

# COMPOSITE ENDPOINTS IN CLINICAL TRIALS.

## Computational Tools, Practical Guidelines and Methodological Extensions.

**Moisés Gómez Mateu**

Department of Statistics and Operations Research  
Doctoral Thesis, 2016



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH

---

*Thesis director:*

**Guadalupe Gómez Melis, PhD**

Department of Statistics and Operations Research  
Universitat Politècnica de Catalunya, Spain

*Thesis co-director:*

**Urania Dafni, PhD**

Department of Public Health  
University of Athens, Greece

---



*A mis padres, avis, y Laura*



## ACKNOWLEDGEMENTS

A menudo, cuando uno empieza un doctorado se pregunta qué lo empuja a hacerlo. Aún ahora, a punto de depositar esta tesis, no sé del todo el verdadero motivo. Lo que sí tengo claro, sin embargo, es que sin el apoyo de los que me rodean hubiera sido imposible llegar hasta aquí.

En primer lugar, quisiera dar las gracias a la que considero mi mentora, Lupe. Ella es la persona que ha guiado mis pasos en este camino no siempre llano, donde a veces la distancia más corta para llegar a la meta no es la línea recta. Gràcies per ensenyar-me a veure la ciència amb rigor i alhora divertida. De veritat, moltes gràcies per la teva incombustible energia positiva i extraordinàries ganes de transmetre nous coneixements. I sobretot, gràcies per la teva immensa implicació per fer-me créixer tant a nivell professional com personal.

Hay una inmensa cantidad de personas con las que he disfrutado y aprendido durante estos años, y diría que la mayoría tienen como nexos de unión algún grupo de investigación. Grupos como el GRASS y GRBIO me han permitido conocer a personas increíbles. Gracias Klaus, por tu excelencia y honradez, y por estar siempre dispuesto a echar una mano. A Nuria Pérez, quien considera acertadamente haber ganado un Oscar con su tesis! Y a su hermana Susana, sois dos grandes ejemplos para mí. A Nuria Porta, quien me animó a empezar esta aventura. A Víctor Urrea, siempre dispuesto a ayudarme con R; y a Xavier de Pedro y Alex Sánchez a quien les agradezco su gran ayuda en los comienzos con la plataforma Tiki. A Carles Serrat, Montse Rué, Olga Julià y Malu Calle, por vuestra increíble capacidad de transmitir excelencia. A los compañeros de departamento y sus cafés en el bar: Erik Cobo, Toni Font, Sonia Navarro, Jordi Cortés, Marta Bofill, Hajar, entre otros. También a Celia y Carme por hacer más ágil el engorroso papeleo.

A la vez que he aprendido, me lo he pasado en grande en los congresos y jornadas. He conocido también a gente maravillosa. Altea, Urko, Martí, Natalia, Irantzu y un gran largo etc. Sois todos geniales! Entre muchos recuerdos imborrables, nunca olvidaré el congreso en Turquía, en especial el viaje en globo por la Capadocia, jejeje.

También a Carmen Cadarso, por impulsar la red Biostatnet; y a las sociedades SCE y la SEB, todas ellas en especial por apoyar a los jóvenes investigadores. También quisiera dar las gracias a los proyectos cuya financiación ha permitido realizar esta tesis: MTM2012-38067-C02-01, MTM2015-64465-C2-1-R del Ministerio de Economía y Competitividad; y 2014 SGR 464 del Departament d'Economia i Coneixement de la Generalitat de Catalunya.

I would also like to express my gratitude to the Society for Clinical Trials for the *2015 Thomas Chalmers Student Scholarship for Research Excellence* in Virginia. It was one of the most gratifying moments since I began my PhD. Thanks also to KyungMann Kim; it is an honor to collaborate in the current and future work.

My sincere gratitude to my co-director Urania Dafni. I have learnt a lot from her, and from her expertise. I have had a very good time in Barcelona, Galicia and Turkey, and I hope we can also meet in Greece someday! Thank you Ranny!

Gràcies avi, tu vas ser el meu gran exemple d'honradesa, humilitat i saber estar, i crec que en el fons, la llavor que m'impulsa a continuar aprenent des que era ben petit. Gràcies iaia, per animar-me a estudiar. La teva llarga experiència i històries de la guerra m'ha permès veure la vida des de diferents prismes. D'alguna manera els dos sou la llum que em guia.

Gràcies també a tu, mama. No podré agrair mai suficientment l'esforç incondicional per ajudar-me a ser millor persona, que no és fàcil. Y a ti, papa, por ayudarme a veces a tomarme los problemas con más humor, y a ver el mundo menos gris en momentos duros.

Le dedico también esta tesis a mi hermana Eva. Siempre, siempre, ha estado ahí para apoyarme en todo y darme consejo, y para animarme en todo momento. Muchísimas gracias. Gracias también a mis hermanas Lali y Mari. Gracias a las tres por ser como sois. Y a mi sobrino Adrià, gran cocinero donde los haya; y a mis sobrinas, Andrea, Carla, Ana y Ainhoa que sois unos soles para mí. También a mis suegros Antonio y Mari; y a todos mis cuñad@s, en especial a Lito, que lo considero un hermano mayor. Y también a Héctor, amigo de la infancia y con el cual hemos compartido y seguimos compartiendo infinitas birras.

Finalmente, esta tesis no hubiera sido posible sin el apoyo incondicional de Laura. Ella es la que siempre me ha acompañado en los momentos más dulces pero sobretodo en los más duros. No sólo en estos últimos cinco años, sino desde que está conmigo. A ella, a ti cariño, te dedico especialmente esta tesis. Contigo, mi vida tiene sentido.

## ABSTRACT

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

Composite endpoints (CE) consist of the union of two or more outcomes and are often used in RCT. When the focus is time-to-event analysis, CE refer to the elapse time from randomization until the first component of the CE. In oncology trials, for instance, progression-free survival is defined as the time to disease progression or death.

The decision on whether to use a CE versus a single component as the PE is controversial. The advantages and drawbacks regarding the use of CE have been extensively discussed in the literature. Gómez and Lagakos develop a statistical methodology to evaluate the convenience of using a relevant endpoint RE versus a CE consisting of the union of the RE plus another additional endpoint (AE). Their strategy is based on the value of the asymptotic relative efficiency (ARE), which relates the efficiency of using the logrank test based on the RE versus the efficiency based on the CE. The ARE is expressed as a function of the marginal laws of the time to each component RE and AE, the probabilities of observing each component in the control group, the hazard ratios measured by each component of the CE between the two treatment groups, and the correlation between components.

This thesis explores, elaborates on, implements and applies the ARE method. We have also developed a new online platform named *CompARE* that facilitates the practical use of this method. The ARE method has been applied to cardiovascular studies. We have made further progress into the theoretical meaning of the ARE and have explored how to handle the probability and the hazard ratio of a combination of endpoints.

In cardiovascular trials, it is common to use CE. We systematically examine the use of CE in this field by means of a literature search and the discussion of several case studies. Based on the ARE methodology, we provide guidelines for the informed choice of the PE.

We prove that the usual interpretation of the ARE as the ratio of sample sizes holds and that it can be applied to evaluate the efficiency of the RE versus the CE. Furthermore, we carry out a simulation study to empirically check the proximity between the ratio of finite sample sizes and the ARE.

