d_{i1} - R_{i1} \frac{d_i}{R_i} \right)}{\sqrt{\sum_{i=1}^{D} \frac{R_{i1}}{R_i} \left(1 - \frac{R_{i1}}{R_i} \right) \frac{R_i - d_i}{R_i - 1} d_i}}$$

where each time t_i ($i = 1, \ldots, D$) represents the moment where an event is observed in the joint sample, d_{i1} and R_{i1} are the observed events and individuals at risk at moment t_i in the sample 1 and d_i and R_i are the observed events and individuals at risk at moment t_i in the joint sample.

One can observe that the numerator of $Z(\tau)$ represents the sum of the differences between observed and expected events under H_0 in each moment. Therefore, under H_0, this number should be small and Z follows a normal standard distribution for n big enough [11].

Efficiency is a term used in comparison of an hypothesis testing procedure. A more efficient test needs a fewer number of individuals in the sample size than a less efficient test to achieve the given power. The relative efficiency of two tests is the ratio of their efficiencies and theoretically depends on the sample size available for the given procedure. It is often possible to use the asymptotic relative efficiency, defined as the limit of the relative efficiency as the sample grows to infinity. When comparing two tests Z_1 and Z_2, Z_1 is more efficient whenever $ARE(Z_1, Z_2) > 1$ and Z_2 is more efficient otherwise. For example, an $ARE(Z_1, Z_2) = 0.5$ means that Z_1 needs the double of individuals sampled in Z_2 to achieve the same power.

Logrank test for the relevant endpoint E_1

Recall that T_1 is the time to the relevant endpoint E_1 and that individuals have been randomized to one of two groups. Let $\lambda_1^{(j)}(t)$ be the hazard function of T_1 for someone in group j ($j = 0, 1$) when there are not competing causes. The null hypothesis of no treatment differences based on the relevant relevant endpoint E_1 is stated as

$$H_0 : \lambda_1^{(0)}(\cdot) = \lambda_1^{(1)}(\cdot)$$

or, equivalently, $H_0 : S_1^{(0)}(\cdot) = S_1^{(1)}(\cdot)$, where $S_1^{(j)}(\cdot)$ is the survival function of T_1 for someone in group j ($j = 0, 1$) when there are no competing causes.

The logrank test Z, under the null hypotheses of no treatment difference, is asymptotically $\mathcal{N}(0, 1)$ [11]. The efficiency of Z is studied by examining the large sample behavior of Z when H_0 does not hold. It is not useful to think about any fixed alternative to H_0 because the power of Z will typically go to 1 as $n \to \infty$. Thus, it is considered a sequence of contiguous alternatives to H_0 which approach H_0 as $n \to \infty$. We consider $\lambda_1^{(0)}(\cdot)$ as fixed and let $\lambda_1^{(1)}(\cdot)$ vary with n defining the sequence of contiguous alternatives to H_0 as

$$H_{a,n} : \lambda_{1,n}^{(1)}(t) = \lambda_1^{(0)}(t) e^{g(t)/\sqrt{n}}$$

For any finite n, the two groups have hazard ratio at time t equal to $\log\left(\frac{\lambda_{1,n}^{(1)}(t)}{\lambda_1^{(0)}(t)}\right) = g(t)/\sqrt{n}$.

Under these conditions, [12] and [13] showed that Z is asymptotically normal with unit variance and mean μ given by

$$\mu = \frac{\int_0^\infty g(t)p(t)[1-p(t)]Pr_{H_0}\{U \geq t\}\lambda_1^{(0)}(t)dt}{\sqrt{\int_0^\infty p(t)[1-p(t)]Pr_{H_0}\{U \geq t\}\lambda_1^{(0)}(t)dt}} \tag{2.1}$$

where $g(t) = \lim_{n\to\infty} \sqrt{n} \log\left(\frac{\lambda_{1,n}^{(1)}(t)}{\lambda_1^{(0)}(t)}\right)$, U is the observed outcome given by $U = \min\{T_1, C\}$, $p(t) = Pr_{H_0}\{X = 1 | U \geq t\}$ is the null probability of, someone at risk at time t, is in treatment group 1, $Pr_{H_0}\{U \geq t\}$ is the null probability that someone is still at risk at time t and $Pr_{H_0}\{U \geq t\}\lambda_1^{(0)}(t)$ corresponds to the probability, under the null hypotheses, of observing event E_1 by time t.

Logrank test for the composite endpoint E_*

If T_2 denotes the time to the event E_2 and $T_* = \min\{T_1, T_2\}$ the time to the composite endpoint E_*, the logrank test statistic Z_* is used to test the null hypotheses of no treatment difference ($H_0^* : \lambda_*^{(0)}(\cdot) = \lambda_*^{(1)}(\cdot)$).

Analogously as above, under H_0, the logrank test statistic Z_* is asymptotically $\mathcal{N}(0, 1)$ and under a sequence of contiguous alternatives to H_0^*, it is asymptotically normal with unit variance and

mean μ_* given by

$$\mu_* = \frac{\int_0^\infty g_*(t)p_*(t)[1-p_*(t)]Pr^*_{H_0}\{U_* \geq t\}\lambda_*^{(0)}(t)dt}{\sqrt{\int_0^\infty p_*(t)[1-p_*(t)]Pr^*_{H_0}\{U_* \geq t\}\lambda_*^{(0)}(t)dt}} \tag{2.2}$$

where $g_*(t) = \lim_{n\to\infty} \sqrt{n}\log\left(\frac{\lambda_{*,n}^{(1)}(t)}{\lambda_*^{(0)}(t)}\right)$, U_* is the observed outcome given by $U_* = min\{T_1, T_2, C\}$, $p_*(t) = Pr_{H_0^*}\{X = 1|U_* \geq t\}$ is the null probability of, someone at risk at time t, is in treatment group 1, $Pr_{H_0^*}\{U \geq t\}$ is the null probability that someone is still at risk at time t and $Pr_{H_0^*}\{U_* \geq t\}\lambda_1^{(0)}(t)$ corresponds to the probability, under the null hypotheses, of observing event E_* by time t.

Asymptotic relative efficiency

The behavior of the ARE of the logrank test Z_* based on E_* versus the logrank test Z based on E_1 is used to assess the difference in efficiency between Z_* and Z. Given that both tests Z and Z_* are asymptotically $\mathcal{N}(0,1)$ under H_0 and H_0^*, respectively, and are asymptotically normal with variance 1 under a sequence of contiguous alternatives to the null hypothesis, their ARE [12] is given by:

$$\text{ARE}(\text{Z}_*, \text{Z}) = \left(\frac{\mu_*}{\mu}\right)^2 \tag{2.3}$$

where μ_* and μ are to be replaced by (2.1) and (2.2).

It is understood that the composite endpoint will be chosen whenever $ARE(Z_*, Z) > 1$.

To achieve the designated goal of deriving an expression of ARE useful for designing clinical trials, the computations of $ARE(Z_*, Z)$ would need to be based on easily interpretable parameters such as:

- The frequencies p_1 and p_2 of observing endpoint E_1 and E_2 in treatment group 0.
- The relative treatment effects on E_1 and E_2 given by hazard ratios HR_1 and HR_2.
- The degree of dependence between T_1 and T_2 given by Spearman's rank correlation coefficient ρ.

It is important to state the main assumptions being established to compute ARE and setting out the main steps to express it in terms of these interpretable parameters.

Assumptions

In the planning of the sample size of a randomized clinical trial in time-to-event study, apart from fixing type I error probability and power to detect the specific alternative, one has to set the marginal law of the survival time in the control group, the censoring distributions, the probability π of being in group 1 under the null hypothesis and the log hazard ratio. It is often assumed

that the censoring rates are the same for both groups and that the hazard ratio is constant. Thus, the mild and reasonable assumptions being adopted in [1] merely reproduce what has often been common practice in most of the clinical trials.

It is assumed that there are two independent samples and a total sample size of n individuals and that a proportion of individuals π is allocated to group 1. Efficiency calculations will be evaluated based on the following other assumptions:

- **Assumption 1**: End-of-study censoring at time τ is the only non-informative censoring cause and, without loss of generality, $\tau = 1$ is taken for computation purposes.
- **Assumption 2**: End-of-study censoring is identical across grows, that is, $Pr\{C > t|X = 0\} = Pr\{C > t|X = 1\} = Pr\{C > t\} = \mathbf{1}_{[0,\tau]}(t)$.
- **Assumption 3**: Treatment groups have proportional hazards. The proportionality assumption is given by the hazard ratios $\frac{\lambda_1^{(1)}(t)}{\lambda_1^{(0)}(t)} = HR_1$ and $\frac{\lambda_2^{(1)}(t)}{\lambda_2^{(0)}(t)} = HR_2$ for all t where $\lambda_i^{(j)}(t)$ is the hazard function of T_i for someone in treatment group j.

2.3.3 Computation of ARE for cases 1 and 3

It is here where the computations are different depending on if there are competing risks. As introduced before, in this master thesis, the censoring cases of interest are cases 1 and 3. Therefore, given the assumptions stated above and, as it is developed in [1], the non-centrality parameters μ and μ_* given in (2.1) and (2.2) can be expressed as

$$\mu = \frac{\sqrt{n\pi(1-\pi)}\int_0^1 \log\left(\frac{\lambda_1^{(1)}(t)}{\lambda_1^{(0)}(t)}\right) f_1^0(t)dt}{\sqrt{\int_0^1 f_1^{(0)}dt}} \tag{2.4}$$

$$\mu_* = \frac{\sqrt{n\pi(1-\pi)}\int_0^1 \log\left(\frac{\lambda_*^{(1)}(t)}{\lambda_*^{(0)}(t)}\right) f_*^0(t)dt}{\sqrt{\int_0^1 f_*^{(0)}dt}} \tag{2.5}$$

where $f_1^{(0)}(t)$ and $f_*^{(0)}(t)$ are the marginal density functions for T_1 and T_* in group 0.

Replacing (2.4) and (2.5) in (2.3), and $\frac{\lambda_1^{(1)}(t)}{\lambda_1^{(0)}(t)}$ by HR_1, the ARE in cases 1 and 3 is given by

$$\text{ARE}(Z_*, Z) = \left(\frac{\mu_*}{\mu}\right)^2 = \frac{\left(\int_0^1 \log\left(\frac{\lambda_*^{(1)}(t)}{\lambda_*^{(0)}(t)}\right) f_*^{(0)}(t)dt\right)^2}{(\log HR_1)^2\,(\int_0^1 f_*^{(0)}(t)dt)(\int_0^1 f_1^{(0)}(t)dt)} \tag{2.6}$$

From (2.6), the ARE expression only depends on:

- The marginal law of T_1 in group 0 ($f_1^{(0)}(t)$)

- The law of T_* in group 0 ($f_*^{(0)}(t)$) and the hazard functions of T_* in both treatment groups ($\lambda_*^{(0)}(t)$ and $\lambda_*^{(1)}(t)$)
- The hazard ratio for the relevant endpoint E_1 (HR_1)

Law of (T_1, T_2) and T_*: the use of copulas to model the bivariate survival function

The law of T_* for each group can be obtained from the bivariate distribution of (T_1, T_2) since $T_*^{(j)} = \min\{T_1^{(j)}, T_2^{(j)}\}$ for $j = 0, 1$ and

$$S_*^{(j)}(t) = Pr\{T_*^{(j)} > t\} = Pr\{\min\{T_1^{(j)}, T_2^{(j)}\} > t\} = Pr\{T_1^{(j)} > t, T_2^{(j)} > t\} = S_{(1,2)}^{(j)}(t, t) \tag{2.7}$$

Law of (T_1, T_2)

If E_1 and E_2 are independent, it has been shown that the beneficial effect of the treatment on E_* does not imply the beneficial effect on each component E_1 and E_2 [7]. On the other hand, most of the times the two endpoints E_1 and E_2 are correlated and the hazard of the composite endpoint E_* cannot be decomposed into the sum of the two marginal hazards. In this situation, the joint law of (T_1, T_2) is needed and a useful way of approaching it is specifying the joint law of (T_1, T_2) as a function of the marginal densities $f_1(t_1)$ and $f_2(t_2)$ and an association parameter θ via a copula model.

A copula is defined as a function that joins or couples multivariate distribution functions to their one-dimensional marginal distribution functions [3]. The copula parametrises the dependence between the marginals, while the parameters of each marginal distribution function can be estimated separately. A key aim of this master thesis is the study of copulas, thus further study is postponed to chapter 3.

In the methodology given by Gómez and Lagakos [1], T_1 and T_2 are assumed to be binded by Frank Archimedean survival copula, given by

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

where θ is the association parameter between T_1 and T_2. There is a bijective relationship between θ and Spearman's rank correlation ρ between T_1 and T_2.

Assuming equal association parameter θ for groups 0 and 1, the joint survival and joint density functions for (T_1, T_2) in group j ($j = 0, 1$) are given by

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

$$f_{(1,2)}^{(j)}(t_1, t_2; \theta) = \frac{\theta}{e^{-\theta} - 1} \frac{e^{-\theta(S_1^{(j)}(t_1) + S_2^{(j)}(t_2))}}{e^{-2\theta S_{(1,2)}^{(j)}(t_1, t_2; \theta)}} [f_1^{(j)}(t_1)][f_2^{(j)}(t_2)]$$

where $S_1^{(j)}(t_1)$ and $f_1^{(j)}(t_1)$, $S_2^{(j)}(t_2)$ and $f_2^{(j)}(t_2)$ are the survival and marginal densities of T_1 and T_2, respectively, in group j.

Law of T_*

As seen in (2.7), $S_*^{(j)}(t;\theta) = S_{(1,2)}^{(j)}(t,t;\theta)$, and having $f_*^{(j)}(t;\theta) = -\partial S_*^{(j)}(t;\theta)/\partial t$, we have

$$S_*^{(j)}(t;\theta) = \frac{-1}{\theta}\log\left(1 + \frac{(e^{-\theta S_1^{(j)}(t)} - 1)(e^{-\theta S_2^{(j)}(t)} - 1)}{e^{-\theta} - 1}\right)$$

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

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

Hence, in order to compute ARE(Z_*, Z) assuming Frank copula, we need to specify:

- $f_1^{(j)}(t)$ and $S_1^{(j)}(t)$: The marginal density and survival functions of T_1 in group j ($j = 0, 1$)
- $f_2^{(j)}(t)$ and $S_2^{(j)}(t)$: The marginal density and survival functions of T_2 in group j ($j = 0, 1$)
- θ: The copula association parameter between T_1 and T_2.
- HR_1: The constant hazard ratio of T_1, $HR_1 = \lambda_1^{(1)}(t)/\lambda_1^{(0)}(t)$

Choice of marginal laws for the computations

To derive the ARE(Z_*, Z) in terms of the above listed interpretable parameters, we have to specify marginal parametric laws for $T_1^{(j)}$ and $T_2^{(j)}$ for both treatment groups 0 and 1 and we have to relate their parameters to the frequencies p_1 and p_2, the hazard ratios HR_1 and HR_2 and the Spearman's coefficient ρ.

Gómez and Lagakos [1] chose Weibull distributions for the endpoints T_1 and T_2 with scale parameters $b_1^{(j)}$ and $b_2^{(j)}$ for groups $j = 0, 1$ and shape parameters β_1 and β_2 chosen equal for both groups so that the proportionality of the hazards holds.

In this case, the survival function $S_k^{(j)}(t)$ is defined by $S_k^{(j)}(t) = \exp\left(-(t/b_k^{(j)})^{\beta_k}\right)$ and, then, there is a direct relationship between $(f_1^{(j)}, S_1^{(j)}, f_2^{(j)}, S_2^{(j)}, \theta, HR_1)$ and $(b_1^{(0)}, b_2^{(0)}, b_1^{(1)}, b_2^{(1)}, \beta_1, \beta_2, \theta, HR_1)$. These parameters can be related to the parameters of interest $(p_1, p_2, \beta_1, \beta_2, \rho, HR_1, HR_2)$ by means of:

- The scale parameters $b_1^{(0)}$ and $b_2^{(0)}$ are functions of p_1, p_2, β_1 and β_2 and, depending on the censoring case, are given by
  - For case 1, $b_1^{(0)}$ and $b_2^{(0)}$ are given by

$$b_1^{(0)} = \frac{1}{(-\log(1-p_1))^{1/\beta_1}} \text{ and } b_2^{(0)} = \frac{1}{(-\log(1-p_2))^{1/\beta_2}}.$$

– For case 3, $b_1^{(0)}$ and $b_2^{(0)}$ are given by

$$b_1^{(0)} = \frac{1}{(-\log(1-p_1))^{1/\beta_1}}$$

$b_2^{(0)}$ is found as the solution of $p_2 = \int_0^1 \int_v^1 f_{(1,2)}^{(0)}(u, v; \theta) du dv$.

- For $k = 1, 2$, the scale parameter $b_k^{(1)}$ is function of the scale parameter $b_k^{(0)}$, the shape parameter β_k and the hazard ratio HR_k as follows: $b_k^{(1)} = \frac{b_k^{(0)}}{HR_k^{\frac{1}{\beta_k}}}$

- The independence parameter θ is a function of ρ for the Frank copula case given by

$$\rho = \rho(\theta) = 1 - \frac{12}{\theta}[\frac{1}{\theta}\int_0^\theta \frac{t}{e^t - 1} - \frac{2}{\theta^2}\int_0^\theta \frac{t^2}{e^t - 1} dt]$$

2.3.4 Cases studies

Gómez and Lagakos [1] and Moises Gómez [2] made some simulation studies with different values for β_1, β_2, p_1, p_2, HR_1, HR_2 and ρ in order to develop statistical guidelines to help physicians decide whether or not use a composite endpoint when designing a clinical trial. It is understood that the composite endpoint will be chosen whenever $ARE(Z_*, Z) > 1$. However, it is possible to recommend a more flexible rule such that the composite endpoint is always used if $ARE(Z_*, Z) > 1.25$ and it is never used if $ARE(Z_*, Z) < 1.1$. In that case, for values of ARE between 1.1 and 1.25 it is not clear which endpoint should be used. Gómez and Lagakos [1] used the rule to choose the composite endpoint if $ARE(Z_*, Z) > 1.1$ and the relevant endpoint otherwise.

The values in [1] for cases 1 and 3 were 0.5, 1 and 2 for β_1 and β_2, representing decreasing, constant and increasing hazard functions, respectively. The possible values for p_1 and p_2 were 0.05, 0.15, 0.30 and 0.50 while the values of HR_1 and HR_2 were $HR_1 = 0.5$ (with $HR_2 = 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$) and $HR_1 = 0.7$ (with $HR_2 = 0.5, 0.6, 0.7, 0.8, 0.9, 1$). The values for Spearman's rank correlation ranged from $\rho = 0.15$ to $\rho = 0.75$ in order to consider weak, moderate and strong correlations between T_1 and T_2. A total of 6048 combinations were studied for both censoring case 1 and case 3. For each such case, these combinations are grouped for specific values of β_1, β_2, p_1, p_2 and HR_1 yielding a total of 144 scenarios for each censoring case (Table 2.1).

A further analysis of the results can be found in [1] but the general pattern is similar for cases 1 and 3: ARE decreases when the Spearman's rank correlation between the two endpoints increases; and also when the relative effect of treatment on the additional endpoint is smaller than that on the relevant endpoint.

On the other hand, it has been made a search of all randomized clinical trials on cardiovascular disease carried out in 2008 [2]. The search was restricted to randomized clinical trials using the

β_1	0.5	1	2				
β_2	0.5	1	2				
(p_1, p_2)	(0.05,0.05)	(0.15,0.05)	(0.15,0.15)	(0.3,0.3)	(0.5,0.3)	(0.5,0.5)	
(p_1, p_2)	(0.05,0.15)	(0.3,0.5)	(0.05,0.3)	(0.15,0.3)			
ρ	0.15	0.25	0.35	0.45	0.55	0.65	0.75
$HR_1 = 0.5$ and $HR_2 =$	0.3	0.4	0.5	0.6	0.7	0.8	
$HR_1 = 0.7$ and $HR_2 =$	0.5	0.6	0.7	0.8	0.9	1	

Table 2.1: Possible values simulated in Gómez and Lagakos [1]

logrank test for comparison between the two groups and it has been observed if the main outcome was a simple or composite endpoint. From here, it has been studied in which cases a more efficient design of clinical trials could have been done and it has been observed that it is necessary to follow some sort of criteria in order to avoid designing the trials wrongly. Preliminary results showed no differences between different values of β_1 and β_2 considering decreasing, constant and increasing hazards and hence $\beta_1 = \beta_2 = 1$ was considered. The values for p_1, p_2, HR_1, HR_2 and ρ were the values shown in Table 2.2. In that case, the threshold for choosing the single o composite endpoint was set to $ARE(Z_*, Z) > 1$ or $ARE(Z_*, Z) \leq 1$. Among the conclusions of this study it is worth mentioning that composite endpoints should always be used if $HR_1 \geq 0.8$ and $HR_2 \leq 0.35$. On the other hand, values of HR_1 chosen to 0.5 and $HR_2 \geq 0.55$ often provide ARE values near . A total of 48 scenarios cases were studied for censoring case 3 in this work.

β_1	1						
β_2	1						
p_1	0.035	0.05	0.09	0.125			
p_2	0.1	0.15	0.2				
ρ	0.15	0.25	0.35	0.45	0.55	0.65	0.75
HR_1	0.5	0.6	0.7	0.8			
HR_2	0.4	0.5	0.6	0.7	0.8	0.9	

Table 2.2: Possible values simulated in Moisés Gómez's master thesis [2]

The general guidelines for using a composite endpoint can be checked in [2] but they are similar to the general rules obtained in [1].

2.4 Applicability of this method

In this methodology presented by Gómez and Lagakos [1], it is assumed that T_1 and T_2 follow a Weibull distribution and the copula used for the joint law is the Frank Archimedean copula.

As said above, it is crucial to make this methodology the widest and most applicable possible studying and testing other families of copulas and marginal laws.

Therefore, the copula's theory will be studied (chapter 3) and we will try the methodology for Gumbel copula for censoring cases 1 and 3 (chapter 4). It is left for future research the censoring cases 2 and 4, with competing risks, and the study of other families of copulas and marginal distribution laws.

Chapter 3

Copulas

In this chapter, an introduction to copulas will be presented summarizing Nelsen (1999) [3] and Trivedi and Zimmer (2007) [4]. This is a summary of those concepts which, from my point of view, are necessary to understand the main aim of this master thesis. Theorems are stated without proofs and further information can be found in [3] and [4].

3.1 Introduction

The study of copulas in statistics is a modern phenomenon that started in the decade of the nineties with the interest of studying distributions with *fixed* or *given* marginal distributions. Copulas are defined as functions that join or couple multivariate distribution functions to their one-dimensional marginal distribution functions. According to Fisher (1997) [14], copulas are of interest to statisticians for two main reasons: firstly, as a way of studying scale-free measures of dependence; and secondly, as a starting point for constructing families of multivariate distributions, sometimes with a view to simulation.

The word *copula* is a latin noun that means "link, tie, bond" [15] and it is used in grammar and logic to describe "that part of a proposition which connects the subject and the predicate" [16]. A brief history of the development and study of copulas can be found in [3] but, in a few words, the word *copula* was first used in a mathematical sense in the theorem that describes the functions which "join together" one-dimensional distribution functions to form multivariate distribution functions. This theorem was proposed by Abe Sklar in 1959 but the functions themselves existed before the use of the term copula because they appeared in the study of multivariate distribution with fixed univariate marginal distributions made by Fréchet, Dall'Aglio, Féron, Hoeffding and many others. On the other hand, the earliest paper explicitly relating copulas to the study of dependence among random variables appears in 1981. In that paper, Schweizer and Wolff discussed and modified a criteria for measures of dependence between pairs of random variables but copulas had already appeared implicitly in earlier work on dependence by many other authors.

The main objective of this master thesis is to study the joint distribution function of two random variables and, therefore, all definitions and results take into account bivariate copulas. An extension to multivariate copulas can be found in [3].

3.2 Definitions and Basic Properties

Consider a pair of random variables X and Y, with distribution functions $F(x) = Pr\{X \leq x\}$ and $G(y) = Pr\{Y \leq y\}$, respectively, and a joint distribution function $H(x, y) = Pr\{X \leq x, Y \leq y\}$. For each pair of real numbers (x, y) we can associate three numbers: $F(x)$, $G(y)$ and $H(x, y)$, each of them lying in the interval [0,1]. Each pair (x, y) leads to a point $(F(x), G(y))$ in the unit square. In plane words, a copula is the function that links the value of the joint distribution $H(x, y)$ to each pair of values of the marginal distribution functions $(F(x), G(y))$.

3.2.1 Preliminaries

A 2-place real function H is a function whose domain, Dom$H$, is a subset of $\overline{\mathbb{R}}^2$ and whose range, Ran$H$, is a subset of $\mathbb{R}$, where $\mathbb{R} = (-\infty, \infty)$ and $\overline{\mathbb{R}} = [-\infty, \infty]$.

Definition 3.2.1. *Let S_1 and S_2 be nonempty subsets of $\overline{\mathbb{R}}$ and let H be a function such that Dom$H$ $= S_1 \times S_2$. Let $B = [x_1, x_2] \times [y_1, y_2]$ a rectangle in Dom$H$. Then the H-volume of B is given by*

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1) \tag{3.1}$$

A 2-place real function H is 2-increasing if $V_H(B) \geq 0$ for all rectangle B in Dom$H$. This definition could be understood as a two-dimensional analog of a nondecreasing function of one variable but it is important to remark that "H is two-increasing" neither implies nor is implied by "H is nondecreasing in each argument".

A function H from $S_1 \times S_2$ is said to be grounded if $H(x, a_2) = H(a_1, y) = 0$ for all $(x, y) \in S_1 \times S_2$ where a_1 and a_2 are the least elements of S_1 and S_2, respectively. In this case, the property of 2-increasing of H implies that H is nondecreasing in each argument.

Analogously, a function H from $S_1 \times S_2$ is said to have marginals, that are functions F and G, if

$$\text{Dom}F{=}S_1 \text{ and } F(x) = H(x, b_2)$$
$$\text{Dom}G{=}S_2 \text{ and } G(y) = H(b_1, y)$$

where b_1 and b_2 are the greatest elements of S_1 and S_2, respectively.

3.2.2 Copulas

Definition 3.2.2. *Let I be the interval $[0, 1]$. A two-dimensional copula is a function C from I^2 to I with the following properties:*

- $\forall u, v \in I$, $C(u, 1) = u$ *and* $C(1, v) = v$.
- $\forall u, v \in I$, $C(u, 0) = 0 = C(0, v)$
- C *is 2-increasing.*

This is the formal definition of a copula. Let us now see some properties and basic results about copulas. First of all, let us study the bounds of these functions.

Theorem 3.2.3. *Given $u, v \in I$, any copula C satisfies*

$$\max\{u + v - 1, 0\} \leq C(u, v) \leq \min\{u, v\}$$

It is trivial to prove that $W(u, v) = \max\{u + v - 1, 0\}$ and $M(u, v) = \min\{u, v\}$ are also copulas and they are called the *Fréchet-Hoeffding bounds*. Another example of a copula is the product copula, defined by $\prod(u, v) = u \times v$.

The graph of any copula is a continuous surface within I^3 whose boundary is the skew quadrilateral with vertices $(0, 0, 0), (1, 0, 0), (0, 1, 0)$ and $(1, 1, 1)$. Given any copula C and $t \in I$, the graph of the level set is defined as $\{(u, v) \in I^2 | C(u, v) = t\}$. Figure 3.1 shows the graph and level sets for some t for copulas $\prod(u, v)$, $M(u, v)$ and $W(u, v)$. The graph of $\prod(u, v)$ lies between the graphs of $M(u, v)$ and $W(u, v)$ and the level set of $\prod(u, v)$ for a given t must lie in the shaded triangle shown in Figure 3.2, whose boundaries are the level sets determined by $M(u, v) = t$ and $W(u, v) = t$. Since $M(u, v)$ and $W(u, v)$ are the bounds for any copula, these properties holds for any copula C, and not only for $\prod(u, v)$.

3.2.3 Joint distributions

The joint distribution of a set of random variables $(Y_1, \ldots, Y_m)$ is defined as $F(y_1, \ldots, y_m) = Pr\{Y_i \leq y_i; i = 1, \ldots, m\}$ and the survival function is given by $\overline{F}(y_1, \ldots, y_m) = Pr\{Y_i > y_i; i = 1, \ldots, m\}$. Let us focus on the 2-dimensional case.

The following conditions are necessary and sufficient for a right-continuous function F to be a bivariate distribution function:

1. $\lim_{y_j \to -\infty} F(y_1, y_2) = 0$ for $j = 1, 2$
2. $\lim_{y_j \to \infty \forall j} F(y_1, y_2) = 1$
3. For all (a_1, b_1), (a_2, b_2) with $a_1 \leq a_2$ and $b_1 \leq b_2$, $F(a_2, b_2) - F(a_1, b_2) - F(a_2, b_1) + F(a_1, b_1) \geq 0$.

Conditions 1 and 2 imply $0 \leq F \leq 1$. Condition 3 is referred to as the property that F is 2-increasing, as seen in (3.1). If F has second derivatives, then the 2-increasing property is equivalent to $\partial^2 F / \partial y_1 \partial y_2 \geq 0$.

Given a bivariate distribution function F, the marginal distribution functions F_1 and F_2 are $F_1(y_1) = \lim_{y_2 \to \infty} F(y_1, y_2)$ and $F_2(y_2) = \lim_{y_1 \to \infty} F(y_1, y_2)$.