We discuss how to derive the probabilities and hazard ratios when they come from a combination of several components. Furthermore, it is shown that the combined hazard ratio (HR\*) is, in general,

not constant over time, even if the hazard ratio of the marginal components are. This non-constant behaviour might have a strong influence on the interpretation of treatment effect and on sample size assessment. We evaluate the behaviour of the  $HR^*$  in respect to the marginal parameters, and we study its departure from constancy, depending on different scenarios.

This thesis has implemented the ARE methodology on the online platform *CompARE*. Clinicians and biostatisticians can use *CompARE* to study the performance of different endpoints in a variety of scenarios. *CompARE* has an intuitive interface and it is a convenient tool for better informed decisions regarding the PE. Results from different parameter settings are shown immediately by means of tables and plots. *CompARE* is extended to quantify specific values for the combined probability and hazard ratios. When the user cannot anticipate some of the needed parameters, *CompARE* provides a range of plausible values. Moreover, the departure from constancy of a combined hazard ratio can be explored by visualizing its shape over time. Sample size computations are implemented as well.



## RESUMEN

Los eventos compuestos consisten en la unión de dos o más eventos, y son utilizados usualmente en ensayos clínicos aleatorizados. A menudo, los análisis se basan en el tiempo hasta que se produce el evento de interés; en ese caso hablaríamos del tiempo hasta el primero de los componentes. En ensayos oncológicos, por ejemplo, la supervivencia libre de progresión se define como el tiempo hasta la progresión o la muerte.

La decisión entre utilizar un evento compuesto o un componente de este como variable principal es controvertida. Gómez y Lagakos desarrollan una metodología estadística para evaluar la conveniencia de utilizar un evento relevante frente a un evento compuesto consistente en la unión del evento relevante más un evento adicional. Su estrategia se basa en el valor de la eficiencia relativa asintótica (ARE, usando el acrónimo en inglés), la cual relaciona la eficiencia de utilizar el test logrank basado en el evento relevante frente a la eficiencia basada en el evento compuesto. La ARE se expresa en función de las leyes marginales correspondientes al tiempo hasta cada componente relevante y adicional, las probabilidades de observar cada componente en el grupo control, los hazard ratios medidos para cada componente del evento compuesto entre los dos grupos de tratamiento y la correlación entre los componentes.

Esta tesis explora, profundiza, implementa y aplica la metodología ARE. También hemos creado una nueva plataforma en línea, *CompARE*, que facilita el uso práctico de esta metodología.

Examinamos sistemáticamente el uso de eventos compuestos en ensayos cardiovasculares a partir de una búsqueda en la literatura existente y discutimos diferentes casos. Basándonos en la metodología ARE, aportamos guías para la elección informada de la variable principal.

Probamos que la interpretación usual de la ARE como el ratio de los tamaños de muestra se sustenta y puede ser aplicado para evaluar la eficiencia del evento relevante frente al evento compuesto. Asimismo, llevamos a cabo una simulación para estudiar empíricamente cuán cerca está el ratio de tamaños de muestra finitos de la ARE.

Discutimos cómo derivar las probabilidades y hazard ratios cuando provienen de una combinación de varios componentes. También mostramos que el hazard ratio combinado es, en general, no constante a lo largo del tiempo, incluso cuando los hazard ratios de los componentes marginales lo son. Este comportamiento no constante puede tener una gran influencia en la interpretación del efecto del tratamiento y en el cálculo de los tamaños de muestra. Evaluamos el comportamiento del

hazard ratio combinado respecto a los parámetros marginales y lo estudiamos para diferentes escenarios.

En esta tesis se ha implementado la metodología ARE en la plataforma en línea *CompARE*. Clínicos y bioestadísticos pueden utilizar *CompARE* para estudiar el comportamiento de diferentes eventos en un gran abanico de escenarios. *CompARE* contiene una interfaz intuitiva y es una herramienta conveniente para tomar una mejor decisión informada sobre la variable principal. Los resultados provenientes de diferentes escenarios son mostrados instantáneamente a partir de tablas y gráficos. *CompARE* se ha ampliado para cuantificar valores específicos para la probabilidad combinada y el hazard ratio. Cuando el usuario no puede anticipar alguno de los parámetros necesarios, *CompARE* facilita un rango de valores posibles. Asimismo, el hazard ratio puede ser explorado visualizando su forma a lo largo del tiempo y, por lo tanto, proporciona una ayuda gráfica para posibles desviaciones de proporcionalidad de los hazards. Cálculos sobre el tamaño de muestra también han sido implementados en la plataforma.

## RESUM

Els esdeveniments compostos consisteixen en la unió de dos o més esdeveniments, i són utilitzats usualment en assajos clínics aleatoritzats. Sovint, les anàlisis es basen en el temps fins que es produeix l'esdeveniment d'interès; en aquest cas parlariem del temps fins al primer dels components. En assajos oncològics, per exemple, la supervivència lliure de progressió es defineix com a temps fins a la progressió o la mort.

La decisió entre utilitzar un esdeveniment compost o un component d'aquest com a variable principal és controvertida. Gómez i Lagakos desenvolupen una metodologia estadística per avaluar la conveniència d'utilitzar un esdeveniment rellevant enfront d'un esdeveniment compost consistent en la unió de l'esdeveniment rellevant més un esdeveniment addicional. La seva estratègia es basa en el valor de l'eficiència relativa asimptòtica (ARE, fent servir l'acrònim en anglès), la qual relaciona l'eficiència d'utilitzar la prova logrank basada en l'esdeveniment rellevant enfront de l'eficiència basada en l'esdeveniment compost. L'ARE s'expressa com a funció de les lleis marginals corresponents al temps fins a cada component rellevant i addicional, les probabilitats d'observar cada component en el grup control, els hazard ratios mesurats per a cada component de l'esdeveniment compost entre els dos grups de tractament i la correlació entre els components.

Aquesta tesi explora, aprofundeix, implementa i aplica la metodologia ARE. També hem creat una nova plataforma en línia, *CompARE*, que facilita l'ús pràctic d'aquesta metodologia.

Examinem sistemàticament l'ús d'esdeveniments compostos en assajos cardiovasculars a partir d'una recerca en la literatura existent i en discutim diferents casos. Basant-nos en la metodologia ARE, aportem guies per a l'elecció informada de la variable principal.

Provem que la interpretació usual de l'ARE com la ràtio de les mides mostrals se sustenta i pot ser aplicada per avaluar l'eficiència de l'esdeveniment rellevant enfront de l'esdeveniment compost. A més, portem a terme una simulació per estudiar empíricament com n'està, de prop, la ràtio de mides mostrals finites respecte de l'ARE.

Discutim com es poden derivar les probabilitats i els hazard ratios quan provenen d'una combinació de diversos components. També mostrem que el hazard ratio combinat és, en general, no constant al llarg del temps, fins i tot quan els hazard ratios dels components marginals ho són. Aquest comportament no constant pot tenir una gran influència en la interpretació de l'efecte del tractament i en el càlcul de les mides mostrals. Avaluem el comportament del hazard ratio combinat respecte dels paràmetres marginals i l'estudiem per a diferents escenaris.

En aquesta tesi també s'ha implementat la metodologia ARE en la plataforma en línia *CompARE*. Clínic i bioestadístics poden utilitzar *CompARE* per estudiar el comportament de diferents esdeveniments en un gran ventall d'escenaris. *CompARE* conté una interfície intuïtiva i és una eina convenient per prendre una decisió informada millor sobre la variable principal. Els resultats provinents de diferents escenaris són mostrats instantàniament a partir de taules i gràfics. *CompARE* s'ha ampliat per quantificar valors específics per a la probabilitat combinada i el hazard ratio. Quan l'usuari no pot anticipar algun dels paràmetres necessaris, *CompARE* facilita un rang de valors possibles. A més, el hazard ratio pot ser explorat visualitzant-ne la forma al llarg del temps i, per tant, proporciona una ajuda gràfica per a possibles desviacions de la proporcionalitat dels hazards. Càlculs sobre la mida mostral també han estat implementats en la plataforma.

# CONTENTS

<b>1</b>	<b>Introduction. State of the art and main objectives</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Main arguments in favor and cautions in the use of composite endpoints in clinical trials	2
1.3	Testing the statistical treatment effect of the Primary Endpoint . . . . .	5
1.3.1	The logrank test . . . . .	5
1.3.2	Testing treatment effect using the relevant endpoint and the composite endpoint	6
1.4	The Asymptotic Relative Efficiency (ARE) of two tests . . . . .	7
1.5	The ARE method by Gómez and Lagakos . . . . .	8
1.5.1	Assumptions . . . . .	9
1.5.2	The ARE expression as a function of anticipatable parameter values . . . . .	14
1.6	Goals and thesis structure . . . . .	15
<b>2</b>	<b>The composite endpoint in the cardiovascular area</b>	<b>17</b>
2.1	Background . . . . .	19
2.1.1	Composite endpoints in cardiovascular research . . . . .	19
2.2	Survey of use of composite endpoints in the cardiovascular literature . . . . .	19
2.2.1	Identification of published clinical trials that used composite endpoints . . . . .	19
2.2.2	Information abstracted from each RCT . . . . .	21
2.2.3	Method to set the recommendations . . . . .	21
2.3	Case studies . . . . .	21
2.3.1	Case study 1: treating patients after an acute coronary syndrome with Succinobucol . . . . .	21
2.3.2	Case study 2: treating hemorrhagic complications during primary percutaneous coronary intervention in acute MI . . . . .	24
2.3.3	Case Study 3: testing Fondaparinux in patients with ST-segment–elevation MI . . . . .	25
2.3.4	Case study 4: prevention studies . . . . .	25
2.4	Results and recommendations . . . . .	26
2.4.1	Death from cardiovascular or death from any cause as the individual primary or co-PE . . . . .	27
2.4.2	Recommendations for cardiovascular clinical trials . . . . .	28

2.5	Discussion . . . . .	31
<b>3</b>	<b>Relationship between the ARE and sample sizes</b>	<b>33</b>
3.1	Notation, the logrank test and the asymptotic relative efficiency . . . . .	35
3.1.1	The logrank tests for the relevant and for the composite endpoints . . . . .	35
3.1.2	Asymptotic Relative Efficiency . . . . .	36
3.2	Relationship between ARE and sample sizes . . . . .	38
3.3	Simulation . . . . .	41
3.3.1	Simulation . . . . .	41
3.3.2	Results . . . . .	43
3.4	Discussion . . . . .	45
<b>4</b>	<b>CompARE. An on-line platform as a decision tool for investigators</b>	<b>47</b>
4.1	Software used to build <i>CompARE</i> . . . . .	49
4.2	<i>CompARE</i> step by step . . . . .	50
4.3	Summary . . . . .	57
<b>5</b>	<b>Practical issues to assign combined probabilities and hazard ratios</b>	<b>59</b>
5.1	Combined probability . . . . .	61
5.1.1	Boundaries for the combined probability . . . . .	61
5.1.2	Expression of the combined probability $p_*$ as a function of the correlation coefficient . . . . .	62
5.1.3	Expression of the combined probability $p_*$ as a function of several parameters. Time-to-event data . . . . .	64
5.2	Combined hazard ratio . . . . .	68
5.2.1	Notation and assumptions . . . . .	68
5.2.2	Expression of $HR_*(t)$ as a function of several parameters . . . . .	69
5.2.3	Scenarios with constant $HR_*(t)$ . . . . .	70
5.3	Practical solutions to assign combined hazard ratios . . . . .	74
5.3.1	General behavior of the $HR_*(t)$ over time . . . . .	74
5.4	Example from the cardiovascular area . . . . .	81
5.5	Conclusion . . . . .	83
<b>6</b>	<b>Applied extensions in CompARE</b>	<b>85</b>
6.1	Extended options to change parameter values . . . . .	87
6.2	Computation of HR and probability for combined endpoints . . . . .	90
6.3	Sample size . . . . .	92
6.4	Computations when both the relevant and additional endpoints include terminating events . . . . .	95

---

<b>7</b>	<b>Future research</b>	<b>97</b>
7.1	Extension of the ARE method . . . . .	97
7.2	Non-proportional hazards and alternative measures . . . . .	98
7.3	Composite endpoints in other medical fields . . . . .	100
7.4	Future extensions in CompARE . . . . .	101
<b>8</b>	<b>References</b>	<b>103</b>
<b>A</b>	<b>Appendix: List of references from literature search in Chapter 2</b>	<b>111</b>
<b>B</b>	<b>Appendix for Chapter 5</b>	<b>117</b>
<b>C</b>	<b>Appendix: R code to compute the Asymptotic Relative Efficiency (ARE) values</b>	<b>121</b>
<b>D</b>	<b>Appendix: Publications related to this thesis</b>	<b>133</b>



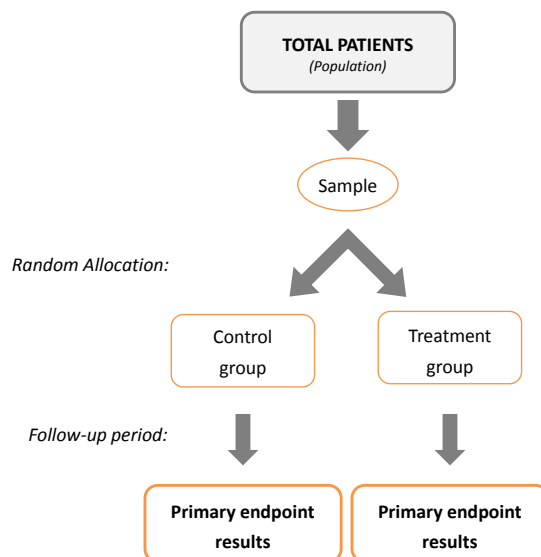


# INTRODUCTION. STATE OF THE ART AND MAIN OBJECTIVES

## 1.1 Background

The broad aim of clinical drug development is to find out whether a treatment can be shown to be simultaneously safe and effective, to the extent that the risk-benefit relationship is acceptable. At phase III of treatment development, randomized clinical trials (RCTs) are commonly designed to demonstrate or confirm any therapeutic benefit (ICH, E8, E9).<sup>36</sup> The conclusions from RCTs rely on the primary endpoint, which is chosen at the design stage of the study, and thus it is of utmost importance to select it appropriately. In these trials, there should generally be only one primary endpoint, and it should be able to provide the most clinically relevant and convincing evidence directly related to the primary objective of the trial. In two-arm clinical trials, primary endpoint observations from control and treatment group are compared in order to provide scientific evidence regarding the potential efficacy of the new treatment after a follow-up period (see Figure 1.1).