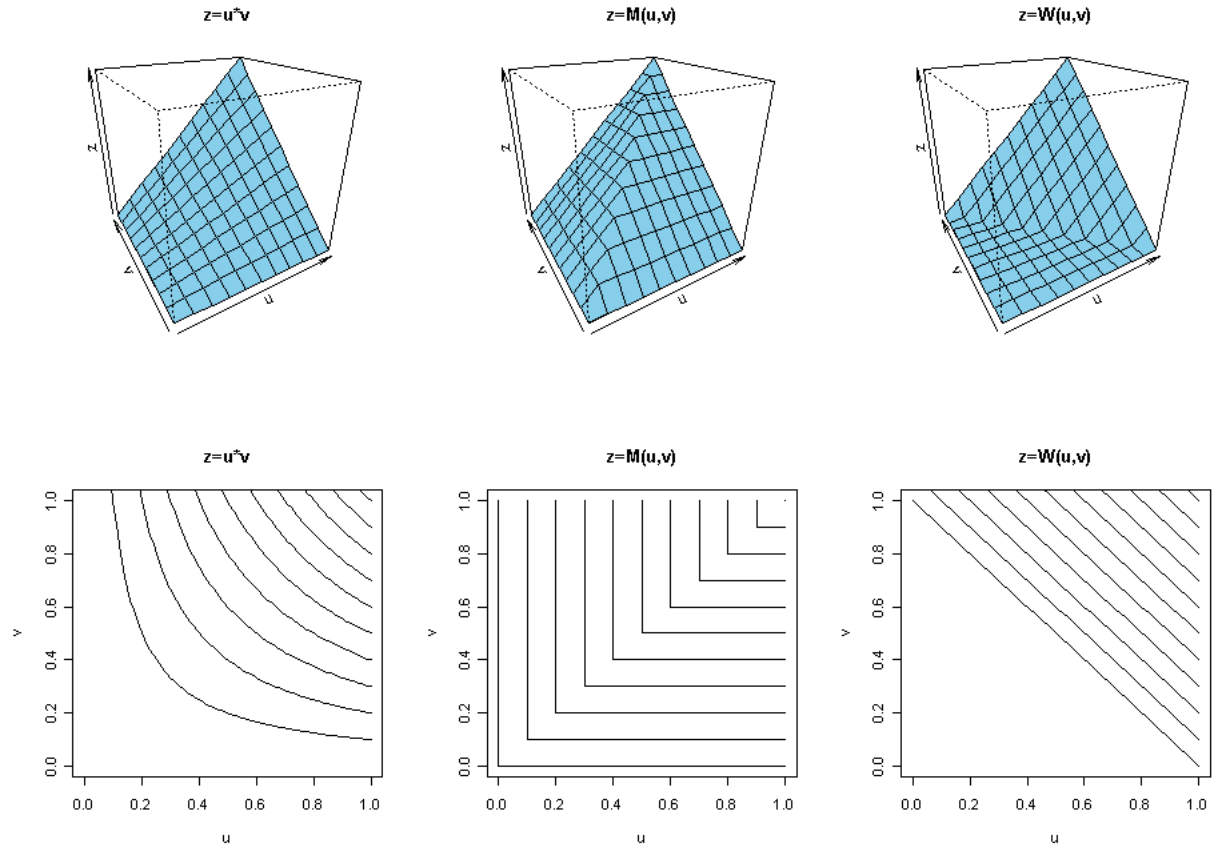


Figure 3.1: Graphs and level sets of $\prod(u, v)$, $M(u, v)$ and $W(u, v)$

3.2.4 Sklar's Theorem

Sklar's theorem elucidates the role that copulas play in the relationship between multivariate distribution functions and their univariate marginals. As we saw in definition 3.2.2, properties 2 and 3 are general properties of multivariate distribution functions. Thus, it follows that a copula can be defined as a bivariate distribution function whose support is contained in I^2 and whose marginals are uniform on I.

Theorem 3.2.4 (Sklar's theorem)**.** *Let H be a joint distribution function with marginals F and G. Then there exists a copula C such that $\forall x, y \in \overline{\mathbb{R}}$,*

$$H(x, y) = C(F(x), G(y))$$

If F and G are continuous, then C is unique; otherwise, C is uniquely determined on $RanF \times RanG$. Conversely, if C is a copula and F and G distribution functions, then the function H defined by $H(x, y) = C(F(x), G(y))$ is a joint distribution function with marginals F and G.

$H(x, y) = C(F(x), G(y))$ gives an expression for joint distribution functions in terms of a copula and two univariate functions. It can be inverted to express copulas in terms of a joint distribution

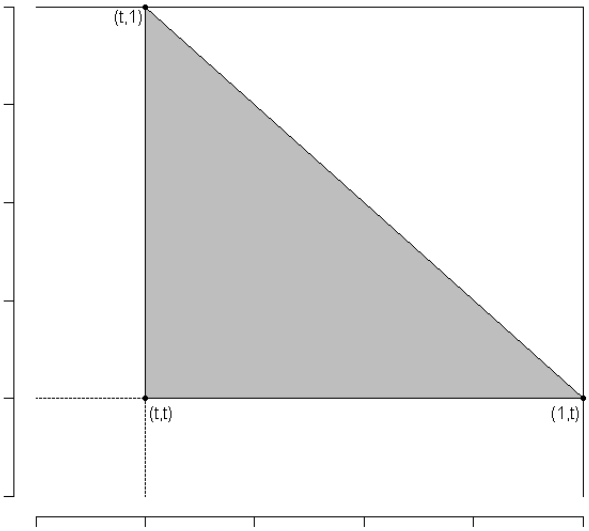


Figure 3.2: Graphs of the level set $\{(u,v) \in I^2 | C(u,v) = t\}$ of $M(u,v)$ and $W(u,v)$

function and the "inverses" of the two marginals (if a marginal is not strictly increasing, it does not have an inverse in the usual sense and a quasi-inverse must be defined).

Definition 3.2.5. *Let F be a distribution function. Then a quasi-inverse of F is a function $F^{(-1)}$ with domain I such that*

1. *If t is in $RanF$, then $F^{(-1)}(t)$ is any number $x \in \overline{\mathbb{R}}$ such that $F(x) = t$.*
2. *If t is not in $RanF$, then $F^{(-1)}(t) = inf\{x|F(x) \geq t\} = sup\{x|F(x) \leq t\}$*

If F is strictly increasing, $F^{(-1)}(t) = F^{-1}(t)$

Corollary 3.2.6. *Let H be a joint distribution function with marginals F and G and let C be a copula such that $H(x,y) = C(F(x), G(y))$. Let $F^{(-1)}$ and $G^{(-1)}$ be quasi-inverses of F and G. Then $\forall (u,v) \in DomC$, $C(u,v) = H(F^{(-1)}(u), G^{(-1)}(v))$*

The relation between distribution functions and copulas has been stated above. Let us denote $C(F(X), G(Y))$ by C_{XY} and let us now see some properties.

Theorem 3.2.7. *Let X and Y be continuous random variables with distribution functions F and G. Then X and Y are independent if and only if $C_{XY} = \prod$, i.e. $C(F(x), G(y)) = F(x)G(y)$.*

Theorem 3.2.8. *Let X and Y be continuous random variables with copula C_{XY}. Assume that α and β are strictly monotone on RanX and RanY respectively, the following properties relate C_{XY} with $C_{\alpha(X)\beta(Y)}$ when $\alpha(X)$ and $\beta(Y)$ are the corresponding transformed random variables:*

1. *If α and β are both strictly increasing, then $C_{\alpha(X)\beta(Y)} = C_{XY}$.*
2. *If α is strictly increasing and β strictly decreasing, then $C_{\alpha(X)\beta(Y)}(u, v) = u - C_{XY}(u, 1 - v)$.*
3. *If α is strictly decreasing and β strictly increasing, then $C_{\alpha(X)\beta(Y)}(u, v) = v - C_{XY}(1 - u, v)$.*
4. *If α and β are both strictly decreasing, then $C_{\alpha(X)\beta(Y)} = u + v - 1 + C_{XY}(1 - u, 1 - v)$.*

The *Fréchet-Hoeffding bounds* can be defined for joint distribution functions of random variables as $\max\{F(x) + G(y) - 1, 0\} \leq H(x, y) \leq \min\{F(x), G(y)\}$. Now we can introduce some definitions and lemmas preceding a theorem related to these bounds.

Definition 3.2.9. *A subset S of $\overline{\mathbb{R}}^2$ is nondecreasing if $\forall (x, y), (u, v) \in S$, $x < u$ implies $y \leq v$. Similarly, a subset S of $\overline{\mathbb{R}}^2$ is nonincreasing if $\forall (x, y), (u, v) \in S$, $x < u$ implies $y \geq v$.*

Lemma 3.2.10. *Let S be a subset of $\mathbb{R}^2$. Then S is nondecreasing if and only if $\forall (x, y) \in \mathbb{R}^2$, either*

1. *$\forall (u, v) \in S$, $u \leq x$ implies $v \leq y$; or*
2. *$\forall (u, v) \in S$, $v \leq y$ implies $u \leq x$*

Lemma 3.2.11. *Let X and Y be random variables with joint distribution function H. Then H is equal to its Fréchet-Hoeffding upper bound if and only if $\forall (x, y) \in \overline{\mathbb{R}}^2$, either $P[X > x, Y \leq y] = 0$ or $P[X \leq x, Y > y] = 0$.*

Theorem 3.2.12. *Let X and Y be random variables with joint distribution function H. Then,*

- *H is identically equal to its Fréchet-Hoeffding upper bound if and only if the support of H is a nondecreasing subset of $\overline{\mathbb{R}}^2$.*
- *H is identically equal to its Fréchet-Hoeffding lower bound if and only if the support of H is a nonincreasing subset of $\overline{\mathbb{R}}^2$.*

3.2.5 Survival Copulas

The probability of an individual living or surviving beyond time x is given by the survival function $\overline{F(x)} = Pr\{X > x\} = 1 - F(x)$. The natural range of a random variable is often $[0, +\infty)$; however, we will use the term "survival function" for $Pr\{X > x\}$ even when the range is $\overline{\mathbb{R}}$.

For a pair (X, Y) of random variables with joint distribution H, the joint survival function is given by $\overline{H}(x, y) = Pr\{X > x, Y > y\}$. Then the marginals of $\overline{H}$ are the functions $\overline{H}(x, -\infty)$ and $\overline{H}(-\infty, y)$

which are the univariate survival functions $\overline{F}$ and $\overline{G}$.

We can now try to find a relationship between univariate and joint survival functions. Let C be a copula of F and G. Then

$$\begin{aligned}\overline{H}(x,y) &= Pr\{X > x, Y > y\} = 1 - Pr\{X \le x\} - Pr\{Y \le Y\} + Pr\{X \le x, Y \le y\} = \\ &= 1 - F(x) - G(y) + H(x,y) = \overline{F}(x) + \overline{G}(y) - 1 + C(F(x), G(y)) = \\ &= \overline{F}(x) + \overline{G}(y) - 1 + C(1 - \overline{F}(x), 1 - \overline{G}(y))\end{aligned}$$

We can define $\widehat{C}(u,v) = u + v - 1 + C(1-u, 1-v)$ and $\overline{H}(x,y) = \hat{C}(\overline{F}(x), \overline{G}(y))$. Then $\widehat{C}$ is a copula and we refer to it as *survival copula* of X and Y.

It is easy to check the three conditions for $\widehat{C}$ to be a copula following that C is a copula. Indeed,

1. $\widehat{C}(u,1) = u + C(1-u,0) = u + 0 = u$
   $\widehat{C}(1,v) = v + C(0,v) = v + 0 = v$
2. $\widehat{C}(u,0) = u - 1 + C(1-u,1) = u - 1 + (1-u) = 0$
   $\widehat{C}(0,v) = v - 1 + (1-v) = 0$
3. For $0 \le u_1 < u_2 \le 1$ and $0 \le v_1 < v_2 \le 1$,
   $\widehat{C}(u_2,v_2) - \widehat{C}(u_2,v_1) - \widehat{C}(u_1,v_2) + \widehat{C}(u_1,v_1) =$
   $= C(1-u_2, 1-v_2) - C(1-u_2, 1-v_1) - C(1-u_1, 1-v_2) + C(1-u_1, 1-v_1) \ge 0$
   since C is 2-increasing and $0 \le 1-u_2 < 1-u_1 \le 1$ and $0 \le 1-v_2 < 1-v_1 \le 1$.

There are other functions of interest, like the *dual of a copula*, $\widetilde{C}(u,v) = u+v-C(u,v)$; or the *co-copula*, $C^*(u,v) = 1 - C(1-u, 1-v)$. They are functions that are not copulas but express probabilities. Especifically, $Pr\{X \le x \text{ or } Y \le y\} = \widetilde{C}(F(x), G(y))$ and $Pr\{X > x \text{ or } Y > y\} = C^*(\overline{F}(x), \overline{G}(y))$.

3.3 Some Common Bivariate Copulas

Copulas can be used to express a multivariate distribution in terms of its marginal distributions because copulas allow researchers to piece together joint distributions when only marginal distributions are known with certainty. In the bivariate case, the copula associated with H is a distribution function C from $[0,1]^2$ to $[0,1]$ that satisfies

$$H(x,y) = C(F(x), G(y); \theta)$$

where θ is a parameter of the copula called the **dependence parameter**, which measures dependence between the marginals. It could be said that each family of copulas define a concrete dependence structure and θ measures the intensity of this dependence. Five examples of common bivariate copulas are introduced in this section and the dependence properties are stated for each copula. However, a more detailed discussion of dependence is given in section 3.4.

3.3.1 Product copula

As seen above, the product copula has the form

$$C(u, v) = uv$$

where u and v take values in the unit interval of the real line and it corresponds to independence between the two marginal random variables.

3.3.2 Farlie-Gumbel-Morgenstern copula

The Farlie-Gumbel-Morgenstern (FGM) copula takes the form

$$C(u, v; \theta) = uv(1 + \theta(1 - u)(1 - v))$$

This copula is a perturbation of the product copula; if the dependence parameter θ equals zero, then the FGM copula collapses to independence. It is restrictive because this copula is only useful when dependence between the two marginals is modest in magnitude.

3.3.3 Frank copula

The Frank copula takes the form

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

The dependence parameter may assume any real value $(-\infty, \infty)$ and values of $-\infty$, 0 and ∞ correspond to the Fréchet-Hoeffding lower bound, independence and upper bound, respectively. It permits negative dependence between the maginals and it is most appropriate for data that exhibits weak tail dependence.

3.3.4 Gumbel copula

The Gumbel copula takes the form

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

The dependence parameter is restricted to the interval $[1, \infty)$ and values of 1 and ∞ correspond to independence and the Fréchet-Hoeffding upper bound. Gumbel copula does not allow negative dependence and it exhibits strong right tail dependence and relatively weak left tail dependence.

3.3.5 Clayton copula

The Clayton copula takes the form

$$C(u, v; \theta) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$$

with the dependence parameter θ restricted to the region $(0, \infty)$. As θ approaches zero, the marginals become independent. On the other hand, as θ approaches infinity, the copula attains the Fréchet-Hoeffding upper bound. The Clayton copula cannot account for negative dependence and it exhibits strong left tail dependence and relatively weak right tail dependence.

To summarize, Table 3.1 contains the possible direction for the dependence of these copulas and the strength of its tail dependence.

Copula	Dependence	Left Tail	Right Tail
Product	independence	weak	weak
FGM	modest in magnitude	weak	weak
Frank	positive and negative	weak	weak
Gumbel	positive	weak	strong
Clayton	positive	strong	weak

Table 3.1: Dependence and Tail dependence for some common bivariate copulas.

3.4 Dependence

As said above, the concept of dependence is defined in this section. Moreover, section 3.4.4 shows the dependence structure of each copula from a graphical point of view.

The random variables X and Y are said to be dependent or associated if they are not independent in the sense that $H(X, Y) \neq F(X)G(Y)$ where H is the joint distribution function of (X, Y) and F and G are the marginal distribution functions of X and Y.

3.4.1 Correlation

Association can be measured using several alternative concepts but, by far, the most familiar association dependence is the **Pearson correlation coefficient** between a pair of variables (X, Y), defined as

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

where $Cov(X, Y) = E[XY] - E[X]E[Y]$ and $\sigma_X, \sigma_Y > 0$ are the standard deviations of X and Y, respectively. It is well known that ρ_{XY} measures linear dependence, so it is necessary to consider

alternative measures of dependence between X and Y (see [4] for more details).

3.4.2 Concordance

Another measure of dependence is the concordance. A pair of random variables (X, Y) are concordant if large values of one tend to be associated with large values of the other, and small values of one with small values of the other. Two measures of concordance are **Kendall's rank correlation** (Kendall's τ) or **Spearman's rank correlation** (Spearman's ρ). For the following definitions, let X and Y be two random variables with marginal and joint distribution functions F, G and H, respectively. Therefore, $H(X, Y) = C(F(X), G(Y))$.

Kendall's τ

Let (X_1, Y_1) and (X_2, Y_2) be independent and identically distributed pairs of random variables, each with joint distribution function H. The Kendall's rank correlation is defined as the probability of concordance minus the probability of discordance, as follows

$$\tau_{XY} = Pr\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - Pr\{(X_1 - X_2)(Y_1 - Y_2) < 0\}.$$

Spearman's ρ

Let (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) be independent and identically distributed pairs of random variables, each with joint distribution function H. The Spearman's rank correlation is defined as the probability of concordance minus the probability of discordance of the two vectors (X_1, Y_1) and (X_2, Y_3), as follows

$$\rho_{XY} = Pr\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - Pr\{(X_1 - X_2)(Y_1 - Y_3) < 0\}.$$

Another way to define Spearman's ρ is as the Pearson correlation coefficient between the ranked variables.

Relation between copulas and measures of concordance

Copulas play an important role in concordance and measures of association. Let X and Y be two random variables with marginals F and G. Given any copula $C(u, v)$ such that $H(X, Y) = C(F(X), G(Y))$, the relation between C and these measures is defined as follows [3]:

$$\rho_{XY} = 12 \int_0^1 \int_0^1 [C(u, v) - uv] du dv$$

$$\tau_{XY} = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1$$

There is a list of the copulas introduced above and their concordance measures in Table 3.2.

Copula type	$C(u,v;\theta)$	θ-domain	τ	ρ
Product	uv	-	0	0
FGM	$uv(1+\theta(1-u)(1-v))$	$[-1,1]$	$\frac{2}{9}\theta$	$\frac{1}{3}\theta$
Clayton	$\max\{[u^{-\theta}+v^{-\theta}-1]^{-1/\theta},0\}$	$[-1,\infty)\backslash\{0\}$	$\frac{\theta}{\theta+2}$	-
Frank	$-\frac{1}{\theta}\log\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{(e^{-\theta}-1)}\right)$	$(-\infty,\infty)$	$1-\frac{4}{\theta}(1-D_1(\theta))$	$1-\frac{12}{\theta}(D_1(\theta)-D_2(\theta))$
Gumbel	$\exp\left(-[(-\log(u))^\theta+(-\log(v))^\theta]^{1/\theta}\right)$	$[1,\infty)$	$\frac{\theta-1}{\theta}$	-

FGM is the Farlie-Gumbel-Morgenstern copula.
$D_k(x)$ denotes the "Debye" function $k/x^k\int_0^x \frac{t^k}{(e^t-1)}dt$, $k=1,2$.
- denotes that there is not an analytical expression.

Table 3.2: List of Common Bivariate Copulas

Relation between ρ and τ

While both Kendall's τ and Spearman's ρ measure the probability of concordance between random variables with a given copula, the values of ρ and τ are often quite different. This is important in this master thesis because there is a straightforward relation between ρ and θ for Frank copula but it is not for Clayton or Gumbel copulas. Therefore, it is not possible to obtain the dependence parameter from a given ρ in these cases and it is necessary to obtain θ using numerical approximations or obtain it using the relation with Kendall's τ. One could imagine that it is easier to do it using Kendall's τ and then find a relation between ρ nd τ but, as stated above, it is not always true that these two values are similar. Nelsen [3] presents the exact relation between these two measures for members of some of the families of copulas but there are general patterns that work for all the cases.

Theorem 3.4.1. *Let X and Y be continuous random variables, and let τ and ρ denote Kendall's τ and Spearman's ρ, respectively. Then, $-1 \leq 3\tau - 2\rho \leq 1$.*

The relationship between τ and ρ can also be shown by the following pair of inequalities:

$$\frac{3\tau-1}{2} \leq \rho \leq \frac{1+2\tau-\tau^2}{2} \text{ if } \tau \geq 0$$

$$\frac{\tau^2+2\tau-1}{2} \leq \rho \leq \frac{1+3\tau}{2} \text{ if } \tau \leq 0$$

These bounds for the values of τ and ρ are illustrated in Figure 3.3. For any pair X and Y of continuous random variables, the values of the population measures of Kendall's τ and Spearman's ρ must lie in the shaded region.

A further study of the relationship between these two variables can be found in [17], where better approximations of the bounds of this relationship are summarized depending on the joint distribution function and it is introduced that, under mild regularity conditions, the limit of the ratio ρ/τ is 3/2 as the joint distribution function of the random variables approaches independence.

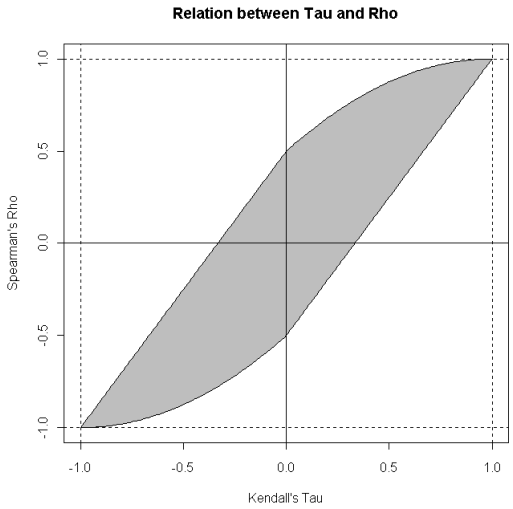


Figure 3.3: Graph of the bounds for ρ and τ for pairs of continuous random variables.

3.4.3 Tail dependence

In some cases the concordance between extreme (tail) values of random variables is of interest.

The tail dependence measure can be defined in terms of the joint survival function $S(u, v)$ for standard uniform random variables u and v. Especifically, λ_L and λ_U are measures of lower and upper tail dependence, respectively, defined by

$$\lambda_L = \lim_{v \to 0^+} \frac{C(v, v)}{v}$$

$$\lambda_U = \lim_{v \to 1^-} \frac{S(v, v)}{1 - v}$$

where the expression $S(v, v) = Pr\{U_1 > v, U_2 > v\}$ represents the joint survival function where $U_1 = F_1^{-1}(X)$ and $U_2 = G_1^{-1}(Y)$. The upper tail dependence measure λ_U is the limiting value of $S(v, v)/(1 - v)$, which is the conditional probability $Pr\{U_1 > v | U_2 > v\}$ $(= Pr\{U_2 > v | U_1 > v\})$. The lower tail dependence measure λ_L is the limiting value of the conditional probability $Pr\{U_1 < v | U_2 < v\}$ $(= Pr\{U_2 < v | U_1 < v\})$.

3.4.4 Visual illustration of dependence

One way of visualizing copulas is to present contour diagrams with graphs of level sets defined as the sets in I^2 given by $C(u, v) = a$ where a is a constant. Graphs in Figure 3.1 show level curves for the upper and lower bounds and the product copula. The shadowed triangle in Figure 3.2 gives the bounds for the level set $\{(u, v) \in I^2 | C(u, v) = t\}$ of any copula, determined by the lower and upper bounds, $W(u, v)$ and $M(u, v)$, respectively.

Unfortunately, this is not always a helpful way of visualizing the data patterns implied by different copulas. One alternative is to present two-way scatter diagrams of realizations from simulated draws from copulas. These scatter diagrams are quite useful in illustrating tail dependence in a bivariate context.

Following [3], Figure 3.4 shows the scatter plots for the different copulas seen in section 3.3. The process of generating this simulated draws is the conditional distribution method. We need to generate a pair (u, v) of observations of uniform (0,1) random variables (U, V) whose joint distribution function is C. The first step is generating a pair (u, t) of uniform random variables U and T in $[0, 1]$. Then (u, v) is the simulated pair where v can be obtained through the conditional distribution function for V given $U = u$, which we denote $c_u(v)$:

$$c_u(v) = Pr\{V \leq v | U = u\} = \frac{Pr\{V \leq v, U = u\}}{Pr\{U = u\}} = \frac{\int_0^v f(u, w)dw}{f(u)} = \int_0^v f(u, w)dw = \frac{\partial}{\partial u}C(u, v)$$

following that $f(u) = 1$ because U is uniform and $v = c_u^{(-1)}(t)$, where $c_u^{(-1)}$ denote a quasi-inverse of c_u. Since T is $U(0, 1)$, $V = c_u^{(-1)}(T)$ is a random variable with distribution c_u. Therefore, the pair (U, V) are uniformly distributed variables drawn from the respective copula $C(u, v; \theta)$.

The equations in Table 3.3 show how v is obtained for the different copulas presented in section 3.3.

Copula	Conditional distribution
Producte	$v = t$
FGM	$v = \frac{2t}{\sqrt{[1-\theta(2u-1)]^2+4\theta t(2u-1)}-\theta(2u-1)}$
Clayton	$v = \left(u^{-\theta}\left(t^{-\theta/(\theta+1)} - 1\right) + 1\right)^{-1/\theta}$
Frank	$v = -\frac{1}{\theta}\log\left(1 + \frac{t(1-e^{-\theta})}{t(e^{-\theta u}-1)-e^{-\theta u}}\right)$

Table 3.3: Selected conditional transforms for copula generation

For the Gumbel copula, the conditional distribution is not directly invertible [18] and, hence, we use another way to generate variables using the following general algorithm [3]:

1. Generate two independent uniform variables (v_1, v_2).
2. Set $w = K_C^{-1}(v_2)$ where $K_C(t) = t - \frac{\varphi(t)}{\varphi'(t)}$.
3. Set $u_1 = \varphi^{-1}[v_1\varphi(w)]$ and $u_2 = \varphi^{-1}[(1 - v_1)\varphi(w)]$.

The desired pair is then (u_1, u_2). In the above algorithm, the function $K_C(t)$ is the distribution function of the random variable $C_\theta(U_1, U_2)$ where U_1 and U_2 are uniform random variables with

an Archimedean copula C generated by φ (Archimedean copulas and their generators will be introduced in section 3.5). For the Gumbel copula, the above algorithm is:

1. Generate two independent uniform variables (v_1, v_2).
2. Set $K_C(w) = w(1 - \frac{\log(w)}{\theta}) = v_2$, and solve it numerically for $0 < w < 1$.
3. Set $u_1 = \exp[v_1^{1/\theta} \log(w)]$ and $u_2 = \exp[(1 - v_1)^{1/\theta} \log(w)]$.

In Figure 3.4, the dependence parameter θ has been set such that $\rho(X, Y) = 0.8$. In the case of the FGM copula, θ has been set such that $\rho(x, y) = 0.3$ because it is unable to accommodate large dependence. This graphic reaffirms what has been introduced in section 3.3 for each copula.

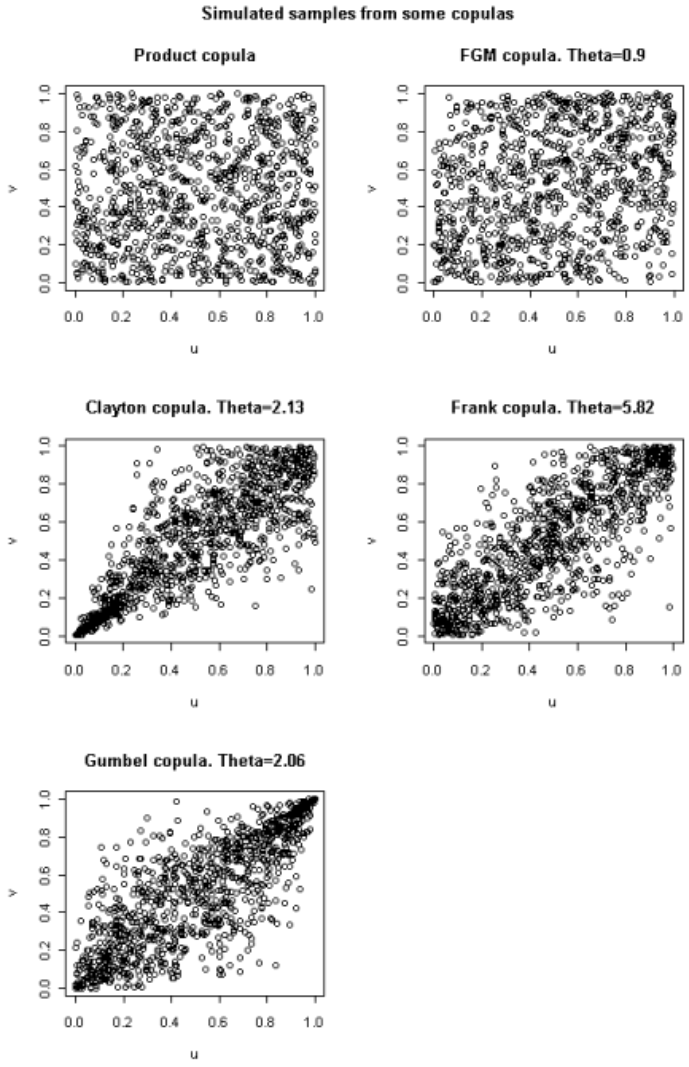


Figure 3.4: Simulated samples from some copulas with $\rho(X, Y) = 0.8$ except for FGM where $\rho(X, Y) = 0.3$ and for Product copula where $\rho(X, Y) = 0$.