Composite endpoints (CEs), also referred as combined outcomes, consist of the union of two or more outcomes and are often used as the primary endpoint in RCTs. Many trials measure dichotomous (binary) endpoints and combine them into a single composite outcome which is considered to have occurred if any of the individual outcomes is observed (Meinert, 2012; Neaton, 2005).<sup>44,46</sup> For example, in the HIV field, the binary composite endpoint consisting of the observation of CD4 cell count  $< 250$  cells/ $\mu$ L or need for antiretroviral therapy (ART) is used as primary endpoint to prove treatment efficacy (Reynolds, 2012).<sup>58</sup> Total scores and health indices, based on rating scales, are also referred to as composite endpoints (Chi, 2005).<sup>6</sup> For instance, the Hamilton Depression Rating Scale (HAMD) is based on multiple items and is a composite endpoint commonly used in depression trials.



*Figure 1.1: Scheme of a randomized confirmatory clinical trial.*

When the focus is time-to-event analysis, composite endpoints use the time from randomization until the time that the first CE component occurs. For example, progression-free survival is defined as a time to disease progression or death in cancer studies. In cardiovascular trials, it is common to use CEs as primary endpoints. They usually incorporate either terminal outcomes such as death from any cause or cardiovascular death, and non-terminal outcomes such as myocardial infarction, stroke and hospitalization. Time to the composite endpoint Major Adverse Cardiovascular Event (MACE), for example, is generally defined as the time to cardiovascular death, myocardial infarction, target vessel revascularization or stroke, whichever occurs first.

## 1.2 Main arguments in favor and cautions in the use of composite endpoints in clinical trials

The decision on whether to use a composite endpoint versus a single component as the primary endpoint in RCTs is controversial. The pros and cons regarding the use of CE have been extensively discussed in the literature, as described in depth by the European Network for Health Technology Assessment (EUnetHTA, 2013)<sup>13</sup> guidelines. We summarize the main arguments discussed by several authors (see Table 1.1).

### Main arguments in favor

Although a simple endpoint has the advantage of simplicity, one of the major arguments for using CEs is the need to combine multiple measurements into a single CE when a single primary variable cannot be selected from several outcomes that are associated with the primary objective, as stated in the guidelines for the International Conference on Harmonisation of Technical Requirements for

### Use of composite endpoints

#### Main arguments in favor

Combination of several outcomes of equal importance in one primary endpoint  
 Net clinical treatment benefit estimation  
 Multiplicity is adequately addressed without type I error adjustments  
 Avoids the problem of bias associated to competing risks  
 Event rate incrementation

#### Cautions

Need for homogeneity between components (similar clinical importance and expected effects)  
 Improvement can be driven by less important components  
 Interpretation problems regarding the the global effect of the CE  
 Multiplicity adjustments when reporting component results  
 Significance of CE does not imply significance of components  
 Including inappropriate components may lead to a loss of power in the CE

**Table 1.1**

Registration of Pharmaceutical for Human Use (ICH).<sup>36</sup> Composite endpoints may help investigators who are having difficulty in deciding which outcome to select as primary endpoint, especially when they are of equal importance (Ferreira-González, 2007; Freemantle, 2003; Ross, 2007)<sup>15, 19, 59</sup> or when no single event alone could be considered an adequate primary endpoint. Composite endpoints can also be useful when an outcome is considered a poor surrogate for the principal outcome of the study. As stated in Amir 2012,<sup>2</sup> progression free survival (PFS), a CE consisting of death and progression, might sometimes be a valid surrogate for overall survival (OS), due to the high correlation between them. Hence, PFS would be an alternative to testing the overall clinical effect of treatment.

The problem of multiple comparisons (multiplicity) is also addressed. Since only one outcome is used as the primary endpoint for statistically testing the treatment effect, there is no need to adjust for the type I error (Kleist, 2006; Wittkop, 2010).<sup>39, 80</sup>

Another major argument is to avoid the problem of competing risks, especially when the target of a treatment is to reduce non-fatal events for diseases in which a fatal event might preclude their observation (Cook, 2007). Thus, endpoints which do not include the mortality component are problematic, since patients who died before the endpoint of interest are likely to not have the same risk as the survivors.

One more rational argument is that the use of CE assures higher event rates. The improvement of treatments and population health in the society fortunately leads to declining event rates. Consequently, a larger number of patients or longer follow-up period is needed in order to detect treatment differences. Using CEs for a fixed sample size and follow-up period, the number of outcome observations (event rates) increases, and this might hopefully lead to an increase in the statistical efficiency of the trial.

#### Cautions

Despite the rationale of using a composite endpoint as described above, several authors have cautioned against their use and interpretation since they do not always prove useful and informative for clinical decision making (Montori, 2005).<sup>45</sup> It is well established that a CE in an RCT should only be

used if the individual components are clinically meaningful and of similar importance to the patient (Tomlinson, 2010);<sup>72</sup> the expected effects on each component are similar based on biological plausibility; and the more important clinical components should -at the very least- not be affected negatively (Kleist, 2007).<sup>38</sup>

Composite endpoints can be used to describe an overall disease process (van Leth, 2003)<sup>76</sup> or a complex disease. However, the interpretation of the global effect of a CE might be confounding when components are quite heterogeneous and do not move along the same lines as each other (Freemantle, 2003).<sup>19</sup> Moreover, the treatment benefit with a CE in which the component endpoints have very different clinical importance can be problematic, because the treatment might have beneficial effects for only the less important endpoints and thus give a misleading impression (Neaton, 2005).<sup>46</sup> As stated in Ferreira-González (2007),<sup>14</sup> the main results of their systematic review show that less important components had higher event rates and larger treatment effects, whereas the most important components (mainly mortality outcomes) provided the lowest event rate and showed the smallest treatment effects.

As several authors propose, the results of each CE component should be reported individually. They should be presented as secondary outcomes, and report the primary CE results that clearly follow the CONSORT guidelines.<sup>8</sup> Some methodologies take into consideration the importance of the CE components, such as those based on priority ranking outcomes (Rauch, 2014).<sup>54</sup> These methods consider endpoints of different scale levels that assign higher weight to the most clinically relevant endpoints. In Pocock (2012),<sup>52</sup> the authors propose using the "win ratio", a measure that takes into account the clinical importance of each component. However, this method does not actually use the precise times from randomization to event occurrence, but only considers whether the outcome is observed sooner in one patient than in another patient with a similar profile.

It has also been discussed that statistical treatment significance on the CE does not imply statistical treatment significance for each component. One way to overcome this problem is to analyze all the composite components separately. However, this would lead to a problem of multiplicity. Consistency-adjusted methods for type I error are proposed to test the efficacy of both the CE and the main component (Rauch, 2014)<sup>57</sup> although it comes with a loss in power. Another strategy is to test the superiority of the CE and the non-inferiority of each one of the components in order to guarantee the overall clinical relevance of the result (Rauch, 2013).<sup>55</sup>

Another rationale in literature for using CE instead of a single event as primary endpoint is the reduction in sample size due to the increase of the event rates (Tomlinson, 2010).<sup>72</sup> Hence, investigators hoping to increase the efficiency of the test can be tempted to add endpoints that will give them a higher number of observed events, such as hospitalizations. In the case of time-to-event endpoints, this increase is expected to be achieved by including component endpoints that occur with higher frequency and/or earlier than the main events of interest (Freemantle, 2003).<sup>19</sup> However, it has been discussed (Montori, 2005)<sup>45</sup> and demonstrated in Gómez and Lagakos (2013)<sup>27</sup> that adding inappropriate components to the relevant endpoint might actually lead to a loss of power in detecting the true treatment differences, consequently leading to a larger chance of failure in detecting any real effect of

the treatment under study.

Another crucial aspect is referred to the joint probability law of the CE. It is the basis for computing the needed sample size in order to detect a prespecified treatment effect size for a specified nominal  $\alpha$ -level for a given power. However, due to the complexity of the composite endpoint definition, a straightforward derivation of the associated joint distribution cannot be based solely on the marginal laws of their multiple single endpoints. The final law for the composite endpoint and, hence, the efficient evaluation of the treatment effect are influenced by: the distribution of every component of the composite endpoint, the probability of observing the outcome in the control group, the expected treatment effects given by the hazard ratio between the two treatment groups, the correlation between the components and, to a lesser extent, the joint distribution between the components.

In the following sections we describe how to assess the statistical treatment effect in clinical trials that involve time-to-event analysis, and the statistical methodology developed by Gómez and Lagakos (2013) for deciding when it would be more appropriate to use a CE instead of a single endpoint.

## 1.3 Testing the statistical treatment effect of the Primary Endpoint

### 1.3.1 The logrank test

The logrank test, also known as the Mantel-Cox test, is commonly used in time-to-event analysis for comparing treatment effects between groups. Assume that we have a two-arm study that involves either a random assignment to an active ( $X = 1$ ) or a control treatment ( $X = 0$ ) that aims to prove the efficacy of the new active treatment. The effect of treatment is evaluated based on the time  $T^{(j)}$  to an endpoint  $\mathcal{E}$ , where the superscript  $j$  indicates the treatment group ( $j = 0$  for the control group and  $j = 1$  for the treatment group). Let  $\lambda^{(j)}(t)$  and  $S^{(j)}(t)$  denote the hazard and survival functions of  $T^{(j)}$  ( $j = 0, 1$ ), respectively. The null hypothesis of no treatment effect is given by  $H_0: S^{(0)}(t) = S^{(1)}(t)$ , or equivalently by  $H_0: HR(t) = \lambda^{(1)}(t)/\lambda^{(0)}(t) = 1$ , where  $HR(t)$  is the hazard ratio. The alternative hypothesis that the new treatment improves survival is given by  $H_a: HR(t) < 1$ .

Consider the observed failure times  $t_1 < t_2 < \dots < t_l$ . At each time  $t_i$  ( $i = 1, 2, \dots, l$ ) we observe  $d_i^{(0)}$  events and  $R_i^{(0)}$  individuals at risk in group 0 (control), and  $d_i^{(1)}$  events and  $R_i^{(1)}$  individuals at risk in group 1 (treatment).

Group	Observed to fail at $t_i$ (events)	Not fail at $t_i$	At risk
Control	$d_i^{(0)}$	$R_i^{(0)} - d_i^{(0)}$	$R_i^{(0)}$
Treatment	$d_i^{(1)}$	$R_i^{(1)} - d_i^{(1)}$	$R_i^{(1)}$
Total	$d_i$	$R_i - d_i$	$R_i$

**Table 1.2**

Note that conditioned on the four marginal totals,  $d_i^{(1)}$  defines the whole table (1.2). Under the null hypothesis  $H_0$  of no treatment effect,  $d_i^{(1)}$  follows a hypergeometric distribution. That is:

$$P(d_i^{(1)} = d) = \binom{d_i}{d} \binom{R_i - d_i}{R_i^{(1)} - d} / \binom{R_i}{R_i^{(1)}},$$

where  $d$  takes values  $d = \max(0, d_i - R_i^{(0)}), \dots, \min(d_i, R_i^{(1)})$ .

Therefore, the mean  $E_i$  and variance  $V_i$  of  $d_i^{(1)}$  under  $H_0$  are:

$$E_i = \left( \frac{R_i^{(1)}}{R_i} \right) d_i,$$

$$V_i = \frac{R_i - R_i^{(1)}}{R_i - 1} \cdot R_i^{(1)} \left( \frac{d_i}{R_i} \right) \left( 1 - \frac{d_i}{R_i} \right).$$

The logrank test is constructed by comparing the observed event minus the expected number of events in one group. In our case, we denote:

$$O = \sum_{i=1}^l d_i^{(1)} \text{ (total failures in group 1),}$$

$$E = \sum_{i=1}^l E_i,$$

$$V = \sum_{i=1}^l V_i,$$

$$\text{and define the logrank statistic as: } Z = \frac{O - E}{\sqrt{V}} = \frac{\sum_{i=1}^l (d_i^{(1)} - E_i)}{\sqrt{\sum_{i=1}^l V_i}}.$$

It is known that under the null hypothesis  $H_0$  of no treatment effect, the test statistic  $Z$  follows a standard normal distribution ( $Z \sim N(0, 1)$ ) (Cox, 1972),<sup>9</sup> or equivalently  $Z^2 \sim \chi_1^2$ . Hence, we would reject the null hypothesis in favor of the alternative for large values of  $Z$ .

### 1.3.2 Testing treatment effect using the relevant endpoint and the composite endpoint

Consider that the effect of treatment is to be evaluated based on the time  $T_R^{(j)}$  to a relevant event  $\mathcal{E}_R$ , where the superscript  $j$  indicates the treatment group ( $j = 0$  for the control group and  $j = 1$  for the treatment group). Assume now that an additional endpoint  $\mathcal{E}_A$  is considered as a component of the primary endpoint and the composite endpoint  $\mathcal{E}_* = \mathcal{E}_R \cup \mathcal{E}_A$  is to be used instead, in order to prove the efficacy of the new treatment. The effect of treatment would then be evaluated with the time  $T_*^{(j)}$  to  $\mathcal{E}_*$ , where  $T_*^{(j)} = \min\{T_R^{(j)}, T_A^{(j)}\}$ , and  $T_A^{(j)}$  stands for the time to  $\mathcal{E}_A$  for each group. We will also assume that end-of-study censoring at time  $\tau$  is the only non-informative censoring cause for both groups; this assumption indirectly implies that the censoring mechanism is the same for both groups.

Let  $\tilde{\lambda}_k^{(j)}(t)$  ( $k = R, A$ ) denote a marginal or cause-specific hazard for the relevant endpoint ( $k = R$ ) or the additional endpoint ( $k = A$ ), such that  $\tilde{\lambda}_k^{(j)}(t) = \lambda_k^{(j)}(t)$  are the marginal hazards of  $T_k^{(j)}$  when there are no competing causes, and  $\tilde{\lambda}_k^{(j)}(t) = \lambda_{Ck}^{(j)}(t)$  denote the cause-specific hazards of  $T_k^{(j)}$  when either  $T_R^{(j)}$  is a competing cause for  $T_A^{(j)}$ , or  $T_A^{(j)}$  is a competing cause for  $T_R^{(j)}$ .