It can be seen that variables drawn from the Frank copula exhibit symmetric dependence in both tails but this dependence seems to be weak. This suggests that the Frank copula is best suited for applications in which tail dependence is relatively weak.

In contrast to Frank copula, the Clayton and Gumbel copulas exhibit asymmetric dependence. Clayton dependence is strong in the left tail but weak in the right tail. The implication is that the Clayton copula is best suited for applications in which two outcomes are likely to experience low values together. On the other hand, the Gumbel copula exhibits strong right tail dependence and weak left tail dependence, although the contrast between the two tails of the Gumbel is not as pronounced as in the Clayton copula. Consequently, Gumbel is an appropriate modeling choice when two outcomes are likely to simultaneously realize upper tail values.

Finally, the FGM copula exhibits symmetry in both tails, but it can not accommodate variables with large dependence. The FGM copula allows negative dependence but it must be weak.

The implication of these graphs is that multivariate distributions with similar degrees of dependence might exhibit substantially different dependence structures.

3.5 Archimedean Copulas

Copulas have been introduced and we have seen the role they play in modeling the dependence structure of two random variables. A family of copulas that allow us to construct joint distribution functions is introduced in this section.

3.5.1 Definitions

We have seen that if there is independence between two continuous random variables X and Y with joint distribution H and marginals F and G, then $H(x, y) = F(x)G(y)\ \forall x, y \in \overline{\mathbb{R}}$.

There are a few cases in which a function of H factors into a product of a function of F and a function of G. For example, as seen in [3], the *Ali-Mikhail-Haq* family of copulas, where the relation between H, F, G and θ is the following

$$\frac{1 - H(x, y)}{H(x, y)} = \frac{1 - F(x)}{F(x)} + \frac{1 - G(y)}{G(y)} + (1 - \theta)\frac{1 - F(x)}{F(x)}\frac{1 - G(y)}{G(y)}$$

can be rewritten as

$$1 + (1 - \theta)\frac{1 - H(x, y)}{H(x, y)} = \left(1 + (1 - \theta)\frac{1 - F(x)}{F(x)}\right)\left(1 + (1 - \theta)\frac{1 - F(x)}{F(x)}\right)$$

That is $\lambda(H(x, y)) = \lambda(F(x)) \cdot \lambda(G(y))$ where $\lambda(t) = 1 + (1 - \theta)\frac{1-t}{t}$.

Let φ be defined as $\varphi(t) = -\log(\lambda(t))$. Then, whenever we can write $\lambda(H(x,y)) = \lambda(F(x)) \cdot \lambda(G(y))$ for a function λ (positive in $(0,1)$), we can write H as a sum of functions of the marginals, i.e. $\varphi(H(x,y)) = \varphi(F(x)) + \varphi(G(y))$ or, for copulas, $\varphi(C(u,v)) = \varphi(u) + \varphi(v)$.

Example 3.5.1. *The copula $\widehat{C}_\theta$ given by $\widehat{C}_\theta(u,v) = (u^{-1/\theta} + v^{-1/\theta} - 1)^{-\theta}$ satisfies $\varphi(C(u,v)) = \varphi(u) + \varphi(v)$ with $\varphi(t) = t^{-1/\theta} - 1$.*

The main point of this kind of copulas is the ease with which they can be constructed but we need to solve the relation $\varphi(C(u,v)) = \varphi(u) + \varphi(v)$ for $C(u,v)$, that is, $C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ for an appropriately defined "inverse" $\varphi^{[-1]}$.

Definition 3.5.2. *Let φ be a continuous, strictly decreasing function from I to $[0,\infty]$ such that $\varphi(1) = 0$. The pseudo-inverse of φ is a function $\varphi^{[-1]}$ with $Dom\varphi^{[-1]} = [0,\infty]$ and $Ran\varphi^{[-1]} = I$ given by*

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t) & \text{if } 0 \le t \le \varphi(0) \\ 0 & \text{if } \varphi(0) \le t \le \infty \end{cases}$$

Note that $\varphi^{[-1]}$ is continuous and non increasing on $[0,\infty]$, and strictly decreasing on $[0,\varphi(0)]$. Furthermore, $\varphi^{[-1]}(\varphi(t)) = t$ on I and

$$\varphi(\varphi^{[-1]}(t)) = \begin{cases} t & \text{if } 0 \le t \le \varphi(0) \\ \varphi(0) & \text{if } \varphi(0) \le t \le \infty \end{cases}$$

Finally, if $\varphi(0) = \infty$, then $\varphi^{[-1]} = \varphi^{-1}$.

Lemma 3.5.3. *Let C be a function from I^2 to I that satisfies $C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$. Then C satisfies the boundary conditions for a copula.*
$C(u,0) = \varphi^{[-1]}(\varphi(u) + \varphi(0)) = 0$
$C(u,1) = \varphi^{[-1]}(\varphi(u) + \varphi(1)) = \varphi^{[-1]}(\varphi(u)) = u$
Analogously, the conditions are satisfied for $C(0,v)$ and $C(1,v)$.

It has been seen that copulas can be constructed with a given function φ but it is not true that they are always copulas. We have seen that a function φ has to be a continuous, strictly decreasing function from I to $[0,\infty]$ such that $\varphi(1) = 0$. Then a function C from I^2 to I that satisfies $C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ satisfies the boundary conditions for a copula. It is still necessary for C to be 2-increasing in order to be a copula. Therefore, the election of φ must be restricted.

Lemma 3.5.4. *C is 2-increasing if and only if whenever $u_1 \le u_2$, $C(u_2,v) - C(u_1,v) \le u_2 - u_1$.*

Theorem 3.5.5. *Let φ be a continuous, strictly decreasing function from I to $[0,\infty]$ such that $\varphi(1) = 0$, and let $\varphi^{[-1]}$ be the pseudo-inverse of φ. Then the function C from I^2 to I given by $C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ is a copula if and only if φ is convex.*

Copulas of the form $C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ are called *Archimedean Copulas* and the function φ is called a generator of the copula. If $\varphi(0) = \infty$, we say that φ is a strict generator and, in this case, $\varphi^{[-1]} = \varphi^{-1}$ and $C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ is said to be a strict Archimedean copula. To

be precise, φ is an additive generator of C. If we set $\lambda(t) = exp(-\varphi(t))$ and $\lambda^{[-1]}(t) = \varphi^{[-1]}(-ln(t))$, then $C(u,v) = \lambda^{[-1]}(\lambda(u)\lambda(v))$ and λ is called a multiplicative generator.

Some properties of Archimedean copulas are stated below. However, more properties can be found in [3].

Theorem 3.5.6. *Let C be an Archimedean copula with generator φ. Then:*

1. *C is symmetric; i.e., $C(u,v) = C(v,u)\ \forall u, v \in I$*
2. *C is associative, i.e, $C(C(u,v),w) = C(u,C(v,w))\ \forall u, v, w \in I$*
3. *if $c > 0$ is any constant, then $c\varphi$ is also a generator of C*

An example of a Archimedean copula is the product copula $\prod(u,v) = uv$. Let $\varphi(t) = -\log(t)$ for t in $[0,1]$. Since $\varphi(0) = \infty$, φ is strict and, then, $\varphi^{[-1]}(t) = \varphi^{-1}(t) = e^{-t}$, and generating C yields $C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)) = \exp(-[(-\log(u) + (-\log(v))]) = \exp(\log(u) + \log(v)) = uv$.

Among copulas introduced in section 3.3, Clayton, Frank and Gumbel are Archimedean copulas with generators $\varphi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$, $\varphi(t) = -\log\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$ and $\varphi(t) = (-\log(t))^{\theta}$, respectively. It is easy to check that the FGM copula is not Archimedean since it is not associative,

$$C_\theta\left(\frac{1}{4}, C_\theta\left(\frac{1}{2}, \frac{1}{3}\right)\right) \neq C_\theta\left(C_\theta\left(\frac{1}{4}, \frac{1}{2}\right), \frac{1}{3}\right)$$

for all $\theta \in [-1,1]$ except 0. Hence, FGM copulas are not Archimedean except the specific case when $\theta = 0$, corresponding to the product copula case.

To summarize, Archimedean copulas can be constructed using functions that will serve as generators, that is, continuous, decreasing convex functions φ from I to $[0,\infty]$ with $\varphi(1) = 0$. With a given φ, families of Archimedean copulas can be generated with different values of the dependence parameter, θ.

3.5.2 *Fréchet-Hoeffding* bounds for Archimedean copulas

The following theorems can often be used to determine whether or not $M(u,v)$, $\prod(u,v)$ and $W(u,v)$ are limiting members of an Archimedean family. We let Ω denote the set of continuous strictly decreasing convex functions φ from I to $[0,\infty]$ with $\varphi(1) = 0$.

Theorem 3.5.7. *Let $\{C_\theta | \theta \in \Theta\}$ be a family of Archimedean copulas with differentiable generators φ_θ in Ω. Then $C = \lim C_\theta$ is an Archimedean copula if and only if there exists a function $\varphi \in \Omega$ such that for all $s, t \in (0,1)$,*

$$\lim \frac{\varphi_\theta(s)}{\varphi'_\theta(t)} = \frac{\varphi(s)}{\varphi'(t)} \tag{3.2}$$

where "lim" denotes the appropriate one-sided limit as θ approaches an endpoint of the parameter interval Θ.

Since the generator of W is $\varphi(t) = 1 - t$, W will be the limit of a family $\{C_\theta|\theta \in \Theta\}$ if $\lim \varphi_\theta(s)/\varphi'_\theta(t) = s - 1$; and since the generator of $\prod$ is $\varphi(t) = -\log(t)$, $\prod$ will be the limit of a family $\{C_\theta|\theta \in \Theta\}$ if $\lim \varphi_\theta(s)/\varphi'_\theta(t) = t\log(s)$.

Theorem 3.5.8. *Let $\{C_\theta|\theta \in \Theta\}$ be a family of Archimedean copulas with differentiable generators φ_θ in Ω. Then $\lim C_\theta(u, v) = M(u, v)$ if and only if*

$$\lim \frac{\varphi_\theta(t)}{\varphi'_\theta(t)} = 0 \text{ for } t \in (0, 1) \tag{3.3}$$

where "lim" denotes the appropriate one-sided limit as θ approaches an endpoint of the parameter's interval Θ.

3.5.3 Kendall's τ for Archimedean copulas

Quantifying dependence is relatively straightforward for Archimedean copulas because Kendall's τ simplifies to a function of the generator function.

Theorem 3.5.9. *Let X and Y be random variables with an Archimedean copula C generated by φ in Ω. The population version of Kendall's τ for X and Y is given by*

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt \tag{3.4}$$

On the other hand, there is not any known analytical relationship between the generator φ of a copula and the Spearman's correlation ρ of X and Y.

Chapter 4

Computing ARE for Gumbel copula

As introduced in chapter 2, the main aim of this master thesis is to study the robustness of the methodology developed by Gómez and Lagakos [1] using the ARE to determine which should be the main endpoint in a randomized clinical trial when changing the copula considered. In the Gómez and Lagakos methodology, the Frank copula was chosen and, in this chapter, the same methodology is developed but for another copula. For this new copula, the results and guidelines for using the composite endpoint or the relevant endpoint are given for censoring cases 1 and 3.

4.1 Stable (Gumbel-Hougaard) copula

The Asymptotic Relative Efficiency (ARE) will be defined for the Stable (Gumbel-Hougaard) copula, also known as Gumbel Copula. As seen in chapter 3, the Gumbel Copula, compared to Frank copula, yields to different families of bivariate distribution functions. The main difference between Frank and Gumbel copula is that the first one exhibits weak tail dependence and the second one shows a strong right tail dependence. On the other hand, Frank copula allows both positive or negative dependence while Gumbel copula only allows positive dependence. This could be a problem when comparing the results for both copulas since it would not be possible to compare the ARE results for two negatively correlated outcomes. It is possible to find a situation where higher values of an outcome are correlated to lower values of the other outcome, for example in HIV area, the CD4 counts and the viral load are negatively correlated because a higher number of CD4 counts is correlated to a lower viral load. However, in this master thesis, the outcomes of interest are the time to the events and, therefore, it is difficult to consider for a composite endpoint in clinical trials two outcomes with negatively correlated times. It has been seen that each one of these two copulas drives us to different families of distribution functions with the same correlation between them. This situation highlights the importance of checking the robustness of the methodology for different copulas since it is not enough to only take into account the correlation between the two possible outcomes. Although the correlation is

the same using one or another copula, it is also important to consider their dependence structure.

The Gumbel copula is an Archimedean copula and its generator is given by $\varphi(t) = (-\log t)^{\varphi}$ for $\theta \geq 1$. The expression of this copula is

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

and the limiting cases of the dependence parameter θ correspond to $C(u, v; 1) = \prod(u, v)$ and $C(u, v; \infty) = M(u, v)$ [3].

4.2 Computing ARE for Gumbel copula

The steps used in chapter 2 are replicated for the computing of ARE with the new copula instead of Frank copula, starting from the expression of ARE in censoring cases 1 and 3, defined in (2.6) as

$$\text{ARE}(Z_*, Z) = \left(\frac{\mu_*}{\mu}\right)^2 = \frac{\left(\int_0^1 \log\left(\frac{\lambda_*^{(1)}(t)}{\lambda_*^{(0)}(t)}\right) f_*^{(0)}(t)dt\right)^2}{(\log HR_1)^2 \left(\int_0^1 f_*^{(0)}(t)dt\right)\left(\int_0^1 f_1^{(0)}(t)dt\right)}$$

where $f_1^{(0)}(t)$ and $f_*^{(0)}(t)$ are, respectively, the densities for T_1 and T_* in group 0.

This expression of the ARE depends, among other things, on the law of T_* and it can be obtained from the bivariate distribution of (T_1, T_2).

Law of (T_1, T_2)

In this case, T_1 and T_2 are assumed to be binded by Gumbel survival copula instead of Frank survival copula. Assuming equal association parameter θ for groups 0 and 1, the joint survival function for (T_1, T_2) in group j ($j = 0, 1$) is given by

$$S_{(1,2)}^{(j)}(t_1, t_2; \theta) = S_1^{(j)}(t_1) + S_2^{(j)}(t_2) - 1 + \exp\left(-\left[(-\log(1 - S_1^{(j)}(t_1)))^{\theta} + (-\log(1 - S_2^{(j)}(t_2)))^{\theta}\right]^{1/\theta}\right)$$

where $S_1^{(j)}(t_1)$ and $S_2^{(j)}(t_2)$ are the survival functions of T_1 and T_2, respectively, in group j.

Law of T_*

As seen in (2.7), $S_*^{(j)}(t; \theta) = S_{(1,2)}^{(j)}(t, t; \theta)$, and having $f_*^{(j)}(t; \theta) = -\partial S_*^{(j)}(t; \theta)/\partial t$, we have

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

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

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

Hence, in order to compute ARE(Z_*, Z) assuming Gumbel copula for both groups with equal association parameter θ, we need to specify the same parameters as in Frank copula's case:

- $f_1^{(j)}(t)$ and $S_1^{(j)}(t)$: The marginal density and survival functions of T_1 in group j ($j = 0, 1$)
- $f_2^{(j)}(t)$ and $S_2^{(j)}(t)$: The marginal density and survival functions of T_2 in group j ($j = 0, 1$)
- θ: The copula association parameter between T_1 and T_2.
- HR_1: The constant hazard ratio of T_1, $HR_1 = \lambda_1^{(1)}(t)/\lambda_1^{(0)}(t)$

These parameters can be computed given the frequencies p_1 and p_2 of observing endpoint E_1 and E_2 in treatment group 0, the relative treatment effects on E_1 and E_2 given by hazard ratios HR_1 and HR_2, the shape parameters of Weibull distribution β_1 and β_2 and the degree of dependence between T_1 and T_2 given by Spearman's rank correlation coefficient ρ.

The relation between those parameters is the same for both Gumbel and Frank copula, given in chapter 2. However, as stated in chapter 3, the dependence parameter θ for Gumbel copula cannot be obtained directly from Spearman's ρ. The first idea was to obtain θ using Kendall's τ and then obtain a relation between ρ and τ. However, as stated in Theorem 3.4.1, it is not true that there is a one-to-one relationship between these two concordance measures (ρ and τ). In order to make the methodology comparable and using the same parameters for Gumbel and Frank copula, including ρ, the dependence parameter has been obtained with ρ using numerical approximations with the R-package `copula` [19, 20, 21].

4.3 Setting for the computations

We have now the expression of the ARE depending on parameters that can be understood from a clinical point of view. Next step is to repeat the simulation studies carried out in [1] and [2] in order to observe whether the composite endpoint should be used in each case or not depending on these parameters. Therefore, the same plausible values for the different parameters have been taken into account for the computations of ARE for both Frank and Gumbel copula and censoring cases 1 and 3:

- Several frequency situations are reproduced for events E_1 and 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 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 is beneficial. Each hazard ratio is combined with eight different relative treatment effects on the additional endpoint 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.
- Values for the shape parameters of Weibull distribution β_1 and β_2 are set to 0.5, 1 and 2 in order to have decreasing, constant and increasing hazards, respectively.
- A range of associations have been considered from weak (Spearman's rank correlation ($\rho = 0.15, 0.25$), through moderate ($\rho = 0.35, 0.45$) to strong ($\rho = 0.55, 0.65, 0.75$).

The proposed settings (shown in Table 4.1) provide 72.576 configurations for both cases 1 and 3.

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 4.1: Possible values for the simulation of ARE

The computations of ARE for both Frank and Gumbel copulas for censoring cases 1 and 3 were done with software R (version 2.10.1) (see code in Appendix A).

4.4 Results for Gumbel Copula

Preliminary analyses show that the ARE for a given combination $(p_1, p_2, HR_1, HR_2, \rho)$ is analogous for the 9 different choices of (β_1, β_2) (this result was also concluded in [1, 2] for the Frank Copula). Thus, we conclude that the behavior of ARE is independent of whether the marginal hazard function are decreasing, constant or increasing. For the moment, no other choices for β_1 and β_2 have been considered.

We present the results for the particular combination of $\beta_1 = \beta_2 = 1$, where there are 8.064 different configurations for each censoring case 1 and 3. For each such case, these are grouped for specific values of p_1, p_2, HR_1 yielding a total number of 144 scenarios for each censoring case 1 and 3. For every scenario, the 7 value of ρ and the 8 values of HR_2 are plotted as described below. Each one of the 144 plots displays 8 curves corresponding to 8 different values of the relative treatment effect on E_2 (HR_2). Each plot has Spearman's ρ ranging from 0.15 to 0.75 on the abscissa and the value of ARE on the ordinate, on a logarithmic scale. A logarithmic scale

has been used since it represents better its significance. For example, an $ARE(Z_*, Z) = 2$ is as relevant as an $ARE(Z_*, Z) = 0.5$. That is, the distance from a point with $ARE(Z_*, Z) = 2$ to 1 is the same as the distance from a point with $ARE(Z_*, Z) = 0.5$ to 1.

Figure 4.1 shows the plot for the particular scenario for $p_1 = 0.1$, $p_2 = 0.05$ and $HR_1 = 0.6$ for censoring case 1 and it can be used by way of illustration. In this case, the ARE is higher than 1 for $HR_2 < 0.7$ and is lower than 1 for $HR_2 > 0.7$. For $HR_2 = 0.7$, the threshold between having the ARE higher or lower than 1 is at $\rho = 0.55$. Therefore, we could conclude that the composite endpoint should be considered for $HR_2 > 0.7$ or $HR_2 = 0.7$ and weak or moderate correlation between the 2 endpoints. The 288 plots (144 for each censoring case) depicting all the scenarios of $\beta_1 = \beta_2 = 1$ can be found in Appendix B.

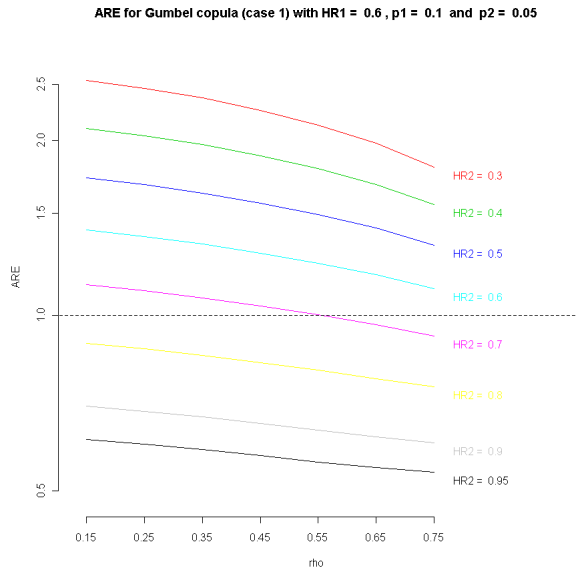


Figure 4.1: ARE for Gumbel copula (case 1) for $HR_1 = 0.6$, $p_1 = 0.1$ and $p_2 = 0.05$.

4.4.1 Case-1 guidelines: non-fatal relevant and additional endpoints

The general pattern in a case where neither the relevant nor the additional endpoint is terminating, is that ARE decreases when the Spearman's rank correlation between them increases; and also when the relative effect of treatment on the additional endpoint decreases (HR_2 increases). This pattern is similar to the one concluded in [1] for Frank Copula. Almost all the plots display a similar parallel behavior as shown, by way of example, in Figure 4.1. Thanks to this parallel behavior, it is easy to develop general guidelines for using the composite endpoint depending on HR_1, HR_2 and ρ. However, when p_1 gets higher and p_2 smaller, different behaviors can be observed. Figure 4.2 is an example of this different behavior for $p_1 = 0.4$ and $p_2 = 0.05$. This could be a problem but it can be observed that it does not affect the decision of whether to use

or not the composite endpoint because the value of ARE for these situations is always smaller than one and, hence, there is no doubt in deciding not to use the composite endpoint.

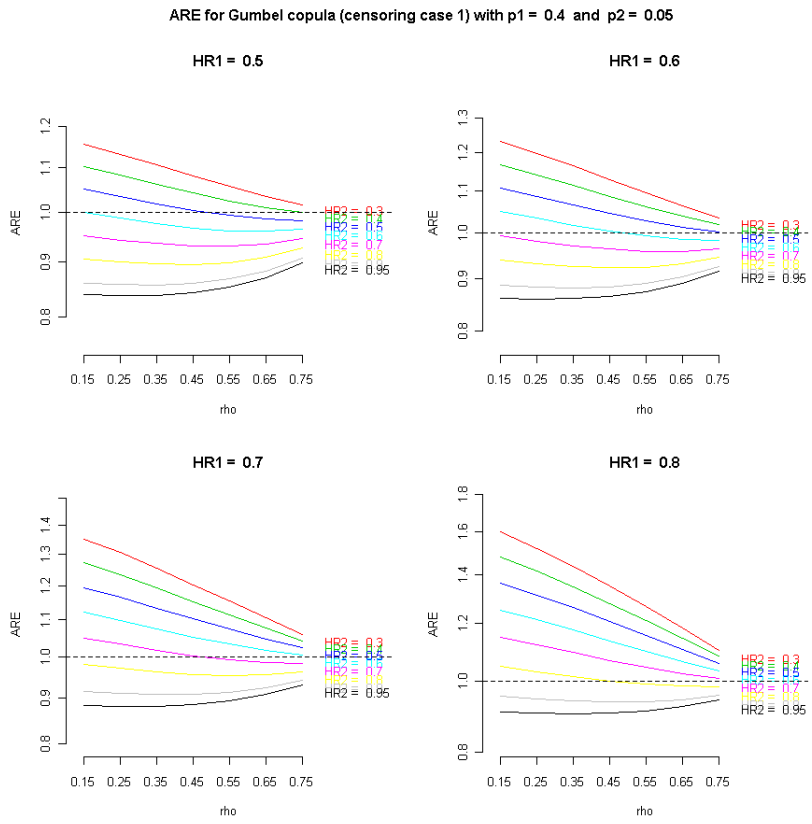


Figure 4.2: ARE for Gumbel copula (case 1) for $p_1 = 0.4$ and $p_2 = 0.05$.

The recommendation to use the composite endpoint is clear when the relative treatment effect HR_2 on the additional endpoint is smaller (higher beneficial effect) than on the relevant endpoint. However, when HR_2 is about the same of HR_1, the composite endpoint should be used if $p_1 < 0.3$. If $p_1 \geq 0.3$ or HR_2 is slightly larger than HR_1, the decision on whether or not to use the composite endpoint depends on the frequency of observing each endpoint together with their correlation.

4.4.2 Case-3 guidelines: relevant endpoint, fatal; additional endpoint, non-fatal

ARE behavior in cases where the relevant endpoint does but the additional endpoint does not include a terminating event is very similar to that of case 1, in which neither event is fatal. Here too it will be observed that ARE decreases: when the correlation between the two endpoint increases; and when HR_2 increases. In that case, it can also be observed that when p_1 increases and p_2 is smaller, the ARE increases with the correlation.

In censoring case 3, the decision is clear when the relative treatment effect is greater on the additional endpoint than on the relevant endpoint ($HR_2 < HR_1$), in which case the composite endpoint should always be used. On the other hand, when $HR_2 \geq HR_1$, the relevant endpoint should always be used when $HR_2 \geq HR_1 + 0.2$. If the additional endpoint one plans to add to the relevant endpoint has approximately the same ($HR_1 = HR_2$) or a slightly smaller effect on treatment ($HR_2 - HR_1 = 0.1$) than does the relevant endpoint, the decision as to whether or not the composite endpoint should be used is not clear, and the choice will depend on the value of these effects on treatment together with the frequency of observations of either endpoint and its correlation.

4.4.3 General Guidelines

The behavior of the ARE has been studied using graphical techniques but it is interesting to see the percentages of situations on which the composite endpoint should be used as a function of the different values of $\beta 1, \beta 2, p_1, p_2, HR_1, HR_2$ and ρ. The following figures consider these proportions considering the 9 different cases for (β_1, β_2) and, hence, 72.576 cases for each censoring case.

Figure 4.3 shows that the percentage of cases in which we should use the composite endpoint ($ARE > 1$) is higher when the Spearman's rank correlation value between the endpoint decreases. However, this is not enough alone to elucidate whether or not to use the composite endpoint and, hence, the influence of other parameters must be evaluate.

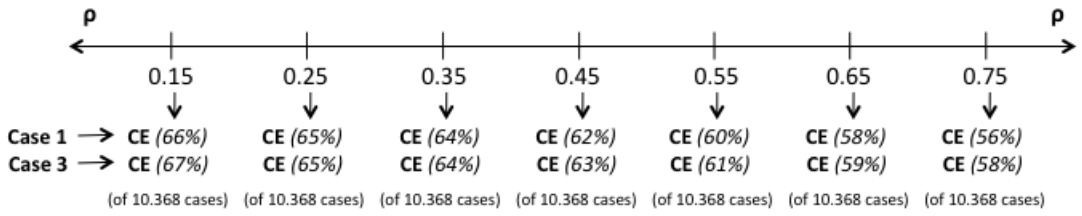


Figure 4.3: Percentage of situations for which the composite endpoint should be used by ρ.

Figure 4.4 shows that using the composite endpoint is less relevant as the probability of observing the relevant endpoint gets larger. On the other hand, the additional endpoint tends to be more necessary when the probability of being observed gets larger. This is not enough to recommend the composite endpoint over the relevant endpoint and next figure shows how the ARE values behave depending on the hazard ratios.

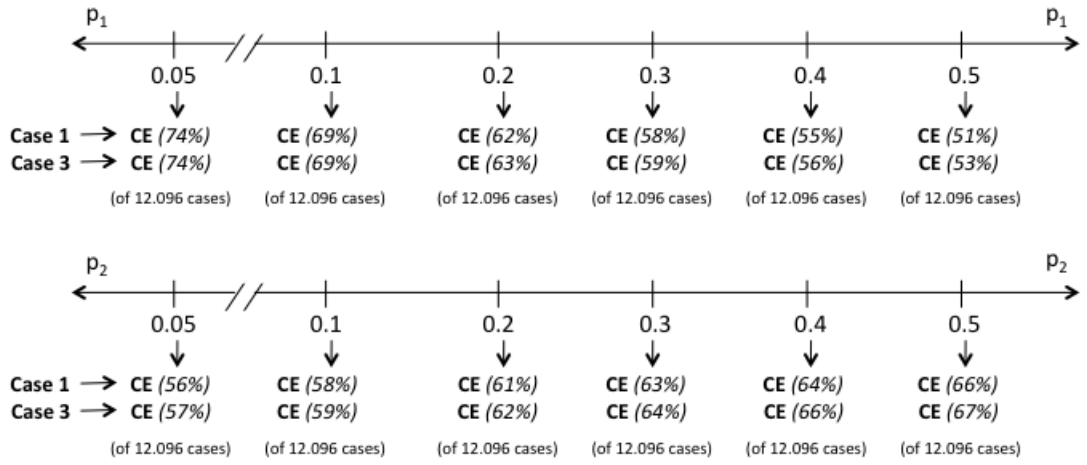


Figure 4.4: Percentage of situations for which the composite endpoint should be used by p_1 and p_2.