If we were to use the relevant endpoint as primary endpoint in a clinical trial, the treatment effect would be tested with the logrank test, being the null hypothesis of no effect given by  $H_0 : \text{HR}_R(t) = \tilde{\lambda}_R^{(1)}(t) / \tilde{\lambda}_R^{(0)}(t) = 1$  and the alternative that the new treatment improves survival by  $H_a : \text{HR}_R(t) < 1$ .

Analogously, if we were to use the composite endpoint as primary endpoint, the treatment effect would be tested with the logrank test to compare  $H_0^* : \text{HR}_*(t) = \lambda_*^{(1)}(t)/\lambda_*^{(0)}(t) = 1$  versus  $H_a^* : \text{HR}_*(t) < 1$ .

## 1.4 The Asymptotic Relative Efficiency (ARE) of two tests

A statistical test is more efficient against another if for a given significance level the former leads to a higher power. However, for any fixed alternative to a null hypothesis, say  $H_0 : \theta = \theta_0$ , the power of any test will typically go to 1 if the number of observations is sufficiently large (Noether, 1954).<sup>48</sup> Pitman defines a sequence of alternatives hypotheses around the null with *asymptotic power* converging to some number less than 1, as  $n \rightarrow \infty$ , to compare the efficiency of two tests. Pitman alternatives are given by:

$$H_{a,n} : \theta_n = \theta_0 + k/\sqrt{n}, \quad \text{for any fixed } k.$$

Since these alternatives change with  $n$  they form a sequence of alternatives. And since  $\theta_n \rightarrow \theta_0$  as  $n \rightarrow \infty$ , it is called a sequence of local alternatives or contiguous alternatives to the null.

Assume we have two tests based on statistics  $T_{1n}$  and  $T_{2n}$  which follow a normal distribution  $N(\mu_{1n}(\theta), \sigma_{1n}^2(\theta))$  and  $N(\mu_{2n}(\theta), \sigma_{2n}^2(\theta))$ , respectively. We want to test the null hypothesis  $H_0 : \theta = \theta_0$  against the alternatives  $H_{a,1n} : \theta_{1n} = \theta_0 + k_1/\sqrt{n_1}$  and  $H_{a,2n} : \theta_{2n} = \theta_0 + k_2/\sqrt{n_2}$  for  $T_{1n}$  and  $T_{2n}$ , respectively, with  $k_1 > 0$  and  $k_2 > 0$ . The power of  $T_{1n}$  and  $T_{2n}$  for a given statistical significance  $\alpha$  is, respectively, given by:

$$\begin{aligned} \Pi_{1n}(\theta_{1n}) &= P(T_{1n} \geq z_\alpha | H_{a,1n}), \\ \Pi_{2n}(\theta_{2n}) &= P(T_{2n} \geq z_\alpha | H_{a,2n}), \end{aligned}$$

where  $z_\alpha$  is the standard normal quantile corresponding to the right tail probability  $\alpha$ .

We define the *relative efficiency* of  $T_{2n}$  with respect to  $T_{1n}$  as the ratio of sample sizes  $n_1/n_2$  that are required to achieve the same power for the same alternative hypothesis for a given significance level (Lehmann and Romano, 2005).<sup>41</sup> For large values of  $n$ , Pitman defines the *asymptotic relative efficiency* as:

$$ARE(T_{2n}, T_{1n}) = \frac{n_1}{n_2} = \lim_{n \rightarrow \infty} \frac{\mu'_{2n}(\theta_0)/\sigma_{2n}(\theta_0)}{\mu'_{1n}(\theta_0)/\sigma_{1n}(\theta_0)} = \lim_{n \rightarrow \infty} \frac{R_{2n}^2(\theta_0)}{R_{1n}^2(\theta_0)},$$

where  $\mu'_{in}(\theta_0) = \partial \mu_{in}(\theta_0)/\partial \theta$  for  $i = (1, 2)$  and  $R_{1n}^2(\theta_0)$ ,  $R_{2n}^2(\theta_0)$  are called the efficacies of the test  $T_{1n}$  and  $T_{2n}$ , respectively. Therefore, the  $ARE(T_{2n}, T_{1n})$  is given by the limit of the ratio of the efficacies of the two tests, which is equivalent to the limit of the corresponding ratio  $n_1/n_2$ .

## 1.5 The ARE method by Gómez and Lagakos

Gómez and Lagakos (2013) develop a statistical methodology that helps to decide between using a relevant endpoint  $\mathcal{E}_R$  instead of a composite endpoint  $\mathcal{E}_*$ , consisting of the union of  $\mathcal{E}_R$  plus another additional endpoint  $\mathcal{E}_A$ , to evaluate the effect of a treatment. Their strategy is based on using the value of the asymptotic relative efficiency to assess the efficiency of the logrank test  $Z_R$ , which is based on the relevant endpoint, and comparing it with the efficiency of the logrank test  $Z_*$ , which is based on the composite endpoint.

Endpoints  $\mathcal{E}_R$  and  $\mathcal{E}_A$  may or may not include terminating events, also called fatal events. This leads to four different censoring situations, referred to as Cases 1, 2, 3 and 4 (see Table 1.3):

**Case 1:** Neither of the two endpoints ( $\mathcal{E}_R, \mathcal{E}_A$ ) includes a terminating event. We observe  $T_k^{(j)}$  ( $k = R, A$ ) if  $\mathcal{E}_k$  occurs before the right-censoring time  $\tau$ .

**Case 2.** The relevant endpoint does not include a terminating event while the additional endpoint does. Hence, we observe  $T_R^{(j)}$  if  $T_R^{(j)} < \min\{T_A^{(j)}, \tau\}$ , and we observe  $T_A^{(j)}$  if  $T_A^{(j)} < \tau$ .

**Case 3.** The relevant endpoint includes a terminating event, but the additional endpoint does not. Thus, we observe  $T_R^{(j)}$  if  $T_R^{(j)} < \tau$ , and we observe  $T_A^{(j)}$  if  $T_A^{(j)} < \min\{T_R^{(j)}, \tau\}$ .

**Case 4.** Both endpoints include a terminating event. We observe  $T_R^{(j)}$  if  $T_R^{(j)} < \min\{T_A^{(j)}, \tau\}$ , and we observe  $T_A^{(j)}$  if  $T_A^{(j)} < \min\{T_R^{(j)}, \tau\}$ .

	Relevant endpoint	Additional endpoint
<b>Case 1</b>	<i>NT</i>	<i>NT</i>
<b>Case 2</b>	<i>NT</i>	<i>T</i>
<b>Case 3</b>	<i>T</i>	<i>NT</i>
<b>Case 4</b>	<i>T</i>	<i>T</i>

**Table 1.3:** Four possible combinations depending on whether the relevant endpoint or the additional endpoint includes a terminating event (*T*). *NT* stands for non-terminating event.