As stated above, Figure 4.5 shows how the values of the relative treatment effect on the relevant (HR_1) and additional (HR_2) endpoint influence the ARE value. It is possible to observe that as HR_1 gets larger (and, hence, having less beneficial effect), adding a additional endpoint is more convenient. On the other hand, the value of HR_2 by itself is relevant. If it is small, the composite endpoint should always be used and, if it is large, the relevant endpoint without the addition of the additional endpoint should be used.

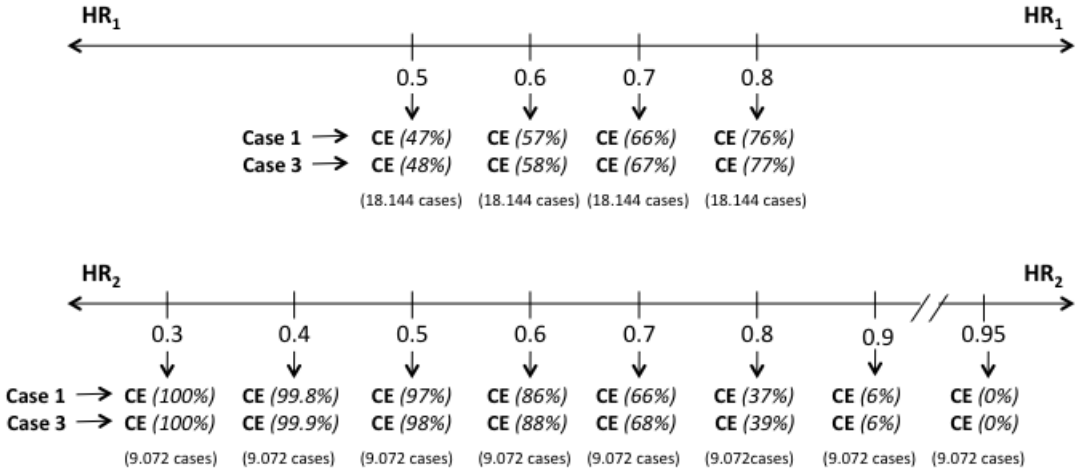


Figure 4.5: Percentage of situations for which the composite endpoint should be used by HR_1 and HR_2.

It is interesting to study the behavior of the ARE for joint combinations of HR_1 and HR_2. Figure 4.6 presents some conclusive results. The composite endpoint should almost always be used if $HR_2 < HR_1$. On the other hand, the relevant endpoint should always be used if $HR_2 > 0.8$ unless perhaps if the effect on E_1 is also very low.

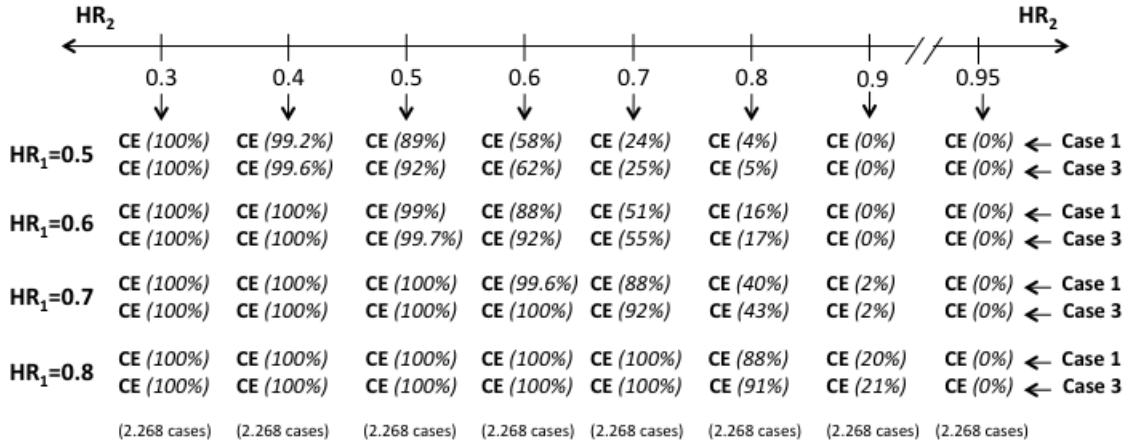


Figure 4.6: Percentage of situations for which the composite endpoint should be used by HR_2 depending on HR_1.

It can be observed in Figure 4.6 that the use of the composite endpoint could be determined by the difference between HR_1 and HR_2. Figure 4.7 shows the behavior of the ARE depending on this difference of hazard ratios without considering $HR_2 = 0.3$ and $HR_2 = 0.95$, where the composite or relevant endpoint, respectively, should always be used. Therefore, the total number of combinations is 54.432 for each censoring case. It can be observed that the composite endpoint should always be used if $HR_2 - HR_1 \leq -0.2$ and the relevant endpoint should always be used if $HR_2 - HR_1 \geq 0.4$ for both censoring cases 1 and 3.

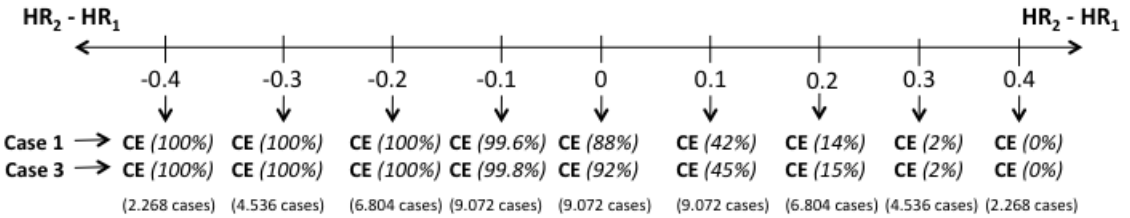


Figure 4.7: Percentage of situations for which the composite endpoint should be used by the difference between HR_2 and HR_1.

After these analysis, the scenarios where adding the additional endpoint is not clear are when

$-0.1 \leq HR_2 - HR_2 \leq 0.3$. Figure 4.8 shows the proportion of cases where the composite endpoint should be used for each one of these differences depending on the value of ρ. The following can be concluded:

- When $HR_2 = HR_1 - 0.1$, the composite endpoint should almost always be used (except some cases when $\rho = 0.75$).
- When $HR_2 = HR_1$, the composite endpoint should be used whenever $\rho < 0.45$).
- The rest of cases should be studied depending on p_1 and p_2.

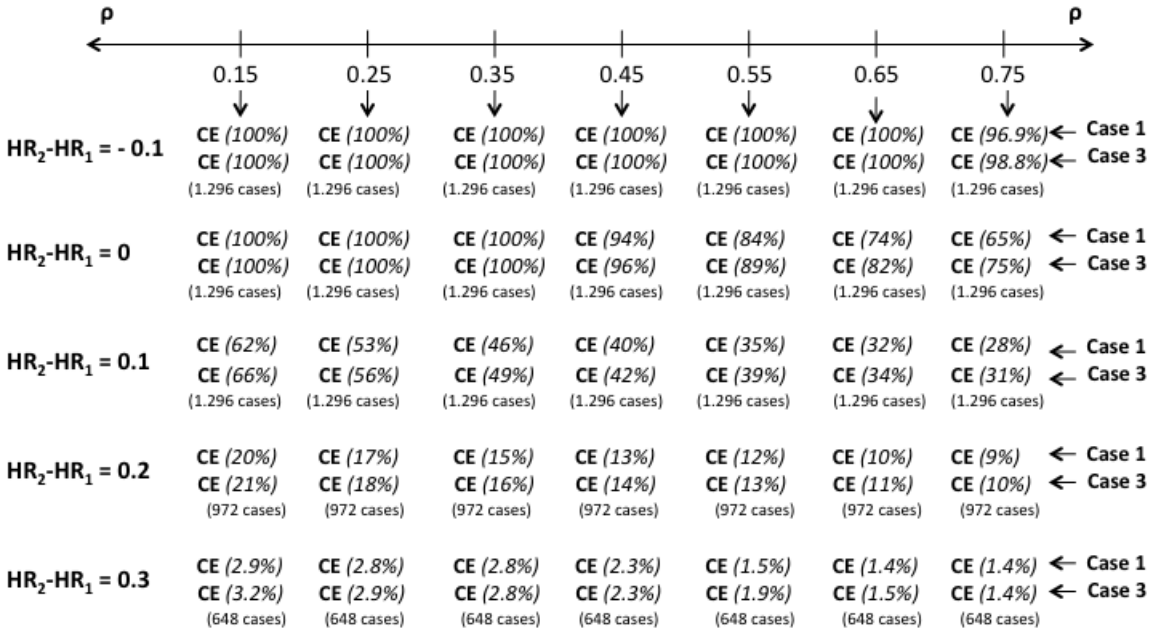


Figure 4.8: Percentage of situations for which the composite endpoint should be used by ρ depending on the difference between HR_2 and HR_1.

Therefore, it is still necessary to study these 5 different scenarios where it is not clear which endpoint should be used and it is necessary to study the behavior of ARE depending on p_1 and p_2:

- $HR_2 = HR_1 - 0.1$ and $\rho = 0.75$.
- $HR_2 = HR_1$ and $\rho \geq 0.45$.
- $HR_2 = HR_1 + 0.1$.
- $HR_2 = HR_1 + 0.2$.
- $HR_2 = HR_1 + 0.3$.

Figure 4.9 summarizes what has been found about the proportion of cases in which the composite endpoint should be used depending on HR_1 and HR_2.

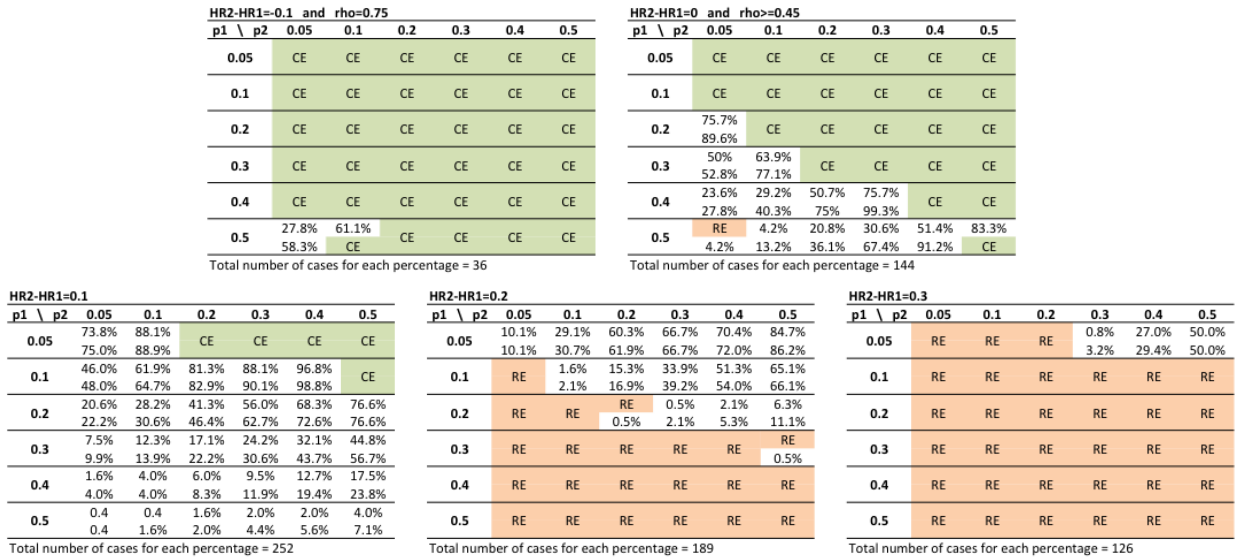
HR1 \ HR2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
0.5	CE	99.2% 99.6%	89% 92%	58% 62%	24% 25%	4% 5%	RE	RE
0.6	CE	CE	99% 99.7%	88% 92%	51% 55%	16% 17%	RE	RE
0.7	CE	CE	CE	99.6% CE	88% 92%	40% 43%	2% 2%	RE
0.8	CE	CE	CE	CE	88% 91%	88% 91%	20% 21%	RE

Total number of cases for each percentage = 2.268

Figure 4.9: Percentage of situations for which the composite endpoint (CE) should be used instead of relevant endpoint (RE) by HR_1 and HR_2.

On the other hand, Figure 4.10 shows the proportion of cases where the composite endpoint should be used for the 5 cases stated above, depending on p_1 and p_2 for both censoring cases 1 and 3. It can be observed that the recommendations are the following:

1. When $HR_2 = HR_1 - 0.1$ and $\rho = 0.75$, use the composite endpoint if:
   - $p_1 \leq 0.4$
   - $p_1 = 0.5$ and $p_2 \geq 0.2$
2. When $HR_2 = HR_1$ and $\rho \geq 0.45$, use the composite endpoint if:
   - $p_1 \leq 0.1$
   - $p_1 = 0.2$ and $p_2 \geq 0.1$
   - $p_1 = 0.3$ and $p_2 \geq 0.2$
   - $p_1 = 0.4$ and $p_2 \geq 0.4$
3. When $HR_2 = HR_1 + 0.1$, use the composite endpoint if:
   - $p_1 = 0.05$ and $p_2 \geq 0.2$
   - $p_1 = 0.1$ and $p_2 = 0.5$
4. When $HR_2 = HR_1 + 0.2$, use the relevant endpoint if:
   - $p_1 \geq 0.3$
   - $p_1 = 0.2$ and $p_2 \leq 0.2$
   - $p_1 = 0.1$ and $p_2 = 0.05$
5. When $HR_2 = HR_1 + 0.3$, use the relevant endpoint if:
   - $p_1 \geq 0.1$
   - $p_1 = 0.05$ and $p_2 \leq 0.2$



HR2-HR1=-0.1 and rho=0.75

p1 \ p2	0.05	0.1	0.2	0.3	0.4	0.5
0.05	CE	CE	CE	CE	CE	CE
0.1	CE	CE	CE	CE	CE	CE
0.2	CE	CE	CE	CE	CE	CE
0.3	CE	CE	CE	CE	CE	CE
0.4	CE	CE	CE	CE	CE	CE
0.5	27.8% 58.3%	61.1% CE	CE	CE	CE	CE

Total number of cases for each percentage = 36

HR2-HR1=0 and rho>=0.45

p1 \ p2	0.05	0.1	0.2	0.3	0.4	0.5
0.05	CE	CE	CE	CE	CE	CE
0.1	CE	CE	CE	CE	CE	CE
0.2	75.7% 89.6%	CE	CE	CE	CE	CE
0.3	50% 52.8%	63.9% 77.1%	CE	CE	CE	CE
0.4	23.6% 27.8%	29.2% 40.3%	50.7% 75%	75.7% 99.3%	CE	CE
0.5	RE 4.2%	4.2% 13.2%	20.8% 36.1%	30.6% 67.4%	51.4% 91.2%	83.3% CE

Total number of cases for each percentage = 144

HR2-HR1=0.1

p1 \ p2	0.05	0.1	0.2	0.3	0.4	0.5
0.05	73.8% 75.0%	88.1% 88.9%	CE	CE	CE	CE
0.1	46.0% 48.0%	61.9% 64.7%	81.3% 82.9%	88.1% 90.1%	96.8% 98.8%	CE
0.2	20.6% 22.2%	28.2% 30.6%	41.3% 46.4%	56.0% 62.7%	68.3% 72.6%	76.6% 76.6%
0.3	7.5% 9.9%	12.3% 13.9%	17.1% 22.2%	24.2% 30.6%	32.1% 43.7%	44.8% 56.7%
0.4	1.6% 4.0%	4.0% 4.0%	6.0% 8.3%	9.5% 11.9%	12.7% 19.4%	17.5% 23.8%
0.5	0.4 0.4	0.4 1.6%	1.6% 2.0%	2.0% 4.4%	2.0% 5.6%	4.0% 7.1%

Total number of cases for each percentage = 252

HR2-HR1=0.2

p1 \ p2	0.05	0.1	0.2	0.3	0.4	0.5
0.05	10.1% 10.1%	29.1% 30.7%	60.3% 61.9%	66.7% 66.7%	70.4% 72.0%	84.7% 86.2%
0.1	RE	1.6% 2.1%	15.3% 16.9%	33.9% 39.2%	51.3% 54.0%	65.1% 66.1%
0.2	RE	RE	RE 0.5%	0.5% 2.1%	2.1% 5.3%	6.3% 11.1%
0.3	RE	RE	RE	RE	RE	RE 0.5%
0.4	RE	RE	RE	RE	RE	RE
0.5	RE	RE	RE	RE	RE	RE

Total number of cases for each percentage = 189

HR2-HR1=0.3

p1 \ p2	0.05	0.1	0.2	0.3	0.4	0.5
0.05	RE	RE	RE	0.8% 3.2%	27.0% 29.4%	50.0% 50.0%
0.1	RE	RE	RE	RE	RE	RE
0.2	RE	RE	RE	RE	RE	RE
0.3	RE	RE	RE	RE	RE	RE
0.4	RE	RE	RE	RE	RE	RE
0.5	RE	RE	RE	RE	RE	RE

Total number of cases for each percentage = 126

Figure 4.10: Percentage of situations for which the composite endpoint (CE) should be used instead of relevant endpoint (RE) by p_1 and p_2 for particular scenarios of HR_1, HR_2 and ρ.

4.5 Conclusions

It is clear that the composite endpoint should be used when HR_2 is small and the relevant endpoint should always be used when it is high. On the other hand, when HR_2 is close to HR_1, the recommendation of whether to use or not the composite endpoint depends on the values of ρ, p_1 and p_2. Most of the cases where it is not clear what should be used are when $HR_2 = HR_1 + 0.1$; when $HR_2 = HR_1$, with p_1 high and p_2 small; and when $HR_2 = HR_1 + 0.2$, with p_1 small and p_2 high. Table 4.2 summarizes which are the cases where the composite and the relevant endpoints should be chosen.

Use the relevant endpoint when:	Use the composite endpoint when:
HR_2=0.95	$HR_2 = 0.3$
$HR_2 \geq 0.9$ and $HR_1 \leq 0.6$	$HR_2 \leq HR_1 - 0.1$
$HR_2 = HR_1 - 0.3$	$HR_2 = HR_1$ with $p_1 \leq p_2$
$HR_2 = HR_1 - 0.2$ with $p_1 \geq 0.3$	

Table 4.2: Cases in which the composite or the relevant endpoint should be chosen

Figure 4.11 shows a decision tree for the selection of the composite endpoint or the relevant endpoint depending on the value of HR_1, HR_2, p_1, p_2 and ρ. In those cases in which it is not clear which one is better, it shows the percentage of cases with the values where the composite endpoint should be chosen.

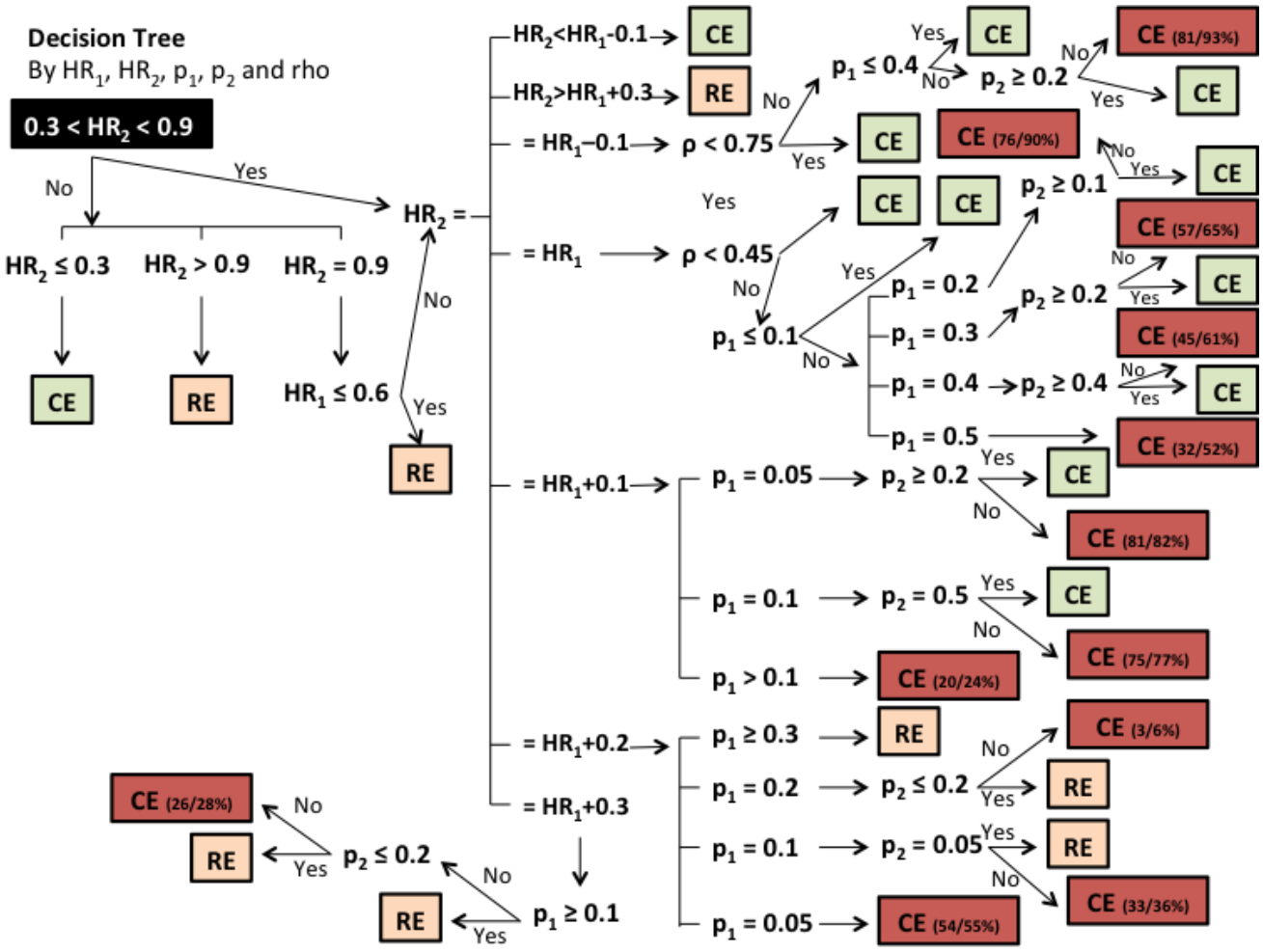


Figure 4.11: Decision tree for the selection of the composite endpoint (CE) or the relevant endpoint (RE).

Chapter 5

Comparation of ARE for Frank and Gumbel copulas

In chapter 4, we calculated the ARE using Frank and Gumbel copula and the results for the latter were analyzed. In this chapter, we compare these results for censoring cases 1 and 3.

These results are compared in two different ways. On one hand, they are described and compared using the absolute difference between both values. This could be useful in order to see if there would be differences between different copulas when calculating a sample size given the parameters and the copula. On the other hand, they are compared in terms of their concordance, studying in which cases the methodology for both copulas yield to the same recommendation on whether or not use the composite endpoint. We will say that there is concordance between both copulas whenever the ARE values computed via Frank or via Gumbel copula are both greater than 1 or both smaller than 1.

The results for the ARE using Frank and Gumbel copula with the same combinations of values for β_1, β_2, HR_1, HR_2, p_1, p_2 and ρ given in Table 4.1 are used to study the robustness of the methodology with respect to the different copulas. A total of 72.576 configurations have been studied for both censoring cases 1 and 3.

5.1 Difference between the ARE values for Frank and Gumbel copulas

In what follows we compare the values of the ARE for both copulas using the absolute difference.

Table 5.1 shows that the ARE value for censoring case 1 ranges from 0.03 to 267.3 and 0.03 to 272.7 for Frank and Gumbel copulas, respectively. The mean values for the ARE for both copulas are 4.9 and 5.1, respectively, with standard deviations of 15.2 and 15.5. We observe that the ARE value for Gumbel copula is slightly higher than the ARE value for Frank copula. Actually, there are 69.948 combinations (96.4%) where the value of the ARE is higher for Gumbel than for Frank

copula and 2.628 combinations (3.6%) where it is the other way around.

	mean (SD)	**min**	Q_1	**median**	Q_3	**max**
ARE_{Frank}	4.95 (15.2)	0.026	0.76	1.18	2.93	267.3
ARE_{Gumbel}	5.08 (15.5)	0.032	0.79	1.22	3.06	272.7
$\lvert ARE_{Frank} - ARE_{Gumbel} \rvert$	0.14 (0.4)	2×10^{-7}	0.02	0.04	0.11	9.529

SD = Standard Deviation. Q_1 and Q_3 are the first and third quartile.

Table 5.1: Descriptive Analysis of the ARE values for censoring case 1

Table 5.2 shows that the ARE value for censoring case 3 ranges from 0.025 to 277.1 and 0.03 to 282.6 for Frank and Gumbel copulas, respectively. The mean values for the ARE for both copulas are 5.47 and 5.52, respectively, with standard deviations of 16.0 and 16.2. We observe again that the ARE value for Gumbel copula is slightly higher than the ARE value for Frank Copula. However, for this censoring case, there are 58.931 combinations (81.2%) in which the value of the ARE is higher for Gumbel than for Frank copula and this percentage is lower than in censoring case 1. On the other hand, there are 13.645 combinations (18.8%) where it is the ARE for Frank copula which is higher.

	mean (SD)	**min**	Q_1	**median**	Q_3	**max**
ARE_{Frank}	5.47 (16.0)	0.025	0.71	1.33	3.68	277.1
ARE_{Gumbel}	5.52 (16.2)	0.030	0.75	1.34	3.68	282.6
$\lvert ARE_{Frank} - ARE_{Gumbel} \rvert$	0.12 (0.3)	1.5×10^{-7}	0.02	0.04	0.10	9.013

SD = Standard Deviation. Q_1 and Q_3 are the first and third quartile.

Table 5.2: Descriptive Analysis of the ARE values for censoring case 3

Despite the fact that the value of ARE is higher for Gumbel copula, this difference is small in almost all the cases. There are only 1.489 combinations (2.1%) and 1.395 combinations (1.9%) for censoring cases 1 and 3, respectively, in which the absolute difference between the two values is higher than one. On the other hand, in the 72.5% and 75.5% of the combinations for both cases, respectively, the absolute difference is lower than 0.1 and, in the 56.4% and 55.5% of the cases, respectively, it is lower than 0.05. It is important to remark that such a small difference would imply an equivalent ratio of sample sizes using both copulas. Therefore, we can claim that the differences between both copulas are negligible.

To complete the study on the relationship between ARE based on Frank and Gumbel copulas, the correlation between them is calculated and it is 0.999 for both censoring cases 1 and 3. Moreover, the relation between the two values has been explored using graphical techniques: Figure 5.1

shows a bivariate plot of the ARE for Frank copula in the abscissa and the ARE for Gumbel copula in the ordinate.

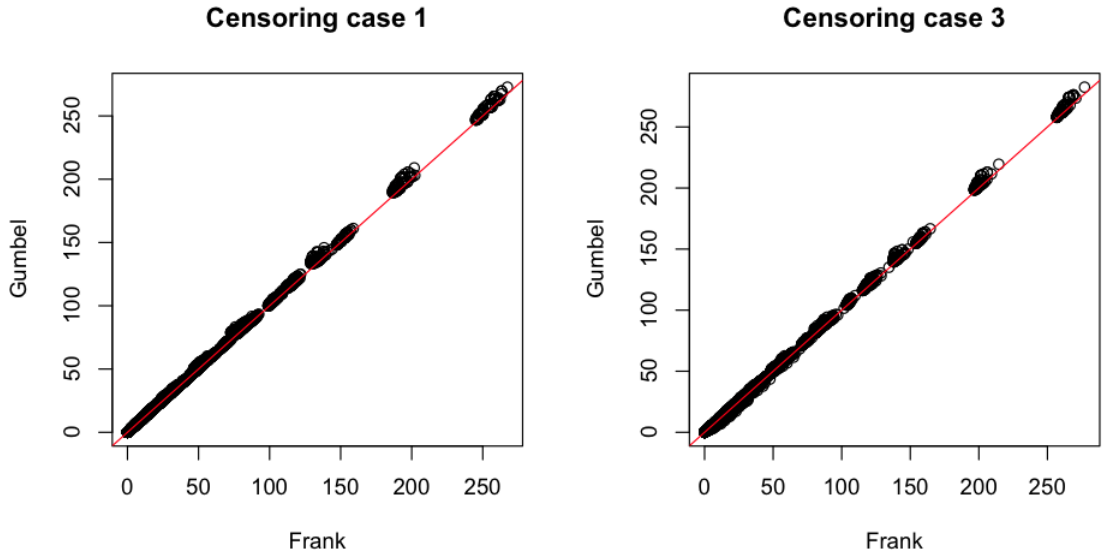


Figure 5.1: Bivariate plot of values of ARE for Frank and Gumbel copulas.

Using graphical techniques, it can be concluded that the value of the ARE using the Gumbel copula is slightly higher than the value of the ARE using the Frank copula but they are similar in all the cases. For censoring case 3, it is clear, as stated above, that there is a higher percentage of cases in which the ARE for Frank copula is higher than the ARE for Gumbel copula than in censoring case 1.

5.2 Concordance

It is relevant and of great practical importance to study in which situations both copulas yield to the same recommendation on whether or not use the composite endpoint. An analysis of concordance is performed and the results are provided in Table 5.3.

The first conclusion is that both copulas agree in recommending the use of the composite or the relevant endpoint in 98.0% and 98.7% of the configurations for censoring cases 1 and 3, respectively. We pay close attention to these situations in which the methodology yield to different results for the two copulas (2.1% and 1.28% of the combinations for censoring cases 1 and 3, respectively).

	$ARE_{Frank} > 1$	$ARE_{Frank} \leq 1$
$ARE_{Gumbel} > 1$	59.5% 61.1%	1.9% 1.2%
$ARE_{Gumbel} \leq 1$	0.02% 0.08%	38.5% 37.6%

Table 5.3: Percentage of cases depending on the ARE value using Frank or Gumbel copula for censoring cases 1 and 3.

5.2.1 Composite endpoint using Frank copula and relevant endpoint using Gumbel copula

Table 5.4 describes the behavior of the ARE value for those cases in which it is higher than 1 for Frank copula and lower than 1 for Gumbel copula. There are 11 cases for censoring case 1, in which the maximum difference between the two values is 0.03 and, hence, it is possible to conclude that they are very similar although they disagree in recommending the use of the composite endpoint. On the other hand, there are 58 cases for censoring case 3, in which the maximum difference is 0.17. A box plot (Figure 5.2) is useful in order to study these differences and it can be observed that there are a few values where this difference is higher than 0.1. Actually, in 75% of the cases, the difference between the two values of the ARE is lower than 0.06.

$ARE_{Frank} > 1$ and $ARE_{Gumbel} \leq 1$		**mean (SD)**	**min**	Q_1	**median**	Q_3	**max**
Censoring case 1	ARE_{Frank}	1.01 (0.01)	1.01	1.01	1.01	1.02	1.02
($n = 11$)	ARE_{Gumbel}	0.99 (0.01)	0.98	0.98	0.99	0.99	0.99
	$ARE_{Frank} - ARE_{Gumbel}$	0.02 (0.01)	0.01	0.01	0.03	0.03	0.03
Censoring case 3	ARE_{Frank}	1.03 (0.03)	1.00	1.01	1.02	1.03	1.15
($n = 58$)	ARE_{Gumbel}	0.98 (0.02)	0.94	0.97	0.98	0.99	1.00
	$ARE_{Frank} - ARE_{Gumbel}$	0.05 (0.03)	0.01	0.03	0.04	0.06	0.17

SD = Standard Deviation. Q_1 and Q_3 are the first and third quartile.

Table 5.4: Descriptive Analysis of the ARE values for those cases in which $ARE_{Frank} > 1$ and $ARE_{Gumbel} \leq 1$

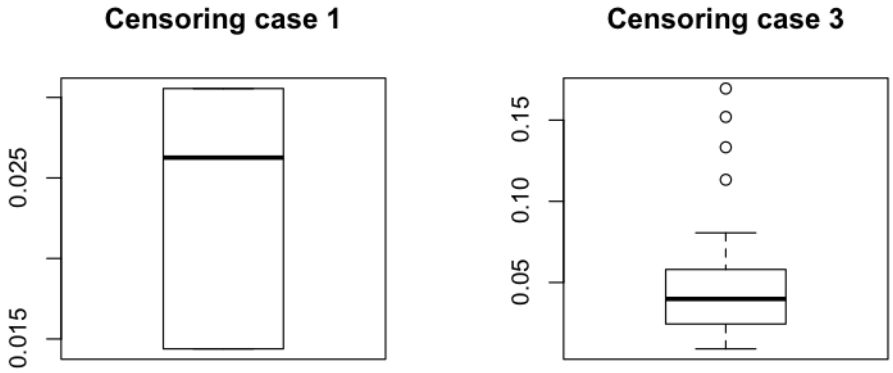


Figure 5.2: Box plot of the difference between ARE values in cases in which $ARE_{Frank} > 1$ and $ARE_{Gumbel} \leq 1$.

5.2.2 Composite endpoint using Gumbel copula and relevant endpoint using Frank copula

Table 5.5 describes the behavior of the ARE value for those cases in which it is higher than 1 for Gumbel copula and lower than 1 for Frank copula. There are 1.415 cases for censoring case 1, in which the maximum difference between the two values is 0.14, and 891 cases for censoring case 3, in which the maximum difference is 0.12. Figure 5.3 shows the distribution of these differences. It is interesting to remark that, in the 75% of the cases, the difference between the two values is lower than 0.06.

$ARE_{Frank} \leq 1$ and $ARE_{Gumbel} > 1$		**mean (SD)**	**min**	Q_1	**median**	Q_3	**max**
Censoring case 1	ARE_{Frank}	0.98 (0.02)	0.88	0.97	0.98	0.99	1.00
(n = 1.415)	ARE_{Gumbel}	1.02 (0.02)	1.00	1.01	1.02	1.03	1.13
	$ARE_{Gumbel} - ARE_{Frank}$	0.05 (0.03)	0.004	0.02	0.04	0.06	0.14
Censoring case 3	ARE_{Frank}	0.98 (0.02)	0.91	0.97	0.98	0.99	1.00
(n = 891)	ARE_{Gumbel}	1.02 (0.02)	1.00	1.01	1.02	1.03	1.09
	$ARE_{Gumbel} - ARE_{Frank}$	0.04 (0.02)	0.003	0.02	0.04	0.06	0.12

SD = Standard Deviation. Q_1 and Q_3 are the first and third quartile.

Table 5.5: Descriptive Analysis of the ARE values for those cases in which $ARE_{Frank} \leq 1$ and $ARE_{Gumbel} > 1$.

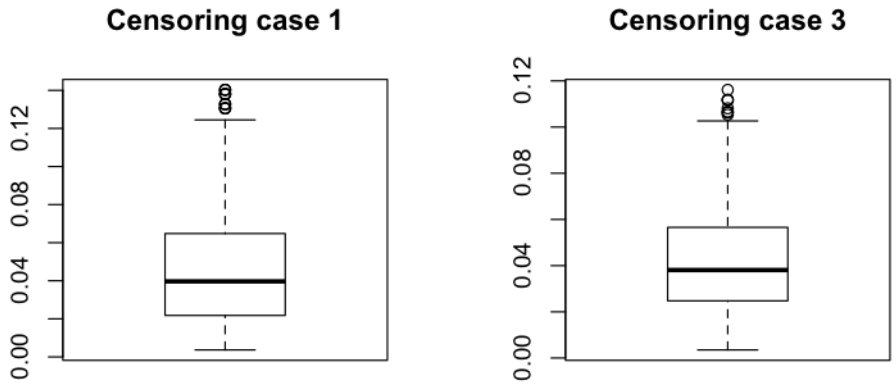


Figure 5.3: Box plot of the difference between ARE values in cases in which $ARE_{Frank} \leq 1$ and $ARE_{Gumbel} > 1$.

5.3 Conclusions

After studying these results, we conclude that the methodology based on the ARE is robust for the choice of the copula when restricted to Frank and Gumbel families. The values for both copulas are highly correlated ($\rho = 0.999$ for both censoring cases 1 and 3) and yield to the same recommendation of whether or not to use the composite endpoint in more than 98% of the cases. In the remaining cases, the difference between both values of the ARE is lower than 0.15. Moreover, this difference is lower than 0.06 in 75% of the situations (Q_3 in Tables 5.4 and 5.5). Considering that the values of the ARE are around 1, we notice that there would be a small effect of the addition of the additional endpoint to the relevant endpoint and, hence, use the composite endpoint in the computation of the sample size. As a remark, in Gómez and Lagakos paper [1], they recommended to use the value of 1.1 as a threshold for using the composite endpoint or the relevant endpoint.

Despite this high concordance, we have described the cases in which there was discordance because it would be useful for future research using other copulas. These analysis can be found in Appendix C and it can be concluded that both copulas disagree when HR_1 and HR_2 are quite similar and for high values of p_1 jointly with low values of p_2. It its interesting to remark that they are the same situations seen in chapter 4 in which the ARE value using Gumbel copula did not yield to a clear recommendation on whether use or not use the composite endpoint.

Furthermore, after observing the high concordance between the values of the ARE for both Frank and Gumbel copula, it could be interesting to study what is different in the computation of the ARE for these two copulas. The main difference is the density function for the composite endpoint T_* and, hence, an analysis of these density functions depending on the copula considered has been performed.

5.3.1 Density functions for the composite endpoint T_*

It is important to remark that the value of the ARE is given by the expression defined in (2.6) as

$$\text{ARE}(Z_*, Z) = \left(\frac{\mu_*}{\mu}\right)^2 = \frac{\left(\int_0^1 \log\left(\frac{\lambda_*^{(1)}(t)}{\lambda_*^{(0)}(t)}\right) f_*^{(0)}(t)dt\right)^2}{(\log HR_1)^2 \left(\int_0^1 f_*^{(0)}(t)dt\right)\left(\int_0^1 f_1^{(0)}(t)dt\right)}$$

where $f_1^{(0)}(t)$ and $f_*^{(0)}(t)$ are, respectively, the densities for T_1 and T_* in group 0.

Therefore, the choice of one copula or another changes the value of the density function $f_*^{(0)}(t)$ for T_* in group 0. It is interesting to study the behavior of this density function depending on the chosen copula. For this reason, some graphical exploratory analysis has been performed.

If the density functions of $f_*^{(0)}(t)$ and $f_*^{(1)}(t)$ are plotted for Frank and Gumbel copula, it can be observed that there are not big differences between them. As a way of example, Figure 5.4 shows these plots for two cases: the first plot corresponds to $\beta_1 = 0.5$, $\beta_2 = 2$, $\rho = 0.45$, $HR_1 = 0.8$, $HR_2 = 0.5$ and $p_1 = 0.05$ and $p_2 = 0.3$ for censoring case 3; and the second one to $\beta_1 = \beta_2 = 2$, $\rho = 0.65$, $HR_1 = 0.7$, $HR_2 = 0.3$ and $p_1 = p_2 = 0.05$ for censoring case 1.

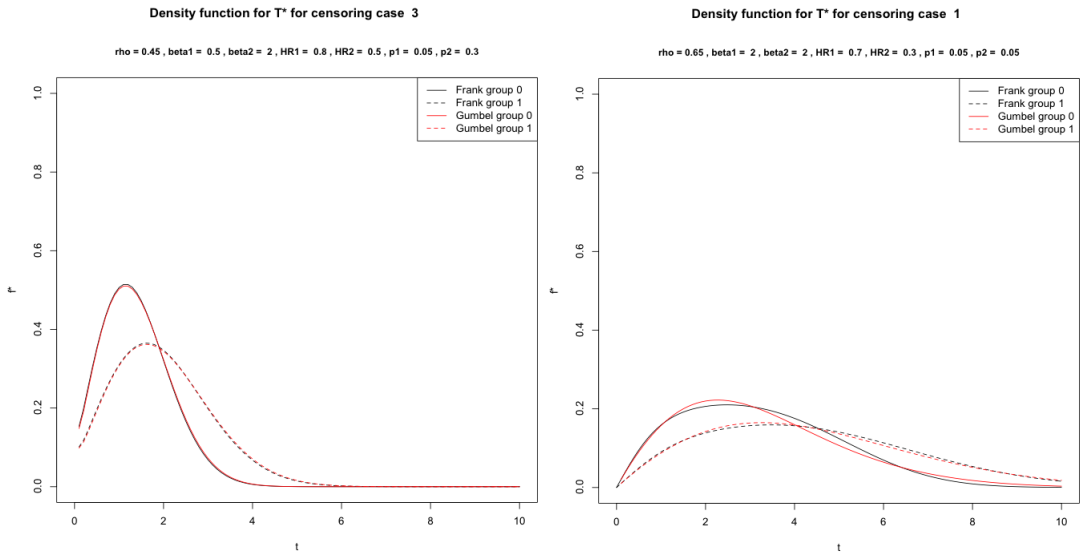


Figure 5.4: Density functions for T_* for treatment groups 0 and 1 and Frank and Gumbel copula.

It has been observed that density functions for both copulas are similar. Therefore, it is still necessary for future research to prove this methodology based on the ARE values for more extreme copulas in order to prove its robustness.

Chapter 6

Concluding remarks and future research

This master thesis has contributed to a deeper understanding of some concepts that I had already seen in some of the courses of the master. For example, I have studied theoretical concepts of survival analysis or theory about asymptotic behavior in order to understand the methodology developed by Gómez and Lagakos [1]. It has allowed me to learn more about some other concepts such as measures of dependence or multivariate distribution or survival functions and, moreover, it has allowed me to discover new ones, such as the copulas. I had never heard about copulas until I started this master thesis and, hence, I needed to do great bibliographic research about this field and study them deeply in order to understand their meaning and the difference between different families of copulas.

The main objective of this master thesis was to study the robustness of the methodology developed by Gómez and Lagakos [1] when changing the copula considered for the construction of the joint distribution function of the two possible endpoints. The results of this project showed that the method is robust considering Frank and Gumbel copulas because they agreed in more than 98% of the cases in both recommending the use of the composite or the relevant endpoint. Furthermore, the ARE values were similar between both copulas and the difference between the two values in discordant cases was very small. However, and as concluded in chapter 5, it is necessary to prove the methodology for more extreme copulas because it has been seen that Frank and Gumbel copulas yield to similar density functions for the composite endpoint. As a first step for future research, the expression of the ARE for Clayton copula has been developed in Appendix D and it has been included in the R function created to compute the ARE value (Code in Appendix A).

In addition, this master thesis refreshed and improved my writing skills using LaTeX and my programming skills using the statistical software R. The PhD student Moises Gómez adapted the Maple code used in Gómez and Lagakos methodology [1] in order to use it with R. One of my contributions to the research in composite endpoints is that I took the codes for censoring cases 1 and 3 and improved it in order to make the iterations faster in simulation studies and to extend

it easily to other copulas or other marginal distribution laws. These improvements enabled me to test the methodology for a higher number of combinations than the ones used until now. On the other hand, I learned how to use the R packages `copula` [19, 20, 21], in order to find a numerical approximation of the dependence parameter θ given the correlation ρ; and `rpanel` [22], a useful tool to construct a graphical display that allows the user to interact with a plot in a very effective manner. The code used for generating pairs of random variables using different copulas or plotting the distribution function of the density function of T_* using `rpanel` can be found in Appendix E. In addition, the software R, jointly with Microsoft Excel and Powerpoint, enabled me to create all the figures of this master thesis.

This project could be considered as a part of the investigations that professor Guadalupe Gómez and the rest of the research group in survival analysis GRASS are doing on composite endpoints. This line of research is one of the mainstays of their research project for the following years. Moreover, I would like to remark that this master thesis made me discover a passion for the research and it gave me the opportunity to work with GRASS group researchers and took benefit of either their knowledge and the work previously done.

It is important to keep on working to make this methodology as extended and applicable possible. Therefore, this master thesis can be considered as the prelude of my Doctoral Thesis, based on this research area. Some of the issues that could be carried out in the near future are the following:

- Extend the comparison between the ARE values computed via Frank or Gumbel copula to censoring cases 2 and 4, with competing risks.
- Study the most common distribution laws for survival analysis and extend this methodology for laws other than Weibull.
- Two different random variables with a given correlation yield to different joint distribution functions depending on the copula used. Therefore, it is necessary to study the robustness of this method for copulas different than Frank and Gumbel.
- Develop recommendations to decide when it is necessary to extend the relevant endpoint to the composite based on the different parameter values amplifying the possibilities by different tests as weighted log-rank test and other from the Fleming-Harrington family.
- Proportional hazards between the two outcomes have been considered in the research done. It is important to develop different hypotheses in the case that proportionality does not hold.
- Develop the statistical methodology if we consider the main outcome as a binary variable instead of times to the event of interest.

I am very pleased of having the opportunity to help in the development of this methodology because it will yield to numerous beneficial repercussions. The application of this new method

of choosing the main endpoint in randomized clinical trials is important for the pharmaceutical industry and other research institutions because it allows the design of better trials and increases their power. It will help in the improvement of the health of the people and will reduce the number of patients to be included in trials. Therefore, it will contribute in reducing unnecessary costs which I think it is very important, especially now that the economic adjustments are in order of the day.

Bibliography

[1] Gómez, G. and Lagakos, S.W. *Statistical considerations when using a composite endpoint for comparing treatment groups*. Statistics in Medicine. Accepted, 2012.

[2] Gómez, M. *Composite Endpoints for Stent Cardiovascular Clinical Trials*. Master Thesis in Interuniversity Master in Statistics and Operations Research, 2011.

[3] Nelsen, Roger B. *An Introduction to Copulas*. Lecture Notes in Statistics. Springer-Verlag, 1999.

[4] Trivedi, P.K. and Zimmer, D.M. *Copula Modelling: An Introduction for Practitioners*. Foundation and Trends in Econometrics, 2007.

[5] Oxford Reference Online Dictionary (*http://www.oxfordreference.com*).

[6] Ferreira-González, I., Permanyer-Miralda, G., Busse, J.W., Bryant, D.M., Montori, V.M., Alonso-Coello, P., Walter, S.D., Guyatt, G.H. *Methodological discussions for using and interpreting composite endpoints are limited, but still identify major concerns*. Journal of Clinical Epidemiology, 2007.

[7] Gómez, G. *Some theoretical thoughts when using a composite endpoint to prove the efficacy of a treatment*. Proceedings of the 26th International Workshop on Statistical Modelling 2011.

[8] Freemantle, N., Calvert, M., Wood, J., Eastaugh, J., Griffin, C. *Composite outcomes in Randomized Trials. Greater precision but with greater uncertainty?*. Journal of the American Medical Association, 2003.

[9] Montori, V.M., Permanyer-Miralda, G., Ferreira-González, I., Busse, J.W., Pachecho-Huergo, V., Bryant, D., Alonso, J., Akl, E.A., Domingo-Salvany, A., Mills, E., Wu, P., Schünemann, H.J., Jaeschke, R., Guyatt, G.H. *Validity of composite endpoints in clinical trials*. British Medical Journal, 2007.

[10] Ferreira-González, I., Busse, J.W., Heels-Ansdell, D., Montori, V.M., Akl, E.A., Bryant, D.M., and others. *Problems with use of composite endpoints in cardiovascular trials: systematic review of randomized controlled trials*. British Medical Journal, 2007.

[11] Gómez, G. *Análisis de la Supervivéncia*. Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya, 2004.

[12] Lagakos, S.W. and Schoenfeld, D. *Properties of Proportional-Hazard Score Tests under Misspecified Regression Models*. Biometrics, 1984.

[13] Schoenfeld, D. *Sample-size formula for the proportional-hazards regression model*. Biometrics, 1983.

[14] Fisher, N.I. *Copulas*. Encyclopedia of Statistical Sciences, Update Vo. 1. John Wiley and Sons, New York, 1997.

[15] Cassell's Latin Dictionary.

[16] Oxford English Dictionary.

[17] Fredricks, Gregory A. and Nelsen, Roger B. *On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables*. Journal of Statistical Planning and Inference, 2006.

[18] Venter, G. G. *Tails of Copulas*. Presented at ASTIN Colloquium, Inernational Actuarial Association, Washington D.C., 2001.

[19] Yan, J. *Enjoy the Joy of Copulas: With a Package copula*. Journal of Statistical Software, 21(4), 1-21, 2007.

[20] Kojadinovic, I. and Yan, J. *Modeling Multivariate Distributions with Contiuous Margins Using the copula R Package*. Journal of Statistical Software, 34(9), 1-20, 2010.

[21] Hofert, M., Maechler, M. *Nested Archimedean Copulas Meet R: The nacopula Pakage*. Journal of Statistical Software, 39(9), 1-20, 2011.

[22] Bowman, A., Crawford, E., Alexander, G., Bowman, R.W. *rpanel: Simple Interactive Controls for R Functions Using the tcltk Package*. Journal of Statistical Software, 17(9), 1-18, 2007.

Appendix A

R code for ARE computations

The PhD student Moises Gómez adapted the Maple code used in Gómez and Lagakos methodology [1] in order to use it with R. My contribution to improve these scripts has been:

- Create a function `ARE(rho,beta1,beta2,HR1,HR2,p1,p2,case,copula)` replacing the several scripts, one for each censoring case, in order to make it easier to compute an ARE value for given parameters instead of making the scripts run for each combination.
- Compute the dependence parameter θ for a given ρ using the R-package `copula` [19, 20, 21] in order to make it faster.
- Compute the ARE value for Gumbel and Clayton copulas.

These improvements have enabled me to test the methodology for a higher number of combinations that the ones used until now and compare the results for the different copulas.

```
1  ##############################################################################
2  #                    ARE_case13.R
3  ##############################################################################
4  # Computation of the Asymptotic relative Efficiency (ARE) values for censoring
5  # cases 1 and 3 and copulas Frank, Gumbel and Clayton.
6  #
7  ##############################################################################
8  # This is an adaptation of the following scripts:
9  #            Case_1_One_Scenario.R
10 #            Case_3_One_Scenario.R
11 #                Last update: 07/01/2012
12 #                R version: R 2.9.2
13 #                Author: MoisÈs GÛmez Mateu (moises.gomez.mateu@upc.edu)
14 #
15 # Improvements:
16 #     Generalize the scripts in one unique function for cases 1 and 3
17 #     Compute the ARE for Gumbel and Clayton copula
18 #     Computation of dependence parameter theta using package copula
19 #     Generalize the scripts in one unique function for Frank, Gumbel and Clayton copula
20 ##############################################################################
21 #
22 # CASE 1: The composite endpoint does not include a fatal event (i.e. Death)
```

```
# in the relevant endpoint neither in the additional.
#
# CASE 3: The composite endpoint does includes a fatal event (i.e. Death)
# in the relevant endpoint but it does not in the additional.
#
# Last update: 05/06/2012
#
# R version: R 2.10.1
#
# Author: Oleguer Plana Ripoll (oleguerplana@gmail.com)
#
################################################################################
#
# Notation:
#
# HR1: Hazard ratio for the relevant endpoint E1.
# HR2: Hazard ratio for the additional endpoint E2.
# p10: Probability of occurence of the relevant event (or endpoint) in group zero (or placebo).
# p20: Probability of occurence of the additional event in group zero.
# p11: Probability of occurence of the relevant event (or endpoint) in treatment group.
# p21: Probability of occurence of the additional event in treatment group.
#
# beta1: Shape parameter for a Weibull law for the relevant event in both groups.
# beta2: Shape parameter for a Weibull law for the additional event in both groups.
#
# b10: Scale parameter for a Weibull law for the relevant event in group zero.
# b20: Scale parameter for a Weibull law for the additional event in group zero.
# b11: Scale parameter for a Weibull law for the relevant event in treatment group.
# b21: Scale parameter for a Weibull law for the additional event in treatment group.
#
# T1: Time to observe the relevant endpoint E1.
# T2: Time to observe the additional endpoint E2.
# rho: Spearman's coefficient between T1 and T2.
#
# References:
# GÛmez G. and Lagakos S. Statistical Considerations in the Use of a Composite
# Time-to-Event Endpoint for Comparing Treatment Groups. Accepted (2012).
#
################################################################################

install.packages("copula")
library(copula)

################################################################################################
# Function: ARE
#
################################################################################################
# Descricption: It computes the ARE value for the given arguments
#
# rho   Spearman's coefficient that we set
# beta1   Shape parameter for a Weibull law for the relevant event
# beta2   Shape parameter for a Weibull law for the additional event
# HR1     Hazard Ratio for a Weibull law for the relevant event
# HR2     Hazard Ratio for a Weibull law for the additional event
# p1      Proportion of the relevant event expected in group zero
# p2      Proportion of the additional event expected in group zero
# case    Censoring case -- > 1 (default) or 3
# copula  Copula used --> "Frank" (default), "Gumbel" or "Clayton"
```

```
################################################################################

ARE<-function(rho, beta1, beta2, HR1, HR2, p1, p2, case = 1, copula="Frank")
{

 ################################################################################
 # Function calibSpearmansRho computes Theta from values of Rho (package copula)
 ################################################################################

 if(copula=="Frank") {
  theta <- calibSpearmansRho(frankCopula(0),rho)
 }

 if(copula=="Gumbel") {
  theta <- calibSpearmansRho(gumbelCopula(1),rho)
 }

 if(copula=="Clayton") {
  theta <- calibSpearmansRho(claytonCopula(1),rho)
 }


 ####### ASSESSMENT OF THE SCALE PARAMETER VALUES b10, b11, b20, b21
 ## b20 is diferent for case 1 or 3

 # b10 and b11 are the same for case 1 or 3
 b10 <- 1/((-log(1-p1))^(1/(beta1)))
 b11 <- b10/HR1^(1/beta1)

 if(case==1) {
  b20 <- 1/(-log(1-p2))^(1/beta2)
 } else
  if (case==3) {
   ################################################################################
   # Function: Fb20
   ################################################################################
   # Description: It computes b20 value for case 3
   # Arguments:
   # b20
   # p2    Probability of observing the additional endpoint
   ################################################################################

   Fb20<-function(b20,p2) {
    integral<-integrate(function(y) {
     sapply(y,function(y) {
      integrate(function(x)((theta*(1-exp(-theta))*exp(-theta*(x+y)))
      /(exp(-theta)+exp(-theta*(x+y))-exp(-theta*x)-exp(-theta*y))^2),lower=0,
      upper=exp(-((((-log(y))^(1/beta2))*b20)/b10)^beta1))$value
     })
    },
    lower= exp(-(1/b20)^beta2), upper=1)$value
    return(integral-p2)
   }
   limits <- c(0.00001,10000)
   b20 <- uniroot(Fb20, interval=limits,p2=p2)$root
  }

 # b21 is the same for case 1 or 3
```

```
b21 <- b20/HR2^(1/beta2)

############## ARE (numerator and denominator) following Gomez and Lagakos paper

numerador <- function(t,b10,b11,b20,b21,beta1,beta2,theta,fT10,fT11,fT20,fT21,
 ST10,ST11,ST20,ST21,Sstar0,fstar0,Lstar0,Sstar1,fstar1,Lstar1,HRstar,logHRstar) {
 fT10 <- (beta1/b10) * ( (t/b10)^(beta1-1) ) * (exp(-(t/b10)^beta1))
 fT11 <- (beta1/b11) * ( (t/b11)^(beta1-1) ) * (exp(-(t/b11)^beta1))
 fT20 <- (beta2/b20) * ( (t/b20)^(beta2-1) ) * (exp(-(t/b20)^beta2))
 fT21 <- (beta2/b21) * ( (t/b21)^(beta2-1) ) * (exp(-(t/b21)^beta2))
 ST10 <- exp(-(t/b10)^beta1)
 ST11 <- exp(-(t/b11)^beta1)
 ST20 <- exp(-(t/b20)^beta2)
 ST21 <- exp(-(t/b21)^beta2)

 if(copula=="Frank") {
  Sstar0 <- (-log(1+(exp(-theta*ST10)-1)*(exp(-theta*ST20)-1)/(exp(-theta)-1))/theta)

  fstar0 <- (exp(-theta*ST10)*(exp(-theta*ST20)-1)*fT10+exp(-theta*ST20)*
  (exp(-theta*ST10)-1)*fT20)/(exp(-theta*Sstar0)*(exp(-theta)-1))

  Sstar1 <- (-log(1+(exp(-theta*ST11)-1)*(exp(-theta*ST21)-1)/(exp(-theta)-1))/theta)

  fstar1 <- (exp(-theta*ST11)*(exp(-theta*ST21)-1)*fT11+exp(-theta*ST21)*
  (exp(-theta*ST11)-1)*fT21)/(exp(-theta*Sstar1)*(exp(-theta)-1))
 }

 if(copula=="Gumbel") {
  Sstar0 <- ST10 + ST20 -1 + exp(-(((-log(1-ST10))^theta+(-log(1-ST20))^theta)^(1/theta)))

  fstar0 <- fT10 + fT20 - (exp(-(((-log(1-ST10))^theta+(-log(1-ST20))^theta)^(1/theta))))*
  (((-log(1-ST10))^theta+(-log(1-ST20))^theta)^((1-theta)/theta))*
  (((-log(1-ST10))^(theta-1))*(fT10/(1-ST10))+((-log(1-ST20))^(theta-1))*(fT20/(1-ST20)))

  Sstar1 <- ST11 + ST21 -1 + exp(-(((-log(1-ST11))^theta+(-log(1-ST21))^theta)^(1/theta)))

  fstar1 <- fT11 + fT21 - (exp(-(((-log(1-ST11))^theta+(-log(1-ST21))^theta)^(1/theta))))*
  (((-log(1-ST11))^theta+(-log(1-ST21))^theta)^((1-theta)/theta))*
  (((-log(1-ST11))^(theta-1))*(fT11/(1-ST11))+((-log(1-ST21))^(theta-1))*(fT21/(1-ST21)))

 }

 if(copula=="Clayton") {
  Sstar0 <- ST10 + ST20 -1 + (((1-ST10)^(-theta))+((1-ST20)^(-theta))-1)^(-1/theta)

  fstar0 <- fT10 + fT20 - ((((1-ST10)^(-theta))+((1-ST20)^(-theta))-1)^(-(1+theta)/theta))*
      (((1-ST10)^(-theta-1))*fT10+((1-ST20)^(-theta-1))*fT20)

  Sstar1 <- ST11 + ST21 -1 + (((1-ST11)^(-theta))+((1-ST21)^(-theta))-1)^(-1/theta)

  fstar1 <- fT11 + fT21 - ((((1-ST11)^(-theta))+((1-ST21)^(-theta))-1)^(-(1+theta)/theta))*
      (((1-ST11)^(-theta-1))*fT11+((1-ST21)^(-theta-1))*fT21)

 }

 Lstar0 <- (fstar0/Sstar0)
 Lstar1 <- (fstar1/Sstar1)
 HRstar <- (Lstar1/Lstar0)
```

```
 logHRstar <- log(HRstar)

 return(logHRstar*fstar0)
}

numerador1<-integrate(numerador,lower=0,upper=1, b10,b11,b20,b21,beta1,beta2,
theta,fT10,fT11,fT20,fT21,ST10,ST11,ST20,ST21,Sstar0,fstar0,Lstar0,Sstar1,fstar1,
Lstar1,HRstar,logHRstar,subdivisions=1000,stop.on.error = FALSE)

numerador2<-(numerador1$value)^2

ST10_1 <- exp(-(1/b10)^beta1)
ST20_1 <- exp(-(1/b20)^beta2)

if(copula=="Frank") {
 Sstar0_1 <- (-log(1+(exp(-theta*ST10_1)-1)*(exp(-theta*ST20_1)-1)/(exp(-theta)-1))/theta)
}

if(copula=="Gumbel") {
 Sstar0_1 <- ST10_1 + ST20_1 -1 + exp(-(((-log(1-ST10_1))^theta+(-log(1-ST20_1))^theta)^(1/
     theta)))
}

if(copula=="Clayton") {
  Sstar0_1 <- ST10_1 + ST20_1 -1 + (((1-ST10_1)^(-theta))+((1-ST20_1)^(-theta))-1)^(-1/theta)
}

ST10_1 <- exp(-(1/b10)^beta1)

denominador <- ((log(HR1))^2)*(1-Sstar0_1)*(1-ST10_1)

AREstarT <- (numerador2/denominador)

# IF THE VALUE THE NUMERATOR IS NOT COMPUTED, THEN WE ASSIGN A MISSING IN THE ARE VALUE
if(numerador1$message!="OK") {AREstarT <- NA}

# ARE VALUE:
return(AREstarT )

}


#############################################################################
#############################################################################
# EXAMPLES
#############################################################################
#############################################################################

ARE(rho=0.2,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Frank")
ARE(rho=0.2,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Gumbel")
ARE(rho=0.2,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Clayton")

ARE(rho=0.5,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Frank")
ARE(rho=0.5,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Gumbel")
ARE(rho=0.5,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Clayton")

ARE(rho=0.8,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Frank")
ARE(rho=0.8,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Gumbel")
```

```
252 ARE(rho=0.8,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Clayton")
253
254 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Frank")
255 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Gumbel")
256 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Clayton")
257
258 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.5,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Frank")
259 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.5,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Gumbel")
260 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.5,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Clayton")
261
262 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.7,HR2=0.6,p1=0.3,p2=0.5,case=1,copula="Frank")
263 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.7,HR2=0.6,p1=0.3,p2=0.5,case=1,copula="Gumbel")
264 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.7,HR2=0.6,p1=0.3,p2=0.5,case=1,copula="Clayton")
265
266 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.9,HR2=0.4,p1=0.6,p2=0.3,case=1,copula="Frank")
267 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.9,HR2=0.4,p1=0.6,p2=0.3,case=1,copula="Gumbel")
268 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.9,HR2=0.4,p1=0.6,p2=0.3,case=1,copula="Clayton")
269
270 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.5,HR2=0.6,p1=0.6,p2=0.3,case=1,copula="Frank")
271 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.5,HR2=0.6,p1=0.6,p2=0.3,case=1,copula="Gumbel")
272 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.5,HR2=0.6,p1=0.6,p2=0.3,case=1,copula="Clayton")
273
274 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.5,HR2=0.6,p1=0.3,p2=0.7,case=1,copula="Frank")
275 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.5,HR2=0.6,p1=0.3,p2=0.7,case=1,copula="Gumbel")
276 ARE(rho=0.2,beta1=2,beta2=2,HR1=0.5,HR2=0.6,p1=0.3,p2=0.7,case=1,copula="Clayton")
277
278 ##############################################################################
279 ##############################################################################
280 # APPENDIX
281 ##############################################################################
282 ##############################################################################
283
284 # WEIBULL DISTRIBUTION
285 ?dweibull
286
287 # DENSITY FUNCTION
288 dweibull(x, shape, scale = 1, log = F)
289 f(t) = (beta1/b10) * ( (t/b10)^(beta1-1) ) * (exp(-(t/b10)^beta1))
290
291 # CUMULATIVE DISTRIBUTION FUNCTION => S(t) = 1 - F(t)
292 pweibull(q, shape, scale = 1, lower.tail = T, log.p = F)
293 F(t) = exp(-(t/b10)^beta1)
```