Consider the marginal or cause-specific hazard function  $\tilde{\lambda}_R^{(0)}(t)$  of the relevant endpoint in the control group as fixed and define a sequence of alternatives  $H_{a,n}$  that consist of instantaneous hazard functions that are close enough to  $\tilde{\lambda}_R^{(0)}(t)$ ; for instance, by taking  $\tilde{\lambda}_{R,n}^{(1)}(t) = \tilde{\lambda}_R^{(0)}(t)e^{g(t)/\sqrt{n}}$  for some  $g(t)$  function. Logrank  $Z_R$  is asymptotically  $N(0, 1)$  under the null hypothesis of no treatment difference ( $H_0 : \text{HR}_R(t) = \tilde{\lambda}_{R,n}^{(1)}(t)/\tilde{\lambda}_R^{(0)}(t) = 1$ ); and under the sequence of alternatives  $H_{a,n}$  it is asymptotically normal with unit variance and mean  $\mu_R$  given in equation (1.1)(Schoenfeld, 1983).<sup>64</sup> Analogously, the hypothesis of no treatment effect when using the composite endpoint  $\mathcal{E}_*$  is stated as  $H_0^* : \lambda_*^{(1)}(t)/\lambda_*^{(0)}(t) = \text{HR}_*(t) = 1$ , and the sequence of alternatives  $H_{a,n}^* : \text{HR}_{*,n}(t) = e^{g_*(t)/\sqrt{n}}$  for a given function  $g_*(t)$ . The statistic  $Z_*$  is asymptotically  $N(0, 1)$  under  $H_0^*$  and asymptotically normal with unit variance and mean  $\mu_*$  given in equation (1.2) under the sequence  $H_{a,n}^*$ . The asymptotic means of  $Z_R$



and  $Z_*$ , called the non-centrality parameters, are given by:

$$\mu_R = \frac{\int_0^\infty g(t)p(t)[1-p(t)]\Pr_{H_0}\{U \geq t\}\tilde{\lambda}_R^{(0)}(t)dt}{\sqrt{\int_0^\infty p(t)[1-p(t)]\Pr_{H_0}\{U \geq t\}\tilde{\lambda}_R^{(0)}(t)dt}}, \quad (1.1)$$

$$\mu_* = \frac{\int_0^\infty g_*(t)p_*(t)[1-p_*(t)]\Pr_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)dt}{\sqrt{\int_0^\infty p_*(t)[1-p_*(t)]\Pr_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)dt}}, \quad (1.2)$$

where  $U = \min\{T_R, \tau\}$  (in Cases 1 and 3),  $U = \min\{T_R, T_A, \tau\}$  (in Cases 2 and 4) and  $U_* = \min\{T_*, \tau\}$  denote the observed outcome;  $\tau$  denotes the censoring time;  $p(t) = \Pr_{H_0}\{X = 1|U \geq t\}$  and  $p_*(t) = \Pr_{H_0^*}\{X = 1|U_* \geq t\}$  are the null probabilities that someone at risk at time  $t$  is in treatment group 1;  $\Pr_{H_0}\{U \geq t\}$  and  $\Pr_{H_0^*}\{U_* \geq t\}$  are the null probabilities that someone is still at risk at time  $t$  and  $\Pr_{H_0}\{U \geq t\}\tilde{\lambda}_R^{(0)}(t)$  and  $\Pr_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)$  correspond to the probabilities, under the null hypothesis, of observing events  $\mathcal{E}_R$  and  $\mathcal{E}_*$ , respectively, at time  $t$ .

To evaluate the difference in efficiency between using the logrank test with the relevant endpoint based on  $Z_R$  versus with the composite endpoint based on  $Z_*$ , Gómez and Lagakos base their strategy on the behavior of the *ARE* of  $Z_*$  versus  $Z_R$  given by:

$$\text{ARE}(Z_*, Z_R) = \left(\frac{\mu_*}{\mu_R}\right)^2, \quad (1.3)$$

where  $\mu_R$  and  $\mu_*$  are to be replaced by expressions (1.1) and (1.2).

Gómez and Lagakos derive an expression of the *ARE* based on parameter values that can be anticipated at the design stage of a clinical trial. Next, we describe the assumptions established and the anticipatable<sup>1</sup> parameters needed to compute *ARE*.

### 1.5.1 Assumptions

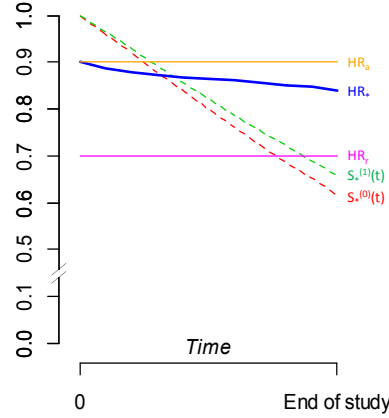
#### Censoring

The end-of-study censoring at time  $\tau$  ( $\tau = 1$  without loss of generality) is the only non-informative censoring cause for both groups. This assumption implies that the censoring mechanism is the same for both groups.

#### Proportional hazards

The hazard ratios between  $T_R^{(0)}$  and  $T_R^{(1)}$  and between  $T_A^{(0)}$  and  $T_A^{(1)}$  are constant, that is,  $\text{HR}_R(t) = \tilde{\lambda}_R^{(1)}(t)/\tilde{\lambda}_R^{(0)}(t) = \text{HR}_R$  and  $\text{HR}_A(t) = \tilde{\lambda}_A^{(1)}(t)/\tilde{\lambda}_A^{(0)}(t) = \text{HR}_A$  for all  $t$ . Note that although we are assuming that the hazard functions  $\tilde{\lambda}_R^{(j)}(t)$  and  $\tilde{\lambda}_A^{(j)}(t)$  ( $j = 0, 1$ ) are proportional, this does not imply the proportionality of hazards  $\lambda_*^{(0)}(t)$  and  $\lambda_*^{(1)}(t)$  for the composite endpoint  $T_*$  (see Figure 1.2). Indeed, the hazard ratio is only constant for specific scenarios, as we will describe further in another chapter of this thesis.

<sup>1</sup>Anticipatable: capable of being anticipated (Source: <http://www.collinsdictionary.com>).



**Figure 1.2:** Hazard ratios  $HR_R$ ,  $HR_A$  and  $HR_*$  of the relevant, additional and composite endpoint; and survival functions  $S_*^{(0)}(t)$  and  $S_*^{(1)}(t)$  of the composite endpoint for each group. Marginal Weibull distributions are assumed for the times to the relevant and additional endpoints.

### Copula assumptions

We can approach the bivariate distribution  $(T_R, T_A)$  by decoupling the joint survival of  $(T_R, T_A)$  into univariate components using a copula model. Gómez and Lagakos consider a Frank Archimedean survival copula (Trivedi PK and Zimmer DM, 2005),<sup>73</sup> given by:

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

where  $\theta (-\infty < \theta < \infty)$  is an association parameter between  $T_R$  and  $T_A$ . Perfect positive and negative dependence between marginals are achieved when  $\theta$  tends to  $\infty$  and  $-\infty$ , respectively. When  $\theta$  tends to 0,  $T_R$  and  $T_A$  are close to being independent. Other copulas could also be considered (Plana-Ripoll and Gómez, 2015).<sup>51</sup> The association parameter  $\theta$  is biunivocally related to Spearman's rank correlation  $\rho (-1 < \rho < 1)$  given by  $\rho = \rho(\theta) = 1 - \frac{12}{\theta} \left[ \frac{1}{\theta} \int_0^\theta \frac{t}{e^t - 1} dt - \frac{2}{\theta^2} \int_0^\theta \frac{t^2}{e^t - 1} dt \right]$ .