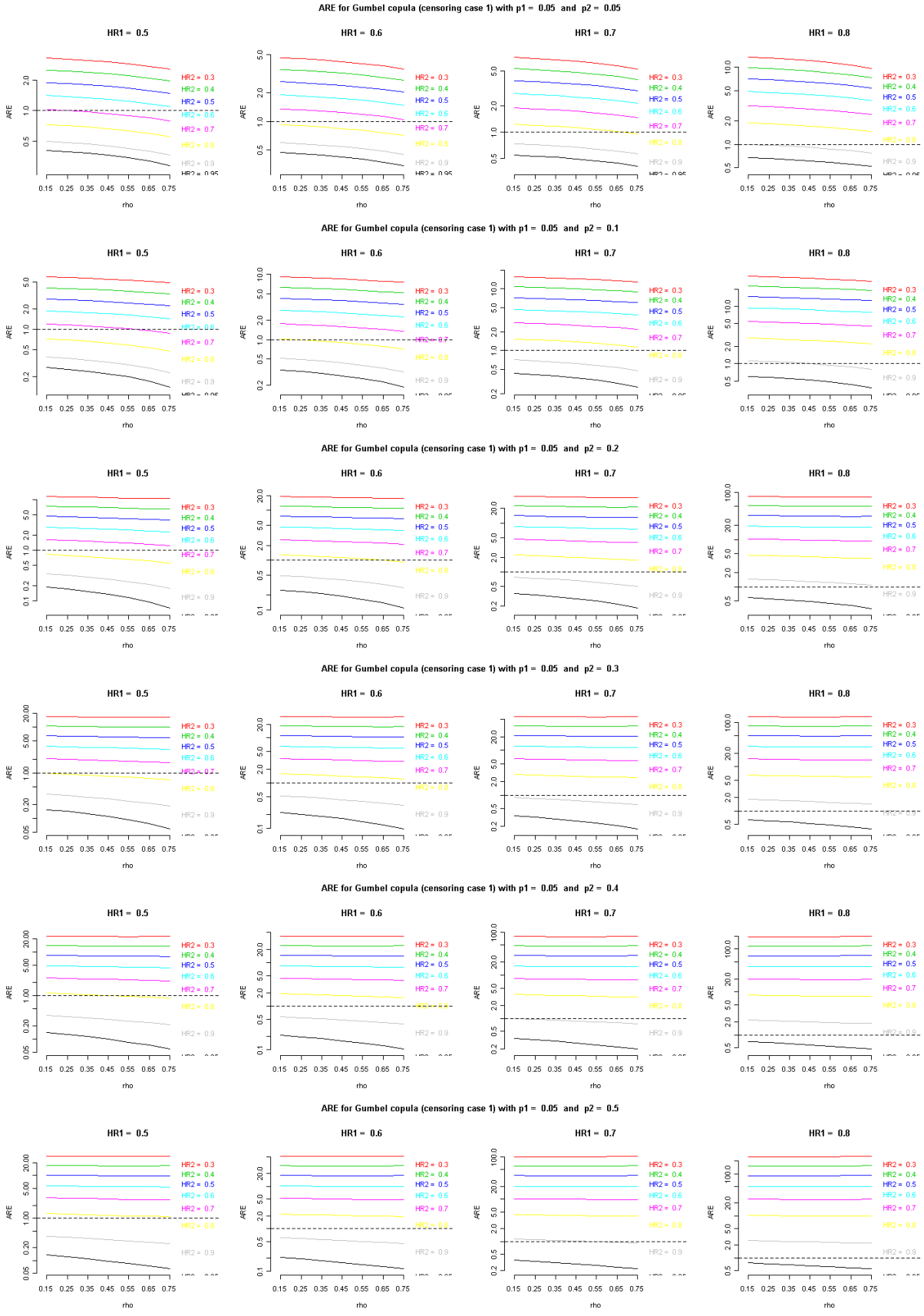
../R/Case_13_One_Scenario_OLEGUER.r

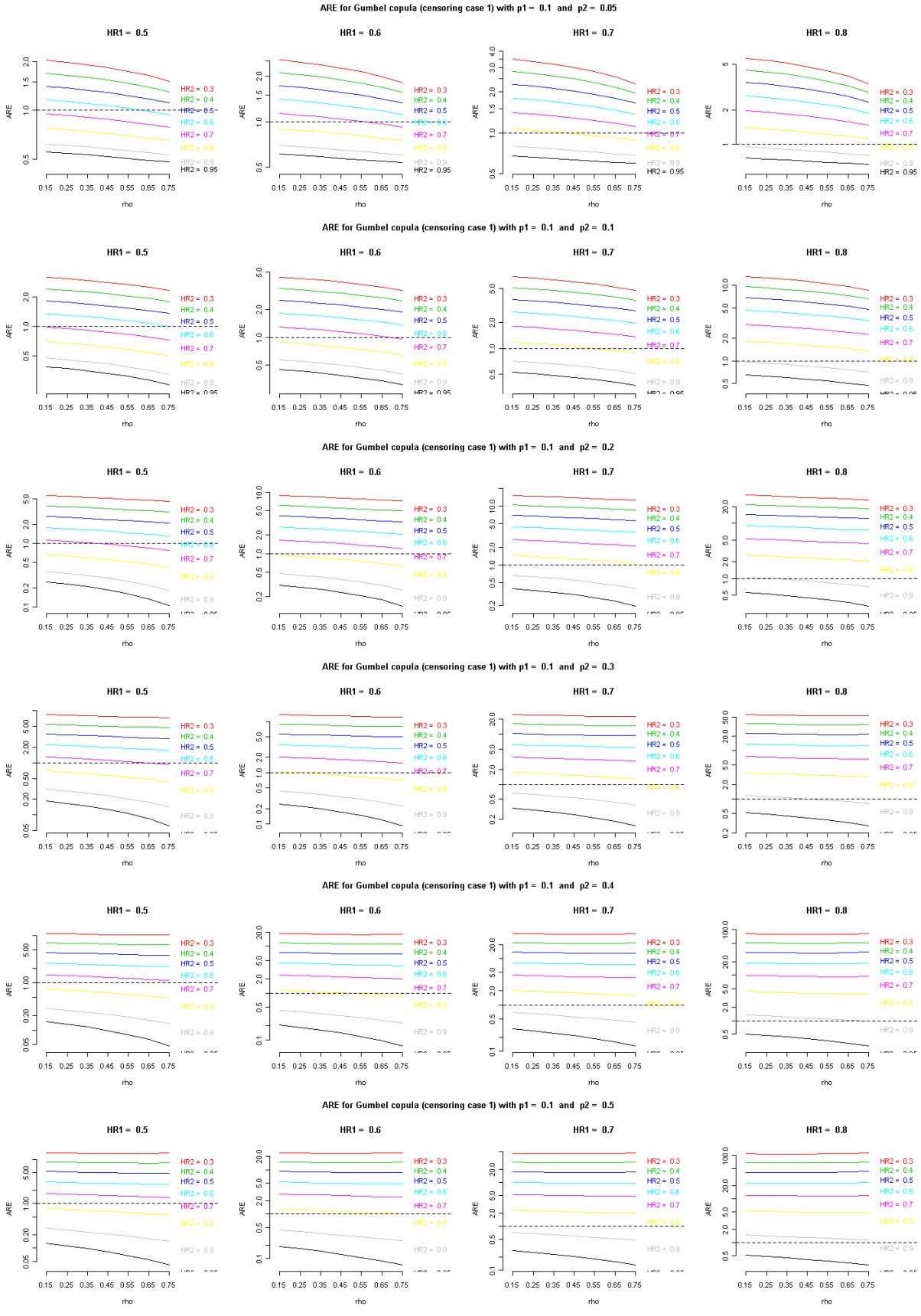
Appendix B

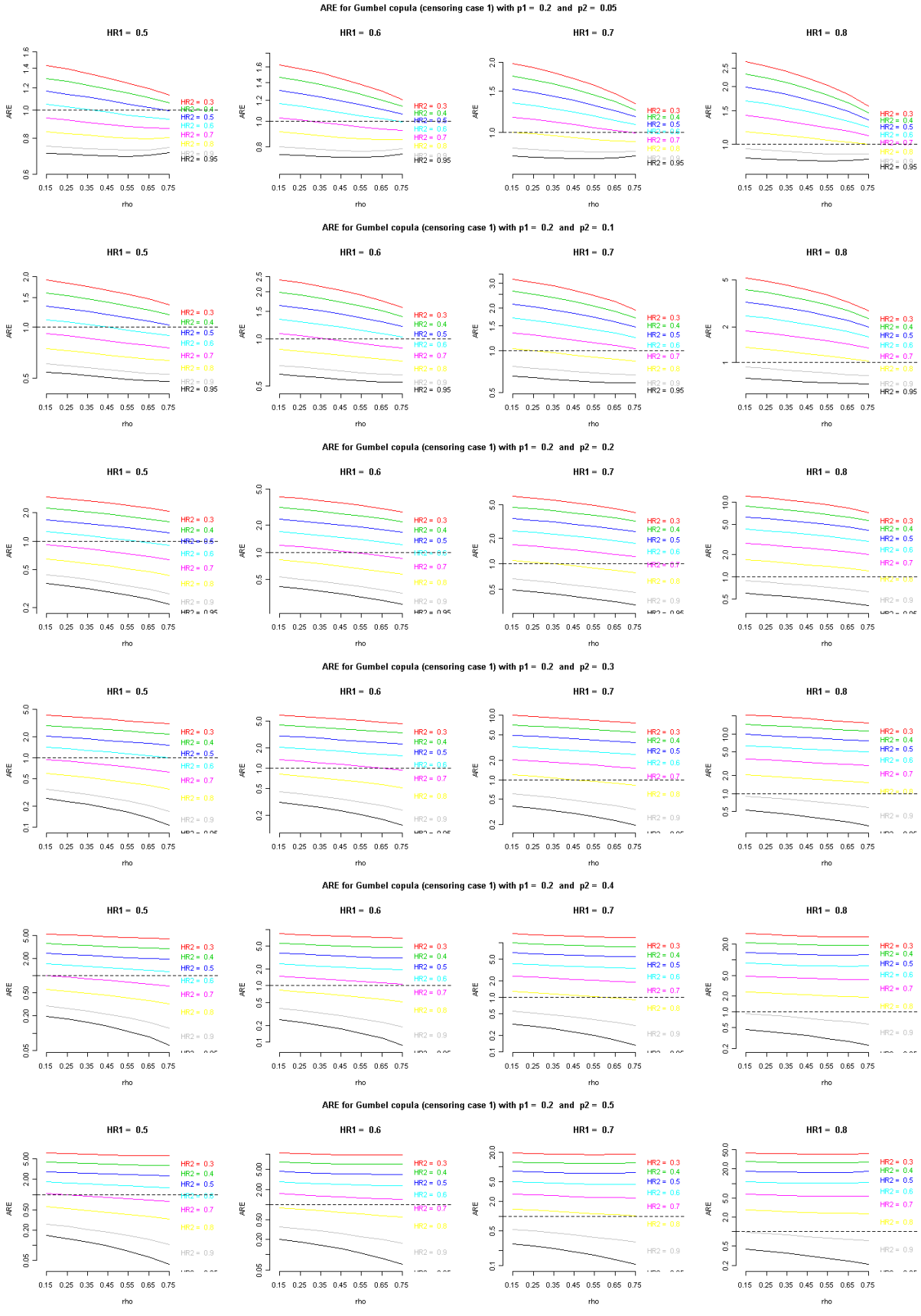
ARE plots depending on HR_1, HR_2, p_1, p_2 and ρ for censoring cases 1 and 3

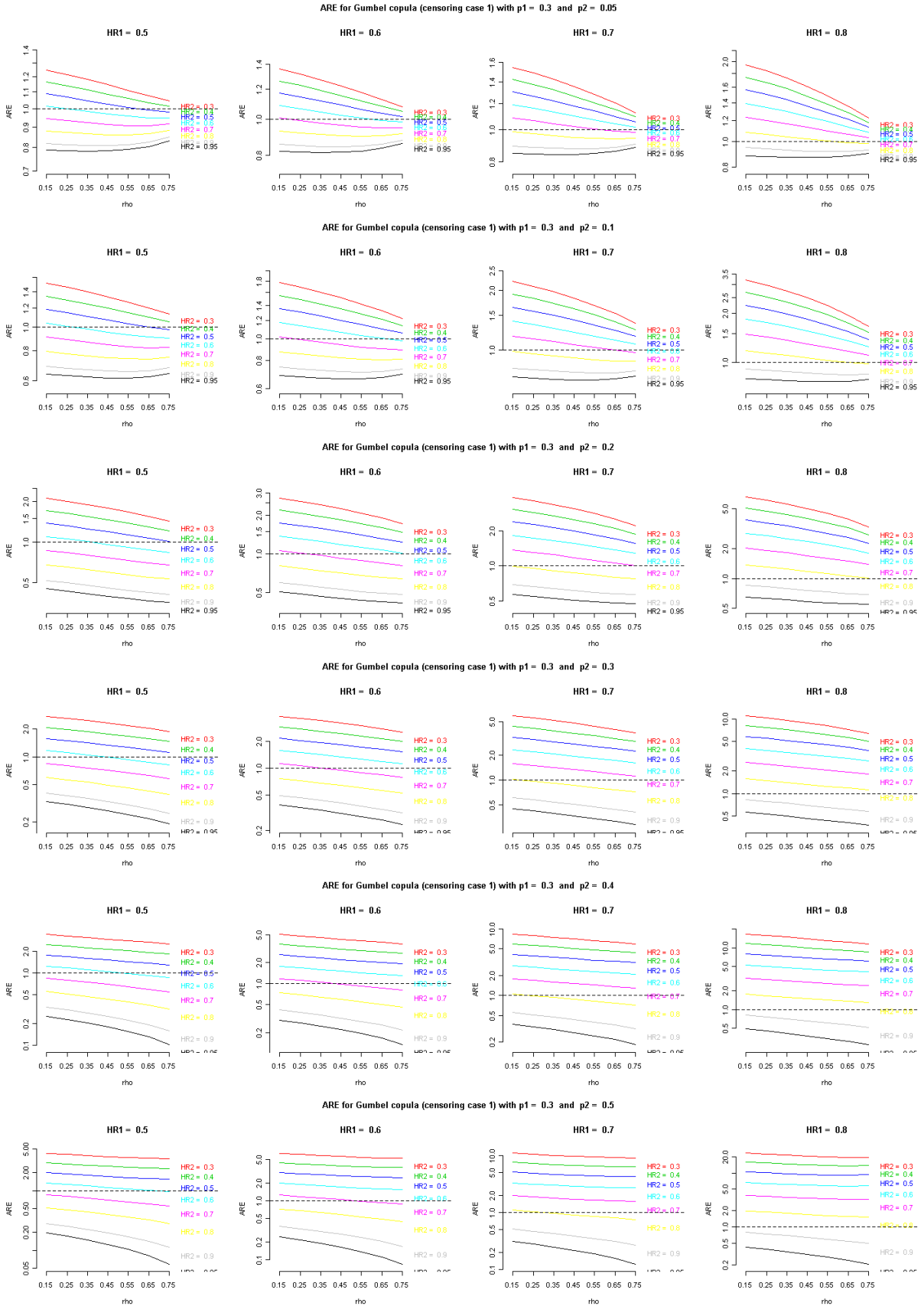
The following pages contain the plots for the particular combination of $\beta_1 = \beta_2 = 1$ and the remaining parameters given in Table 4.1, where there are 8.064 different configurations for each censoring case 1 and 3. For each such case, these are grouped for specific values of p_1, p_2 and HR_1 yielding a total number of 144 scenarios for each censoring case 1 and 3. Therefore, the following 144 plots are grouped in the following way: the first 6 pages correspond to censoring case 1 and the following 6 to censoring case 3; each page correspond to a different value of p_1 and it contains 24 plots; within each page, there are 6 rows, each one corresponding to a different value of p_2; each row contains 4 plots, each one corresponding to a different value HR_1.

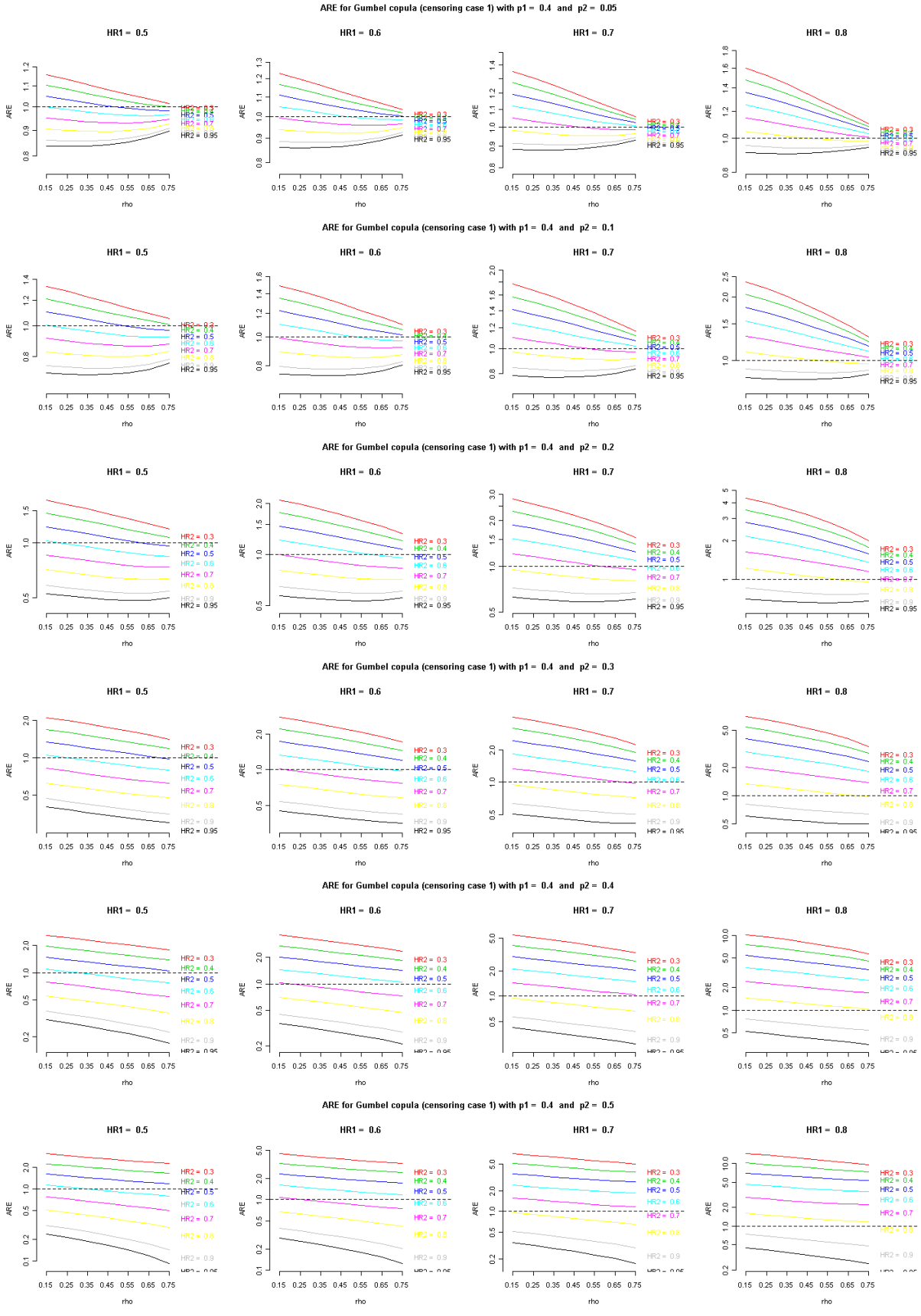
For every scenario, the 7 value of ρ and the 8 values of HR_2 are plotted as described below. Each one of the 144 plots displays 8 curves corresponding to 8 different values of the relative treatment effect on E_2 (HR_2). Each plot have Spearman's ρ ranging from 0.15 to 0.75 on the abscissa and the value of ARE on the ordinate, on a logarithmic scale. A logarithmic scale has been used since it represents better its significance. For example, an $ARE(Z_*, Z) = 2$ is as relevant as an $ARE(Z_*, Z) = 0.5$. That is, the distance from a point with $ARE(Z_*, Z) = 2$ to 1 is the same as the distance from a point with $ARE(Z_*, Z) = 0.5$ to 1.

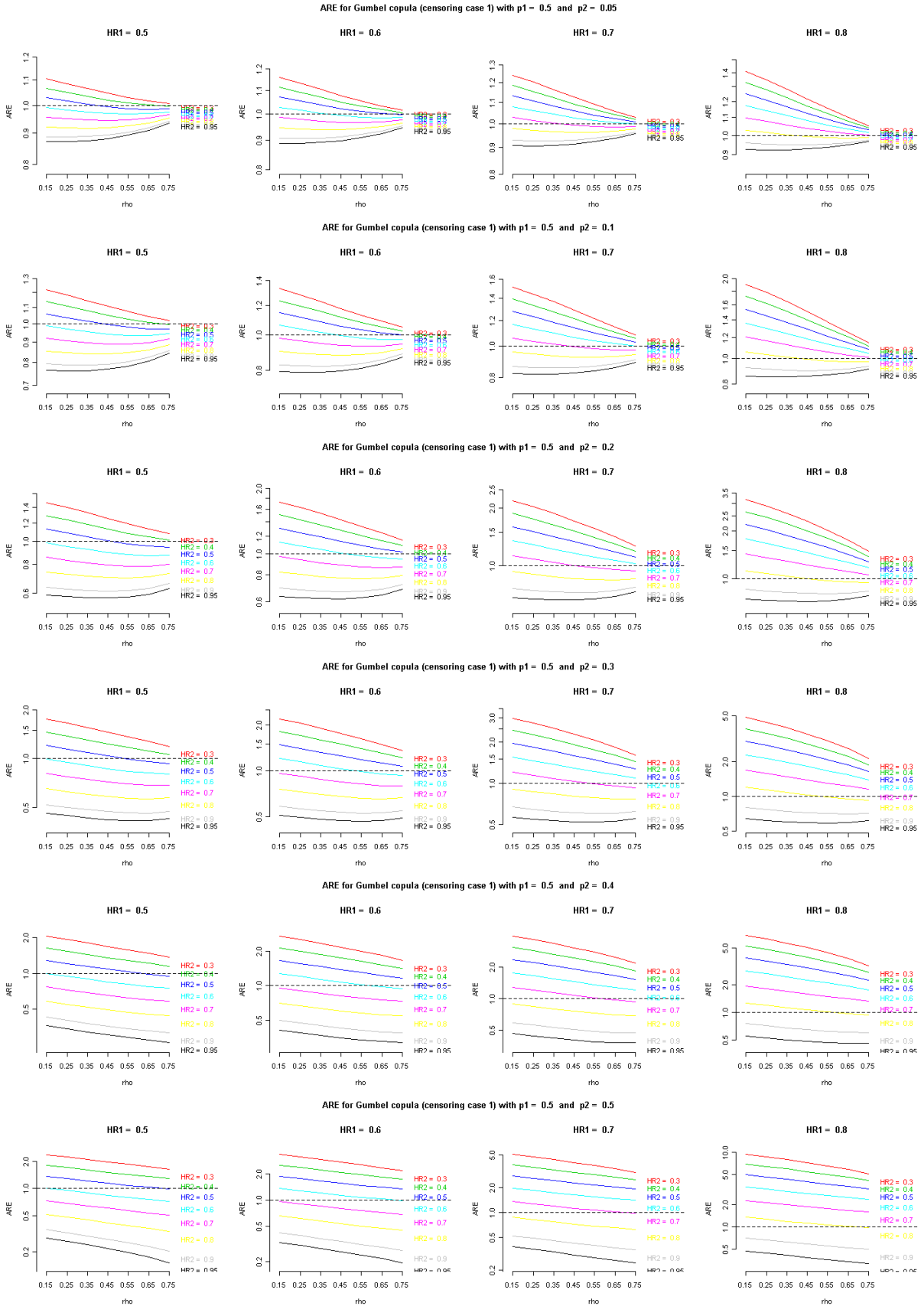

ARE for Gumbel copula (censoring case 1) with p1 = 0.05 and p2 = 0.05
ARE for Gumbel copula (censoring case 1) with p1 = 0.05 and p2 = 0.1
ARE for Gumbel copula (censoring case 1) with p1 = 0.05 and p2 = 0.2
ARE for Gumbel copula (censoring case 1) with p1 = 0.05 and p2 = 0.3
ARE for Gumbel copula (censoring case 1) with p1 = 0.05 and p2 = 0.4
ARE for Gumbel copula (censoring case 1) with p1 = 0.05 and p2 = 0.5
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
HR2 = 0.3
HR2 = 0.4
HR2 = 0.5
HR2 = 0.6
HR2 = 0.7
HR2 = 0.8
HR2 = 0.9
ARE
rho

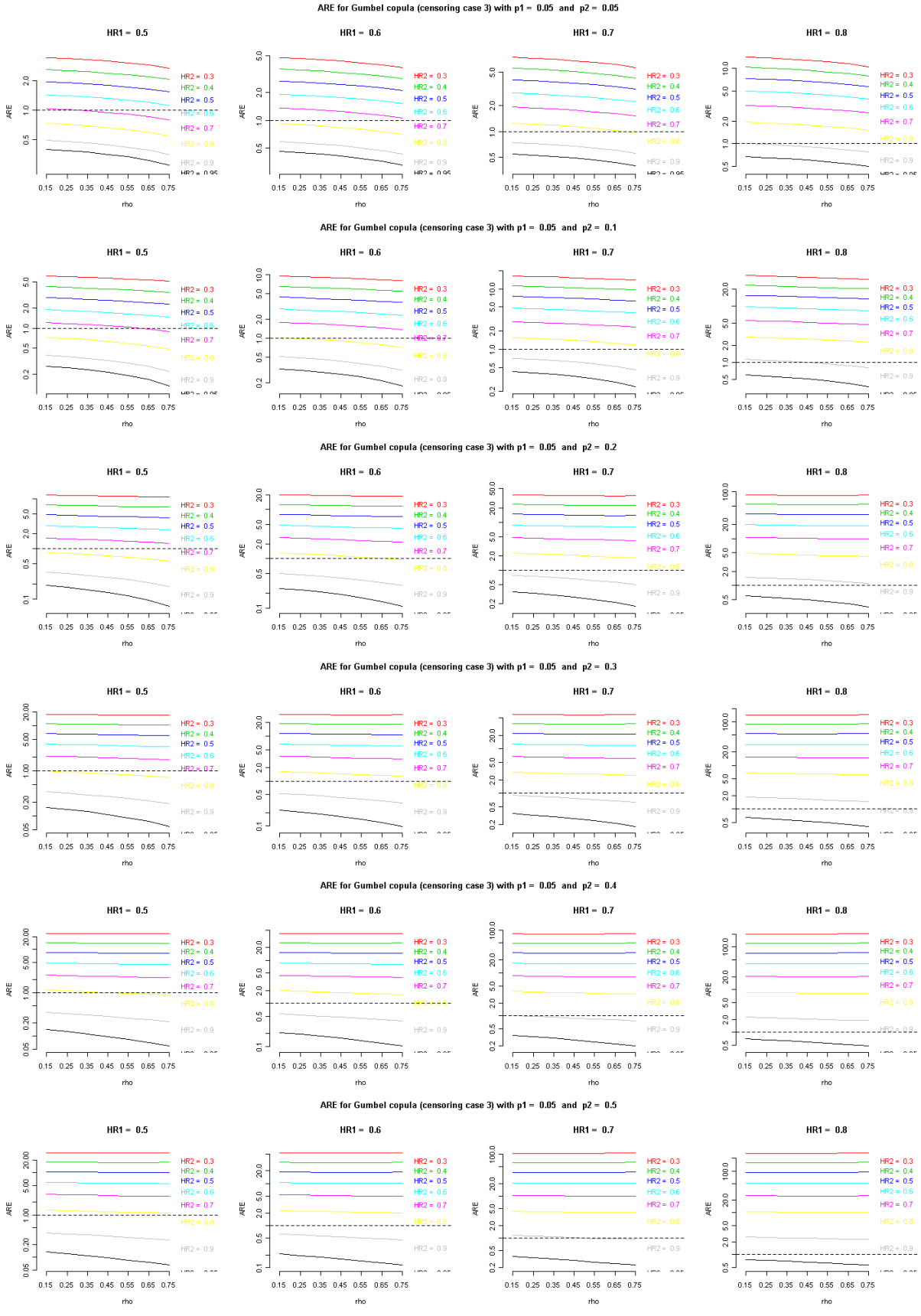

ARE for Gumbel copula (censoring case 1) with p1 = 0.1 and p2 = 0.05
ARE for Gumbel copula (censoring case 1) with p1 = 0.1 and p2 = 0.1
ARE for Gumbel copula (censoring case 1) with p1 = 0.1 and p2 = 0.2
ARE for Gumbel copula (censoring case 1) with p1 = 0.1 and p2 = 0.3
ARE for Gumbel copula (censoring case 1) with p1 = 0.1 and p2 = 0.4
ARE for Gumbel copula (censoring case 1) with p1 = 0.1 and p2 = 0.5
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE
rho
HR2 = 0.3
HR2 = 0.4
HR2 = 0.5
HR2 = 0.6
HR2 = 0.7
HR2 = 0.8
HR2 = 0.9
HR2 = 0.95

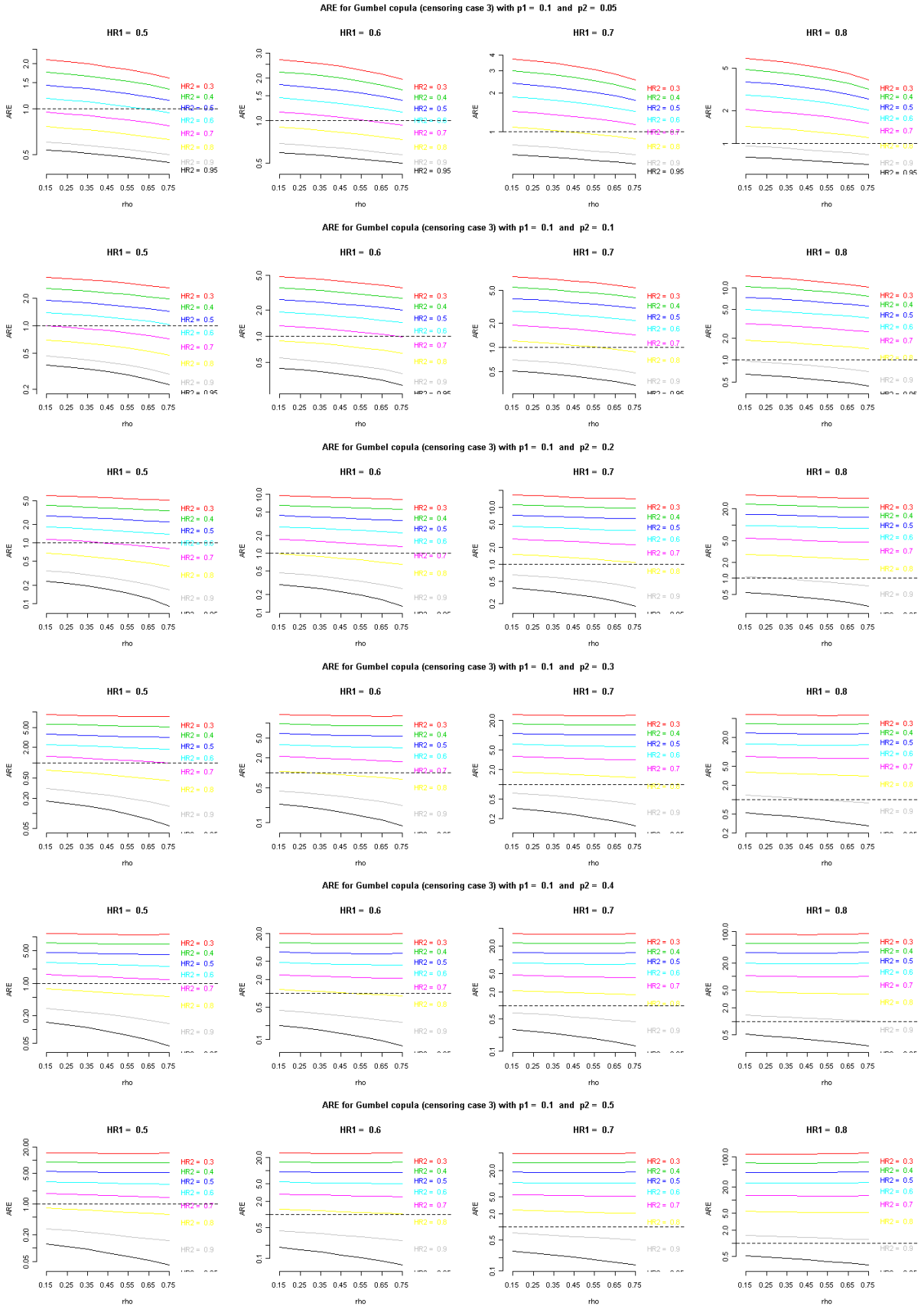

ARE for Gumbel copula (censoring case 1) with p1 = 0.2 and p2 = 0.05
ARE for Gumbel copula (censoring case 1) with p1 = 0.2 and p2 = 0.1
ARE for Gumbel copula (censoring case 1) with p1 = 0.2 and p2 = 0.2
ARE for Gumbel copula (censoring case 1) with p1 = 0.2 and p2 = 0.3
ARE for Gumbel copula (censoring case 1) with p1 = 0.2 and p2 = 0.4
ARE for Gumbel copula (censoring case 1) with p1 = 0.2 and p2 = 0.5
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE
rho
HR2 = 0.3
HR2 = 0.4
HR2 = 0.5
HR2 = 0.6
HR2 = 0.7
HR2 = 0.8
HR2 = 0.9
HR2 = 0.95

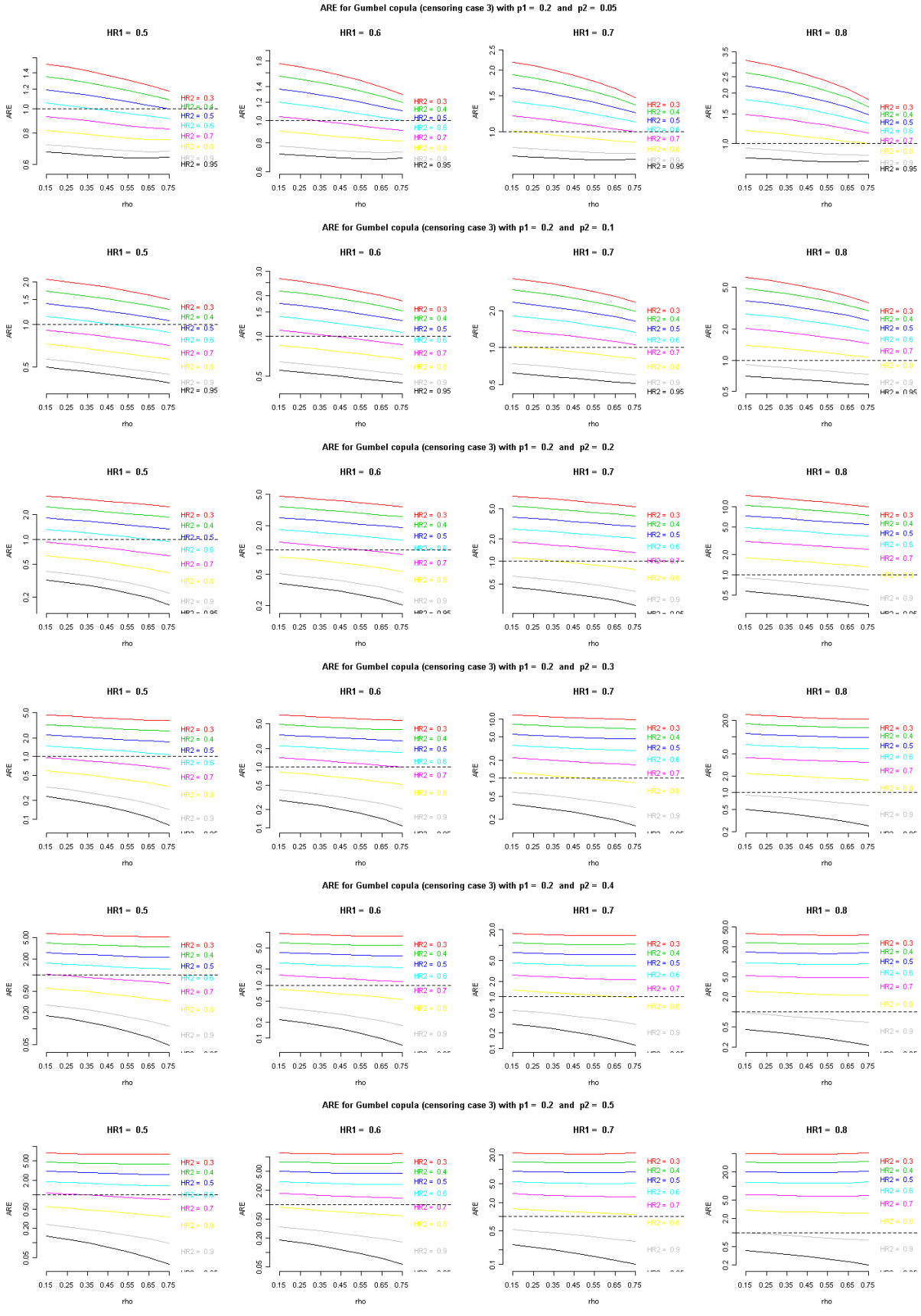

ARE for Gumbel copula (censoring case 1) with p1 = 0.3 and p2 = 0.05
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 1) with p1 = 0.3 and p2 = 0.1
ARE for Gumbel copula (censoring case 1) with p1 = 0.3 and p2 = 0.2
ARE for Gumbel copula (censoring case 1) with p1 = 0.3 and p2 = 0.3
ARE for Gumbel copula (censoring case 1) with p1 = 0.3 and p2 = 0.4
ARE for Gumbel copula (censoring case 1) with p1 = 0.3 and p2 = 0.5
ARE
rho
HR2 = 0.3
HR2 = 0.4
HR2 = 0.5
HR2 = 0.6
HR2 = 0.7
HR2 = 0.8
HR2 = 0.9
HR2 = 0.95

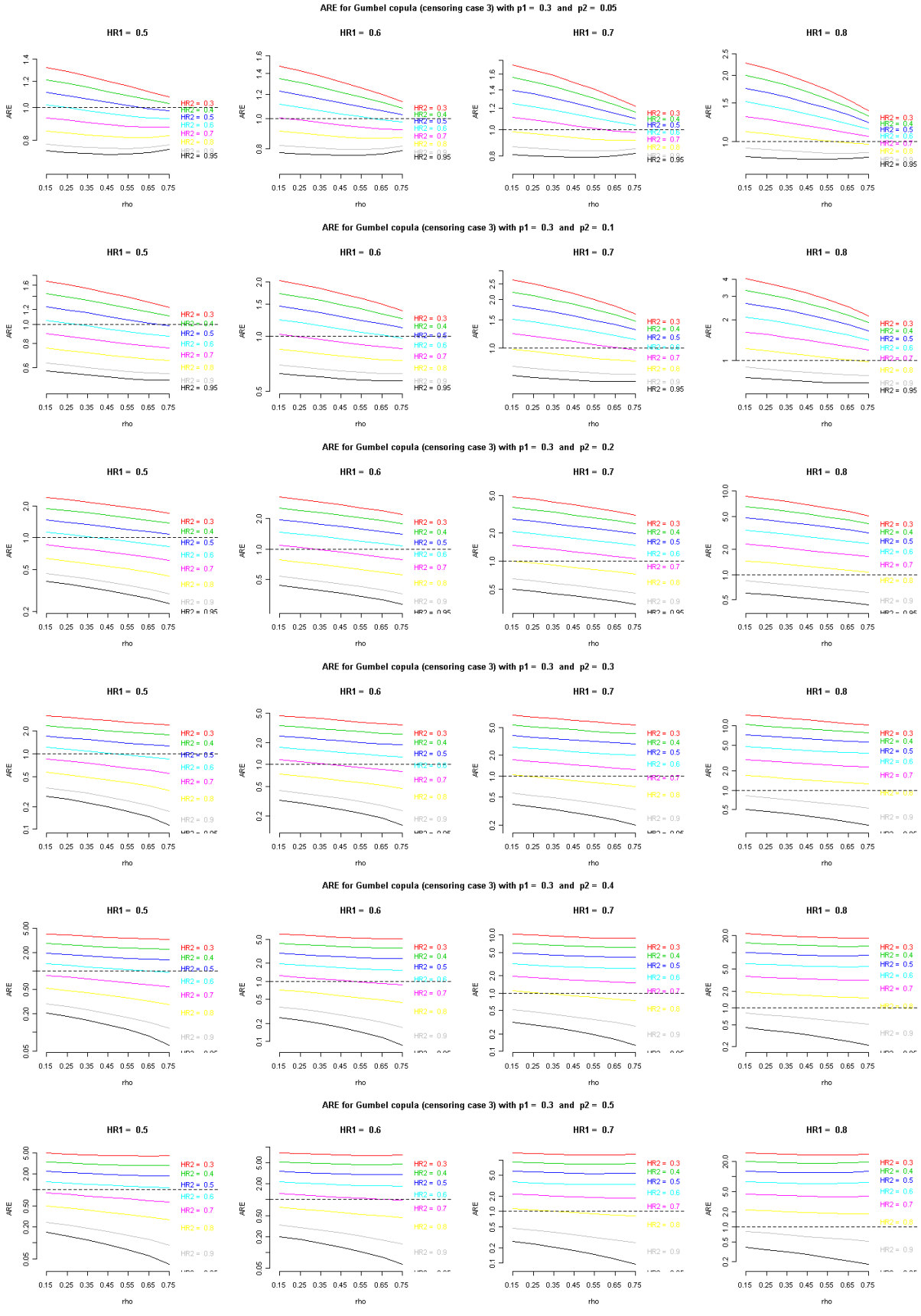

ARE for Gumbel copula (censoring case 1) with p1 = 0.4 and p2 = 0.05
ARE for Gumbel copula (censoring case 1) with p1 = 0.4 and p2 = 0.1
ARE for Gumbel copula (censoring case 1) with p1 = 0.4 and p2 = 0.2
ARE for Gumbel copula (censoring case 1) with p1 = 0.4 and p2 = 0.3
ARE for Gumbel copula (censoring case 1) with p1 = 0.4 and p2 = 0.4
ARE for Gumbel copula (censoring case 1) with p1 = 0.4 and p2 = 0.5
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE
rho
HR2 = 0.3
HR2 = 0.4
HR2 = 0.5
HR2 = 0.6
HR2 = 0.7
HR2 = 0.8
HR2 = 0.9
HR2 = 0.95

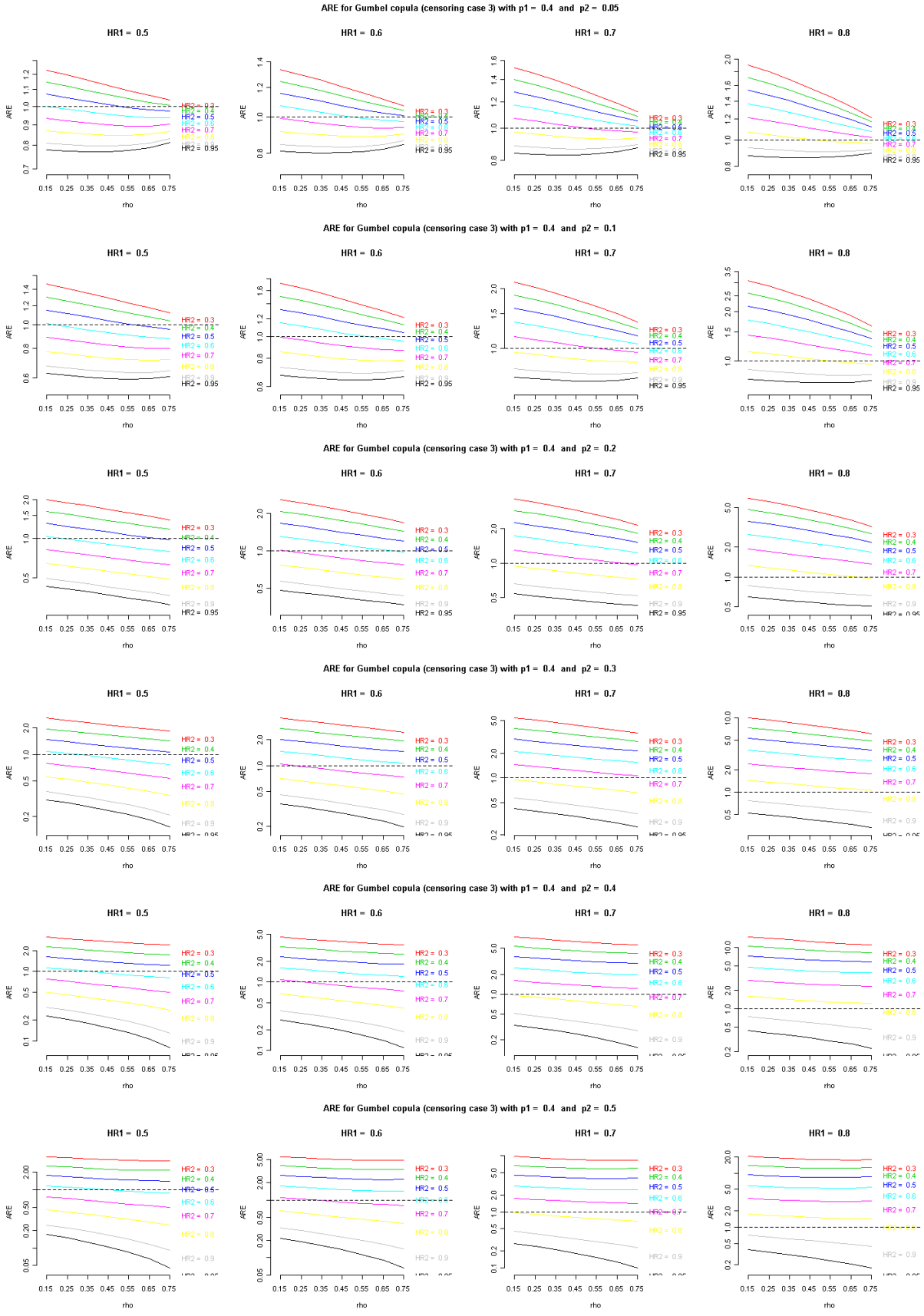

ARE for Gumbel copula (censoring case 1) with p1 = 0.5 and p2 = 0.05
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 1) with p1 = 0.5 and p2 = 0.1
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 1) with p1 = 0.5 and p2 = 0.2
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 1) with p1 = 0.5 and p2 = 0.3
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 1) with p1 = 0.5 and p2 = 0.4
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 1) with p1 = 0.5 and p2 = 0.5
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE
rho
HR2 = 0.3
HR2 = 0.4
HR2 = 0.5
HR2 = 0.6
HR2 = 0.7
HR2 = 0.8
HR2 = 0.9
HR2 = 0.95

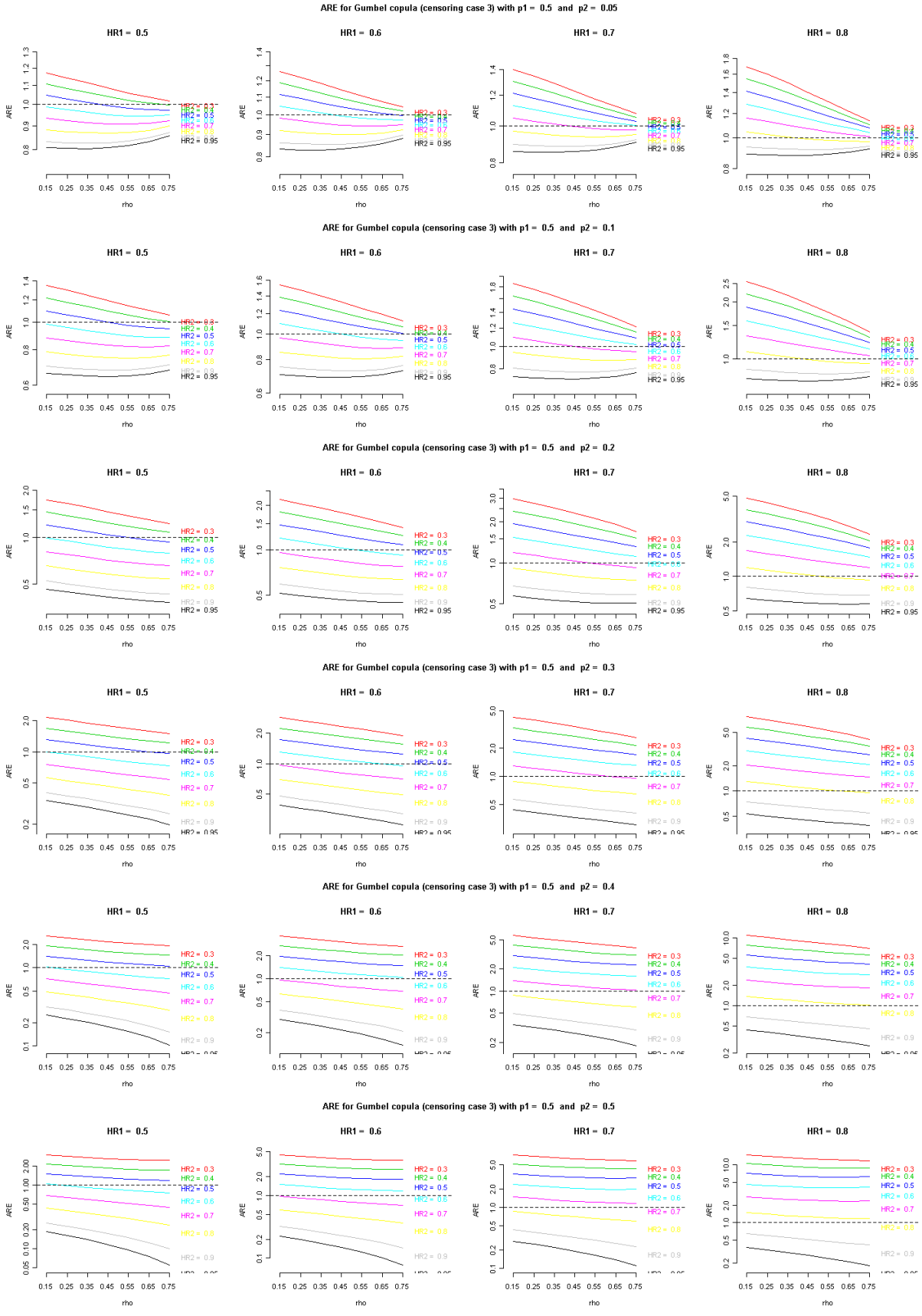

ARE for Gumbel copula (censoring case 3) with p1 = 0.05 and p2 = 0.05
ARE for Gumbel copula (censoring case 3) with p1 = 0.05 and p2 = 0.1
ARE for Gumbel copula (censoring case 3) with p1 = 0.05 and p2 = 0.2
ARE for Gumbel copula (censoring case 3) with p1 = 0.05 and p2 = 0.3
ARE for Gumbel copula (censoring case 3) with p1 = 0.05 and p2 = 0.4
ARE for Gumbel copula (censoring case 3) with p1 = 0.05 and p2 = 0.5
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE
rho
HR2 = 0.3
HR2 = 0.4
HR2 = 0.5
HR2 = 0.6
HR2 = 0.7
HR2 = 0.8
HR2 = 0.9


ARE for Gumbel copula (censoring case 3) with p1 = 0.1 and p2 = 0.05
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 3) with p1 = 0.1 and p2 = 0.1
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 3) with p1 = 0.1 and p2 = 0.2
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 3) with p1 = 0.1 and p2 = 0.3
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 3) with p1 = 0.1 and p2 = 0.4
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 3) with p1 = 0.1 and p2 = 0.5
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE
rho
HR2 = 0.3
HR2 = 0.4
HR2 = 0.5
HR2 = 0.6
HR2 = 0.7
HR2 = 0.8
HR2 = 0.9
HR2 = 0.95


ARE for Gumbel copula (censoring case 3) with p1 = 0.2 and p2 = 0.05
ARE for Gumbel copula (censoring case 3) with p1 = 0.2 and p2 = 0.1
ARE for Gumbel copula (censoring case 3) with p1 = 0.2 and p2 = 0.2
ARE for Gumbel copula (censoring case 3) with p1 = 0.2 and p2 = 0.3
ARE for Gumbel copula (censoring case 3) with p1 = 0.2 and p2 = 0.4
ARE for Gumbel copula (censoring case 3) with p1 = 0.2 and p2 = 0.5
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE
rho
HR2 = 0.3
HR2 = 0.4
HR2 = 0.5
HR2 = 0.6
HR2 = 0.7
HR2 = 0.8
HR2 = 0.9
HR2 = 0.95


ARE for Gumbel copula (censoring case 3) with p1 = 0.3 and p2 = 0.05
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 3) with p1 = 0.3 and p2 = 0.1
ARE for Gumbel copula (censoring case 3) with p1 = 0.3 and p2 = 0.2
ARE for Gumbel copula (censoring case 3) with p1 = 0.3 and p2 = 0.3
ARE for Gumbel copula (censoring case 3) with p1 = 0.3 and p2 = 0.4
ARE for Gumbel copula (censoring case 3) with p1 = 0.3 and p2 = 0.5
ARE
rho
HR2 = 0.3
HR2 = 0.4
HR2 = 0.5
HR2 = 0.6
HR2 = 0.7
HR2 = 0.8
HR2 = 0.9
HR2 = 0.95


ARE for Gumbel copula (censoring case 3) with p1 = 0.4 and p2 = 0.05
ARE for Gumbel copula (censoring case 3) with p1 = 0.4 and p2 = 0.1
ARE for Gumbel copula (censoring case 3) with p1 = 0.4 and p2 = 0.2
ARE for Gumbel copula (censoring case 3) with p1 = 0.4 and p2 = 0.3
ARE for Gumbel copula (censoring case 3) with p1 = 0.4 and p2 = 0.4
ARE for Gumbel copula (censoring case 3) with p1 = 0.4 and p2 = 0.5
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE
rho
HR2 = 0.3
HR2 = 0.4
HR2 = 0.5
HR2 = 0.6
HR2 = 0.7
HR2 = 0.8
HR2 = 0.9
HR2 = 0.95


ARE for Gumbel copula (censoring case 3) with p1 = 0.5 and p2 = 0.05
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 3) with p1 = 0.5 and p2 = 0.1
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 3) with p1 = 0.5 and p2 = 0.2
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 3) with p1 = 0.5 and p2 = 0.3
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 3) with p1 = 0.5 and p2 = 0.4
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE for Gumbel copula (censoring case 3) with p1 = 0.5 and p2 = 0.5
HR1 = 0.5
HR1 = 0.6
HR1 = 0.7
HR1 = 0.8
ARE
rho
HR2 = 0.3
HR2 = 0.4
HR2 = 0.5
HR2 = 0.6
HR2 = 0.7
HR2 = 0.8
HR2 = 0.9
HR2 = 0.95

Appendix C

Discordance between Frank and Gumbel copulas in recommending the use of the composite endpoint

It is interesting to study if discordant cases follow any patron and find which values of β_1, β_2, p_1, p_2, HR_1, HR_2 and ρ yield to different recommendations in whether use or not the composite endpoint.

Composite endpoint using Frank copula and relevant endpoint using Gumbel copula

There are 11 cases in which $ARE_{Frank} > 1$ and $ARE_{Gumbel} \leq 1$ for censoring case 1. In all these cases, $\rho = 0.75$ and $p_1 = p_2 = 0.5$:

- There are 4 cases in which $HR_1 = HR_2 = 0.7$ with $(\beta_1, \beta_2) \in \{(0.5, 1), (1, 0.5), (1, 2), (2, 1)\}$.
- There are 7 cases in which $HR_1 = HR_2 = 0.8$ with $(\beta_1, \beta_2) \in \{(0.5, 1), (1, 0.5), (1, 2), (2, 1)\}$ or $\beta_1 = \beta_2 \in \{0.5, 1, 2\}$.

The number of discordant cases is higher for censoring case 3, with 58 cases in which the ARE is higher than 1 using the Frank copula and lower than 1 using the Gumbel copula. As seen for censoring case 1, most of these cases are when $HR_1 = HR_2$ (35 cases) or $HR_2 = HR_1 + 0.1$ (23 cases). None of the other situations yield to a value of the ARE higher than 1 using Frank copula and lower than 1 using Gumbel copula. When studying these 58 cases depending on p_1 and p_2, it can be observed that despite one case in which $p_1 = p_2 = 0.05$, all other cases correspond to values of p_1 higher than 0.3. On the other hand, the number of discordant cases increases as ρ gets larger, ranging from 1 case for $\rho = 0.15$ to 20 cases for $\rho = 0.75$.

Composite endpoint using Gumbel copula and relevant endpoint using Frank copula

In the situation in which $ARE_{Gumbel} > 1$ and $ARE_{Frank} \leq 1$, there are 1.415 cases for censoring case 1 and 891 cases for censoring case 3. Therefore, it is not useful to list all of them and an analysis

of its parameters is needed. Figure C.1 shows the number of combinations that do not agree depending on HR_1 and HR_2. It can be observed that the cases where there are more discordant situations are when $-0.1 \leq HR_2 - HR_1 \leq 0.1$.

HR1 \ HR2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
0.5	20	112	165	140	25	7	7	0
	0	45	105	114	27	6	0	0
0.6	0	15	101	156	117	30	30	0
	0	0	44	103	96	21	0	0
0.7	0	0	6	76	145	77	77	0
	0	0	0	28	99	69	5	0
0.8	0	0	0	0	44	138	138	0
	0	0	0	0	8	88	34	0

Total number of discordant cases for censoring case 1 = 1.415
Total number of discordant cases for censoring case 3 = 891

Figure C.1: Number of combinations in which $ARE_{Gumbel} > 1$ and $ARE_{Frank} \leq 1$ by HR_1 and HR_2.

If the same analysis is carried out considering the difference $HR_2 - HR_1$ without considering $HR_2 = 0.95$ in which all the cases are concordant, it can be confirmed that the highest number of discordant cases are when HR_1 and HR_2 have similar values (Figure C.2).

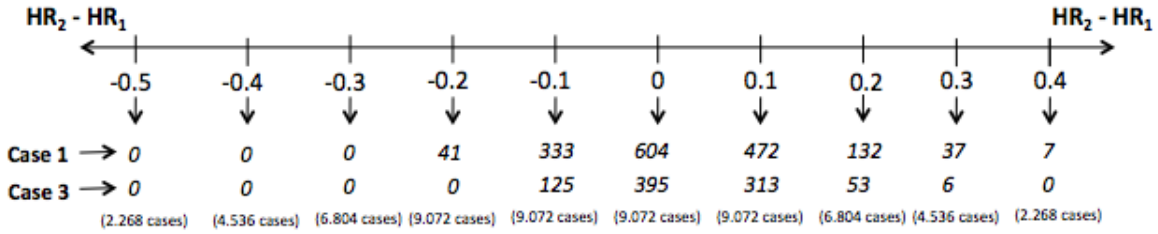


Figure C.2: Number of combinations in which $ARE_{Gumbel} > 1$ and $ARE_{Frank} \leq 1$ by $HR_2 - HR_1$.

When the probabilities of observing the relevant and additional endpoints are considered, it can be observed that the number of discordant cases increases as p_1 gets larger and decreases as p_2 gets larger (Figure C.3).

p1 \ p2	0.05	0.1	0.2	0.3	0.4	0.5
0.05	0	8	5	0	10	5
	3	5	2	0	9	1
0.1	11	26	23	25	18	8
	15	23	17	22	15	3
0.2	56	76	29	37	19	12
	70	48	31	25	14	3
0.3	89	94	89	46	23	17
	71	71	45	20	16	13
0.4	110	97	57	49	38	11
	82	60	36	20	8	2
0.5	113	90	72	35	13	4
	68	51	12	9	1	0

Total number of discordant cases for censoring case 1 = 1.415
Total number of discordant cases for censoring case 3 = 891

Figure C.3: Number of combinations in which $ARE_{Gumbel} > 1$ and $ARE_{Frank} \leq 1$ by p_1 and p_2.

On the other hand, as seen in Figure C.4, the number of discordant cases depending on ρ increases as the correlation gets higher, ranging from 58 and 60 cases for $\rho = 0.15$ to 414 and 194 cases for $\rho = 0.75$, for censoring cases 1 and 3, respectively.

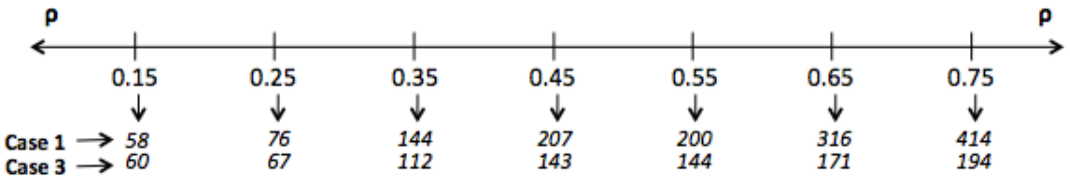


Figure C.4: Number of combinations in which $ARE_{Gumbel} > 1$ and $ARE_{Frank} \leq 1$ by ρ.

It is interesting to remark the behavior of the discordant cases when the shape parameters β_1 and β_2 of the Weibull distribution are taken into account. As seen in Table C.1, there are no big differences in the distribution of the discordant cases but it is interesting to see that the same number of cases are discordant when $(\beta_1, \beta_2) = (0.5, 1)$ and $(\beta_1, \beta_2) = (1, 2)$; when $(\beta_1, \beta_2) = (1, 0.5)$ and $(\beta_1, \beta_2) = (2, 1)$; and when $\beta_1 = \beta_2 = 0.5, 1$ or 2.If these cases are studied, not only the number of cases are the same but the cases themselves are the same. It is also interesting to remark that the number of discordant cases increases as β_1 gets smaller and β_2 gets larger for censoring case 1 but the opposite behavior occurs for censoring case 3.

Censoring case 1	$\beta_2 = 0.5$	$\beta_2 = 1$	$\beta_2 = 2$
$\beta_1 = 0.5$	164	163	176
$\beta_1 = 1$	147	164	163
$\beta_1 = 2$	127	147	164
Censoring case 3	$\beta_2 = 0.5$	$\beta_2 = 1$	$\beta_2 = 2$
$\beta_1 = 0.5$	108	74	66
$\beta_1 = 1$	115	108	74
$\beta_1 = 2$	123	115	108

Table C.1: Number of combinations in which Gumbel and Frank copulas do not agree in recommending the use of the composite endpoint depending on β_1 and β_2.

After these analyses, and as said in chapter 5, it is possible to conclude that both copulas disagree when HR_1 and HR_2 are quite similar and for high values of p_1 jointly with low values of p_2 or strong correlation between the relevant and the additional endpoints.

Appendix D

Computing ARE for Clayton copula

The main aim of this master thesis is to study the robustness of copulas in assessing the efficiency of the main endpoint in a randomized control trial. Gómez and Lagakos [1] used Frank copula and it has been compared to Gumbel copula in this master thesis. In this Appendix, the methodology based on the ARE is developed for Clayton copula in order to compare the results with the previous ones in future research.

As seen in chapter 3, the Clayton Copula conduces to different families of bivariate distribution functions than Frank or Gumbel copulas. The main difference between Clayton, Frank and Gumbel copulas is that the first one exhibits strong left tail dependence while Frank copula exhibit weak tail dependence and the Gumbel copula shows a strong right tail dependence. On the other hand, Frank copula allows both positive or negative dependence while Gumbel and Clayton copulas only allow positive dependence. As stated in chapter 4, this could be a problem but, in clinical trials, it is difficult to find an scenario where the time to the two possible outcomes are correlated negatively.

The Clayton copula is an Archimedean copula and its generator is $\varphi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$. The expression of this copula is

$$C(u, v; \theta) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$$

and the limiting cases of the dependence parameter θ correspond to $C(u, v; 0) = \pi(u, v)$ and $C(u, v; \infty) = M(u, v)$ [3].

The same steps used in chapters 2 and 4 are followed on the computing of ARE with Clayton copula starting from the expression of ARE in censoring cases 1 and 3, defined in (2.6) as

$$\text{ARE}(Z_*, Z) = \left(\frac{\mu_*}{\mu}\right)^2 = \frac{\left(\int_0^1 \log\left(\frac{\lambda_*^{(1)}(t)}{\lambda_*^{(0)}(t)}\right) f_*^{(0)}(t)dt\right)^2}{(\log HR_1)^2 \left(\int_0^1 f_*^{(0)}(t)dt\right)\left(\int_0^1 f_1^{(0)}(t)dt\right)}$$

where $f_1^{(0)}(t)$ and $f_*^{(0)}(t)$ are, respectively, the densities for T_1 and T_* in group 0.

This expression of the ARE depends, among other things, on the law of T_* and it can be obtained, again, from the bivariate distribution of (T_1, T_2). The objective is to obtain the expression of the ARE depending on the same parameters as Frank and Gumbel copulas.

Law of (T_1, T_2)

In this case, T_1 and T_2 are assumed to be binded by Clayton survival copula instead of Frank or Gumbel survival copulas. Assuming equal association parameter θ for groups 0 and 1, the joint survival function for (T_1, T_2) in group j ($j = 0, 1$) is given by

$$S^{(j)}_{(1,2)}(t_1, t_2; \theta) = S^{(j)}_1(t_1) + S^{(j)}_2(t_2) - 1 + [(1 - S^{(j)}_1(t_1))^{-\theta} + (1 - S^{(j)}_2(t_2))^{-\theta} - 1]^{-1/\theta}$$

where $S^{(j)}_1(t_1)$ and $S^{(j)}_2(t_2)$ are the survival functions of T_1 and T_2, respectively, in group j.

Law of T_*

As seen in (2.7), $S^{(j)}_*(t; \theta) = S^{(j)}_{(1,2)}(t, t; \theta)$, and having $f^{(j)}_*(t; \theta) = -\partial S^{(j)}_*(t; \theta)/\partial t$, we have

$$S^{(j)}_*(t; \theta) = S^{(j)}_1(t) + S^{(j)}_2(t) - 1 + [(1 - S^{(j)}_1(t))^{-\theta} + (1 - S^{(j)}_2(t))^{-\theta} - 1]^{-1/\theta}$$

$$f^{(j)}_*(t; \theta) = f^{(j)}_1(t) + f^{(j)}_2(t) - [(1 - S^{(j)}_1(t))^{-\theta} + (1 - S^{(j)}_2(t))^{-\theta} - 1]^{-\frac{1+\theta}{\theta}} \left((1 - S^{(j)}_1(t))^{-(1+\theta)} f^{(j)}_1(t) + (1 - S^{(j)}_2(t))^{-(1+\theta)} f^{(j)}_2(t)\right)$$

$$\lambda^{(j)}_*(t; \theta) = \frac{f^{(j)}_*(t; \theta)}{S^{(j)}_*(t; \theta)}$$

Hence, in order to compute ARE(Z_*, Z) assuming Clayton copula for both groups with equal association parameter θ, we need to specify the same parameters than in the cases assuming Frank or Gumbel copulas.

The relation between those parameters is the same for Clayton, Gumbel and Frank copula, given in chapters 2 and 4. As stated in chapter 3, the dependence parameter θ for Clayton copula cannot be obtained directly from Spearman's ρ as well as for Gumbel Copula. In this case, it is also necessary to use numerical approximations using the R-package `copula` [19, 20, 21].

It is left for future research the simulation cases using Clayton copula and compare its results with the ones obtained from Frank and Gumbel copulas.

Appendix E

R code for interactive graphical display

It has been seen in chapter 3 that two random variables with a fixed correlation yield to a joint distribution function that depends on its dependence structure and, hence, in the copula that links the marginal distribution functions. The R-package `rpanel` [22] is a powerful graphical tool to study how this dependence structure is important. On the other hand, it is also useful in order to study the difference between density functions of T_* for a given copula.

The following scripts have been used in order to study these issues.

```
install.packages("rpanel")
library(rpanel)
install.packages("copula")
library(copula)

#Functions for Gumbel Copula
K<-function(w,theta,v2) {
 return(w*(1-((log(w))/theta))-v2)
}

equacio<-function(gm,K,theta) {
 gm[3]<-uniroot(K, interval=c(0.000000001,0.999999999999), theta=theta, v2=gm[2])$root
}

gumbel_random<-function(n,theta) {
 gm<-cbind(runif(n),runif(n),NA,NA,NA) #gumbel manual
 gm[,3]<-apply(gm,1,equacio,K=K,theta=theta)
 gm[,4]<-exp((gm[,1]^(1/theta))*log(gm[,3]))
 gm[,5]<-exp(((1-gm[,1])^(1/theta))*log(gm[,3]))
 return(gm[,4:5])
}

#rpanel
rp.copula<-function ()
{
    copulaplot.pars <- function(copulaplot) {
       rho2<-as.numeric(copulaplot$rho)
       n2<-copulaplot$n

```

```
30        #Computation of theta
31        copulaplot$theta_frank<-calibSpearmansRho(frankCopula(param=4,dim=2),rho2)
32        copulaplot$theta_clayton<-calibSpearmansRho(claytonCopula(param=4,dim=2),rho2)
33        copulaplot$theta_gumbel<-calibSpearmansRho(gumbelCopula(param=4,dim=2),rho2)
34
35        #Generation of random pairs
36        copulaplot$frank<-rcopula(frankCopula(param=copulaplot$theta_frank,dim=2),n2)
37        copulaplot$clayton<-rcopula(claytonCopula(param=copulaplot$theta_clayton,dim=2),n2)
38        #copulaplot$gumbel<-rcopula(gumbelCopula(param=copulaplot$theta_gumbel,dim=2),n2)
39        #It does not work --> Manual method:
40        copulaplot$gumbel<-gumbel_random(n2,copulaplot$theta_gumbel)
41
42        copulaplot.draw(copulaplot)
43    }
44
45    copulaplot.draw <- function(copulaplot) {
46        copulaplot$n <- max(copulaplot$n, 1)
47        with(copulaplot, {
48         par(mfrow = c(2,2),oma=c(0,0,1,0))
49         if(frank.showing == "TRUE") {
50          plot(frank,main="Frank's copula",xlab="u",ylab="v",xlim=c(0,1),ylim=c(0,1))
51         }
52         else { plot(1, type="n", axes=F, xlab="", ylab="") }
53         if(gumbel.showing == "TRUE") {
54          plot(gumbel,main="Gumbel's copula",xlab="u",ylab="v",xlim=c(0,1),ylim=c(0,1))
55         }
56         else { plot(1, type="n", axes=F, xlab="", ylab="") }
57         if(clayton.showing == "TRUE") {
58          plot(clayton,main="Clayton's copula",xlab="u",ylab="v",xlim=c(0,1),ylim=c(0,1))
59         }
60         else { plot(1, type="n", axes=F, xlab="", ylab="") }
61         if(sobreposar.showing == "TRUE") {
62                   plot(frank,pch=19,cex=0.5,xlab="u",ylab="v",xlim=c(0,1),ylim=c(0,1))
63           points(gumbel,pch=19,cex=0.5,col="red")
64           points(clayton,pch=19,cex=0.5,col="green")
65           legend("topleft",c("Frank","Gumbel","Clayton"),col=c("black","red","green"),pch=19,
                 bg="white")
66         }
67         else { plot(1, type="n", axes=F, xlab="", ylab="") }
68         title(paste("n =",n,"     /     rho =",rho),outer=T)
69
70        })
71        copulaplot
72    }
73
74    #Configure the panel
75    copula.panel <- rp.control("Copula tool", n = 350, rho = 0.8)
76    rp.slider(copula.panel, rho, 0.01, 0.99, initval=0.8, title = "rho", action = copulaplot.
          pars)
77    rp.doublebutton(copula.panel, rho, initval=0.8, range=c(0.01,0.99),showvalue=T,step=0.01,
          action = copulaplot.pars)
78    rp.textentry(copula.panel, n, title = "n", action = copulaplot.pars)
79    rp.checkbox(copula.panel, frank.showing, initval= "TRUE", title = "Frank's Copula",action =
           copulaplot.draw)
80    rp.checkbox(copula.panel, gumbel.showing, initval= "TRUE", title = "Gumbel's Copula",action
           = copulaplot.draw)
81    rp.checkbox(copula.panel, clayton.showing, initval= "FALSE", title = "Clayton's Copula",
          action = copulaplot.draw)
```

```
    rp.checkbox(copula.panel, sobreposar.showing, initval= "FALSE", title = "SOBREPOSAR",
        action = copulaplot.draw)
    rp.do(copula.panel, copulaplot.pars)
}

rp.copula()
```

../R/random_pairs_copulas.r

```
install.packages("copula")
library(copula)

#Draw f*
fstar<-function(rho, beta1, beta2, HR1, HR2, p1, p2, case = 1, copula="Frank",group=1,x)
{
 if(copula=="Frank") {
  theta <- calibSpearmansRho(frankCopula(0),rho)
 }
 if(copula=="Gumbel") {
  theta <- calibSpearmansRho(gumbelCopula(1),rho)
 }
 if(copula=="Clayton") {
  theta <- calibSpearmansRho(claytonCopula(1),rho)
 }

 ####### ASSESSMENT OF THE SCALE PARAMETER VALUES b10, b11, b20, b21
 ## b20 is diferent for case 1 or 3

 # b10 and b11 are the same for case 1 or 3
 b10 <- 1/((-log(1-p1))^(1/(beta1)))
 b11 <- b10/HR1^(1/beta1)

 if(case==1) {
  b20 <- 1/(-log(1-p2))^(1/beta2)
 } else
  if (case==3) {
   ##############################################################################
   # Function: Fb20
   ##############################################################################
   # Description: It computes b20 value for case 3
   # Arguments:
   # b20
   # p2    Probability of observing the additional endpoint
   ##############################################################################

   Fb20<-function(b20,p2) {
    integral<-integrate(function(y) {
     sapply(y,function(y) {
      integrate(function(x)((theta*(1-exp(-theta))*exp(-theta*(x+y)))
      /(exp(-theta)+exp(-theta*(x+y))-exp(-theta*x)-exp(-theta*y))^2),lower=0,
      upper=exp(-((((-log(y))^(1/beta2))*b20)/b10)^beta1))$value
     })
    },
    lower= exp(-(1/b20)^beta2), upper=1)$value
    return(integral-p2)
   }
   limits <- c(0.00001,10000)
   b20 <- uniroot(Fb20, interval=limits,p2=p2)$root
```

```
 }

# b21 is the same for case 1 or 3
b21 <- b20/HR2^(1/beta2)

 fT10 <- (beta1/b10) * ( (x/b10)^(beta1-1) ) * (exp(-(x/b10)^beta1))
 fT11 <- (beta1/b11) * ( (x/b11)^(beta1-1) ) * (exp(-(x/b11)^beta1))
 fT20 <- (beta2/b20) * ( (x/b20)^(beta2-1) ) * (exp(-(x/b20)^beta2))
 fT21 <- (beta2/b21) * ( (x/b21)^(beta2-1) ) * (exp(-(x/b21)^beta2))
 ST10 <- exp(-(x/b10)^beta1)
 ST11 <- exp(-(x/b11)^beta1)
 ST20 <- exp(-(x/b20)^beta2)
 ST21 <- exp(-(x/b21)^beta2)

 if(copula=="Frank") {
  Sstar0 <- (-log(1+(exp(-theta*ST10)-1)*(exp(-theta*ST20)-1)/(exp(-theta)-1))/theta)

  fstar0 <- (exp(-theta*ST10)*(exp(-theta*ST20)-1)*fT10+exp(-theta*ST20)*
  (exp(-theta*ST10)-1)*fT20)/(exp(-theta*Sstar0)*(exp(-theta)-1))

  Sstar1 <- (-log(1+(exp(-theta*ST11)-1)*(exp(-theta*ST21)-1)/(exp(-theta)-1))/theta)

  fstar1 <- (exp(-theta*ST11)*(exp(-theta*ST21)-1)*fT11+exp(-theta*ST21)*
  (exp(-theta*ST11)-1)*fT21)/(exp(-theta*Sstar1)*(exp(-theta)-1))
 }

 if(copula=="Gumbel") {
  Sstar0 <- ST10 + ST20 -1 + exp(-(((-log(1-ST10))^theta+(-log(1-ST20))^theta)^(1/theta)))

  fstar0 <- fT10 + fT20 - (exp(-(((-log(1-ST10))^theta+(-log(1-ST20))^theta)^(1/theta))))*
  (((-log(1-ST10))^theta+(-log(1-ST20))^theta)^((1-theta)/theta))*
  (((-log(1-ST10))^(theta-1))*(fT10/(1-ST10))+((-log(1-ST20))^(theta-1))*(fT20/(1-ST20)))

  Sstar1 <- ST11 + ST21 -1 + exp(-(((-log(1-ST11))^theta+(-log(1-ST21))^theta)^(1/theta)))

  fstar1 <- fT11 + fT21 - (exp(-(((-log(1-ST11))^theta+(-log(1-ST21))^theta)^(1/theta))))*
  (((-log(1-ST11))^theta+(-log(1-ST21))^theta)^((1-theta)/theta))*
  (((-log(1-ST11))^(theta-1))*(fT11/(1-ST11))+((-log(1-ST21))^(theta-1))*(fT21/(1-ST21)))

 }

 if(copula=="Clayton") {
  Sstar0 <- ST10 + ST20 -1 + (((1-ST10)^(-theta))+((1-ST20)^(-theta))-1)^(-1/theta)

  fstar0 <- fT10 + fT20 - ((((1-ST10)^(-theta))+((1-ST20)^(-theta))-1)^(-(1+theta)/theta))*
      (((1-ST10)^(-theta-1))*fT10+((1-ST20)^(-theta-1))*fT20)

  Sstar1 <- ST11 + ST21 -1 + (((1-ST11)^(-theta))+((1-ST21)^(-theta))-1)^(-1/theta)

  fstar1 <- fT11 + fT21 - ((((1-ST11)^(-theta))+((1-ST21)^(-theta))-1)^(-(1+theta)/theta))*
      (((1-ST11)^(-theta-1))*fT11+((1-ST21)^(-theta-1))*fT21)

 }

 if (group==1)  { return(fstar1) }
 if (group==0)  { return(fstar0) }

}
```

```
106
107 #EXAMPLES
108 curve(fstar(rho=0.2,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Frank",group
        =0,x),col=1,from=0,to=10)
109 curve(fstar(rho=0.2,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Frank",group
        =1,x),col=2,add=T)
110
111 curve(fstar(rho=0.2,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Gumbel",group
        =0,x),col=1,add=T,lty=2)
112 curve(fstar(rho=0.2,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Gumbel",group
        =1,x),col=2,add=T,lty=2)
113
114 curve(fstar(rho=0.2,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Clayton",group
        =0,x),col=1,add=T,lty=2)
115 curve(fstar(rho=0.2,beta1=1,beta2=2,HR1=0.9,HR2=0.4,p1=0.3,p2=0.5,case=1,copula="Clayton",group
        =1,x),col=2,add=T,lty=2)
116
117
118 ##############################################################################
119 ##############################################################################
120 # rpanel
121 ##############################################################################
122 ##############################################################################
123
124
125 library(rpanel)
126
127 rp.fstar<-function ()
128 {
129     fstarplot.pars <- function(fstarplot) {
130
131        fstarplot$rho<-as.numeric(fstarplot$rho)
132        fstarplot$beta1<-as.numeric(fstarplot$beta1)
133        fstarplot$beta2<-as.numeric(fstarplot$beta2)
134        fstarplot$HR1<-as.numeric(fstarplot$HR1)
135        fstarplot$HR2<-as.numeric(fstarplot$HR2)
136        fstarplot$p1<-as.numeric(fstarplot$p1)
137        fstarplot$p2<-as.numeric(fstarplot$p2)
138        fstarplot$case<-as.numeric(fstarplot$case)
139
140        fstarplot.draw(fstarplot)
141     }
142     fstarplot.draw <- function(fstarplot) {
143         with(fstarplot, {
144          par(mfrow = c(1,1),oma=c(0,0,2,0))
145
146           curve(fstar(rho,beta1,beta2,HR1,HR2,p1,p2,case,copula="Frank",group=0,x),col=1,lty=1,
                  xlim=c(0,10),ylim=c(0,1),xlab="t",ylab="f*",
147           main=paste("rho =",rho,", beta1 = ",beta1,", beta2 = ",beta2,", HR1 = ",HR1,", HR2 =
                  ",HR2,", p1 = ",p1,", p2 = ",p2),cex.main=0.9)
148           curve(fstar(rho,beta1,beta2,HR1,HR2,p1,p2,case,copula="Frank",group=1,x),col=1,lty=2,
                  add=T)
149           curve(fstar(rho,beta1,beta2,HR1,HR2,p1,p2,case,copula="Gumbel",group=0,x),col=2,lty
                  =1,add=T)
150           curve(fstar(rho,beta1,beta2,HR1,HR2,p1,p2,case,copula="Gumbel",group=1,x),col=2,lty
                  =2,add=T)
151           if(clayton.showing == "TRUE") {
152            curve(fstar(rho,beta1,beta2,HR1,HR2,p1,p2,case,copula="Clayton",group=0,x),col=3,lty
```

```
                =1,add=T)
153             curve(fstar(rho,beta1,beta2,HR1,HR2,p1,p2,case,copula="Clayton",group=1,x),col=3,lty
                =2,add=T)
154             legend("topright",c("Frank group 0","Frank group 1","Gumbel group 0","Gumbel group 1
                ","Clayton group 0","Clayton group 1"),col=c(1,1,2,2,3,3),lty=c(1,2,1,2,1,2))
155           }
156           else {
157             legend("topright",c("Frank group 0","Frank group 1","Gumbel group 0","Gumbel group 1
                "),col=c(1,1,2,2),lty=c(1,2,1,2))
158           }
159           title(paste("Density function for T* for censoring case ",case),outer=T)
160         })
161         fstarplot
162     }
163     fstar.panel <- rp.control("Density functions for T*",size=c(150,570), rho = 0.45, beta1=1,
            beta2=2, HR1=0.8, HR2=0.4, p1=0.3, p2=0.5)
164
165     rp.radiogroup(fstar.panel, rho, c(0.15,0.25,0.35,0.45,0.55,0.65,0.75), title="Rho", action=
            fstarplot.pars, pos=c(1,1,70,175))
166     rp.radiogroup(fstar.panel, beta1, c(0.5,1,2), title="Beta1", action=fstarplot.pars, pos=c
            (75,1,70,85))
167     rp.radiogroup(fstar.panel, beta2, c(0.5,1,2), title="Beta2", action=fstarplot.pars, pos=c
            (75,90,70,85))
168     rp.radiogroup(fstar.panel, HR1, c(0.5,0.6,0.7,0.8), title="HR1", action=fstarplot.pars, pos
            =c(1,180,70,110))
169     rp.radiogroup(fstar.panel, HR2, c(0.3,0.4,0.5,0.6,0.7,0.8,0.9,0.95), title="HR2", action=
            fstarplot.pars, pos=c(75,180,70,195))
170     rp.radiogroup(fstar.panel, case, c(1,3), title="Case", action=fstarplot.pars, pos=c
            (1,295,70,80))
171     rp.radiogroup(fstar.panel, p1, c(0.05,0.1,0.2,0.3,0.4,0.5), title="p1", action=fstarplot.
            pars, pos=c(1,380,70,155))
172     rp.radiogroup(fstar.panel, p2, c(0.05,0.1,0.2,0.3,0.4,0.5), title="p2", action=fstarplot.
            pars, pos=c(75,380,70,155))
173
174      rp.checkbox(fstar.panel, clayton.showing, initval= "FALSE", title = "Clayton",
175         action = fstarplot.draw,pos=c(1,540,145,30))
176
177     rp.do(fstar.panel, fstarplot.pars)
178 }
179
180 rp.fstar()
```

../R/functions_fstar.r