### Spearman's correlation

The spearman rank correlation  $\rho$  between  $T_R$  and  $T_A$  is given by (Schweizer, 1981):<sup>66</sup>

$$\rho(T_R, T_A) = 12 \int_0^{+\infty} \int_0^{+\infty} (F(t_R, t_A) - F_R(t_R)F_A(t_A)) dF_R(t_R) dF_A(t_A), \quad (1.4)$$

where  $F_R(t_R)$  and  $F_A(t_A)$  are the distribution functions of  $T_R$  and  $T_A$ , respectively, and  $F(t_R, t_A)$  is the joint distribution function of  $(T_R, T_A)$ .

The joint survival probability is given by:

$$S(t_R, t_A) = P(T_R > t_R, T_A > t_A) = 1 - F_R(t_R) - F_A(t_A) + F(t_R, t_A) = S_R(t_R) + S_A(t_A) - 1 + F(t_R, t_A), \quad (1.5)$$

where  $S_R(t_R) = 1 - F_R(t_R)$  and  $S_A(t_A) = 1 - F_A(t_A)$  are the survival functions of  $T_R$  and  $T_A$ , respectively.

It follows from (1.5)

$$F(t_R, t_A) = S(t_R, t_A) - S_R(t_R) - S_A(t_A) + 1,$$

and

$$F(t_R, t_A) - F_R(t_R)F_A(t_A) = S(t_R, t_A) - S_R(t_R) - S_A(t_A) + 1 - (1 - S_R(t_R))(1 - S_A(t_A)) = S(t_R, t_A) - S_R(t_R)S_A(t_A).$$

Hence from (1.4), we can express  $\rho(T_R, T_A)$  in terms of the survival functions as follows:

$$\rho(T_R, T_A) = 12 \int_0^{+\infty} \int_0^{+\infty} (S(t_R, t_A) - S_R(t_R)S_A(t_A)) dS_R(t_R) dS_A(t_A). \quad (1.6)$$

We can also define  $\rho$  in terms of the survival copula  $C_S(S_R(t_R), S_A(t_A)) = S(t_R, t_A)$  as:

$$\rho(T_R, T_A) = 12 \int_0^{+\infty} \int_0^{+\infty} (C_S(S_R(t_R), S_A(t_A)) - S_R(t_R)S_A(t_A)) dS_R(t_R) dS_A(t_A).$$

If we make the substitution  $u = S_R(t_R)$  and  $v = S_A(t_A)$ , the  $\rho(T_R, T_A)$  simplifies to:

$$\rho(T_R, T_A) = 12 \int_0^1 \int_0^1 (C_S(u, v) - uv) dudv.$$

Spearman's  $\rho$  can be as well defined in terms of the probability of concordance and discordance of two vectors (Nelsen, 2006).<sup>47</sup> Let  $(T_{R1}, T_{A1})$ ,  $(T_{R2}, T_{A2})$  and  $(T_{R3}, T_{A3})$  be three independent random vectors with common joint distribution function  $F(t_R, t_A)$ . The population version of  $\rho$  is defined to be proportional to the probability of concordance minus the probability of discordance for the two vectors  $(T_{R1}, T_{A1})$  and  $(T_{R2}, T_{A3})$  as:

$$\rho(T_R, T_A) = 3(P[(T_{R1} - T_{R2})(T_{A1} - T_{A3}) > 0] - P[(T_{R1} - T_{R2})(T_{A1} - T_{A3}) < 0]).$$

### Joint distribution of $(T_R^{(j)}, T_A^{(j)})$

Assuming equal association parameter  $\theta$  for groups 0 and 1, the joint survival and joint density for  $(T_R^{(j)}, T_A^{(j)})$  ( $j=0,1$ ) are given by:

$$S_{(R,A)}^{(j)}(t_R, t_A; \theta) = -\frac{1}{\theta} \log \left\{ 1 + \frac{(e^{-\theta S_R^{(j)}(t_R)} - 1)(e^{-\theta S_A^{(j)}(t_A)} - 1)}{e^{-\theta} - 1} \right\}$$

$$f_{(R,A)}^{(j)}(t_R, t_A; \theta) = \frac{\theta e^{-\theta(S_R^{(j)}(t_R) + S_A^{(j)}(t_A))}}{(1 - e^{-\theta}) e^{-2\theta S_{(R,A)}^{(j)}(t_R, t_A; \theta)}} [f_R^{(j)}(t_R)] [f_A^{(j)}(t_A)], \quad (1.7)$$

where  $S_R^{(j)}(t_R)$  and  $f_R^{(j)}(t_R)$ ,  $S_A^{(j)}(t_A)$  and  $f_A^{(j)}(t_A)$  are the survival and marginal densities of  $T_R^{(j)}$  and  $T_A^{(j)}$ , respectively. The survival function of  $T_*^{(j)} = \min\{T_R^{(j)}, T_A^{(j)}\}$  is given by:

$$S_*^{(j)}(t; \theta) = P(T_*^{(j)} > t) = P(T_R^{(j)} > t, T_A^{(j)} > t) = C(S_R^{(j)}(t), S_A^{(j)}(t); \theta) = S_{(R,A)}^{(j)}(t, t; \theta).$$

### Marginal laws of $T_R^{(j)}$ and $T_A^{(j)}$

Regarding the marginal laws of  $T_R^{(j)}$  and  $T_A^{(j)}$ , the Weibull distributions are chosen since they are widely used in survival analysis due to their flexibility, allowing decreasing, constant and increasing hazard functions. Hence, for both treatment groups ( $j = 0, 1$ ) the survival function is given by

$$S_k^{(j)}(t) = \exp\{-(t/b_k^{(j)})^{\beta_k^{(j)}}\} \quad (k = R, A),$$

where  $b_k^{(j)}$  and  $\beta_k^{(j)}$  are the scale and shape parameters, respectively, for  $T_k^{(j)}$ . The shape parameters are chosen equal for both groups, that is  $\beta_k^{(0)} = \beta_k^{(1)} = \beta_k$ , so that the assumption of proportionality of the hazards holds.

Denoting by  $p_R$  and  $p_A$  the probabilities of observing  $\mathcal{E}_R$  and  $\mathcal{E}_A$  in group 0, respectively, they are related to the marginal law of  $T_R^{(0)}, T_A^{(0)}$  and the bivariate law of  $(T_R^{(0)}, T_A^{(0)})$  as follows:

$$p_R = \begin{cases} \Pr\{T_R^{(0)} < 1\} = 1 - S_R^{(0)}(1) & \text{Cases 1, 3} \\ \Pr\{T_R^{(0)} < \min\{T_A^{(0)}, 1\}\} = \int_0^1 \int_u^\infty f_{(R,A)}^{(0)}(u, v; \theta) dv du & \text{Cases 2, 4} \end{cases}$$

$$p_A = \begin{cases} \Pr\{T_A^{(0)} < 1\} = 1 - S_A^{(0)}(1) & \text{Cases 1, 2} \\ \Pr\{T_A^{(0)} < \min\{T_R^{(0)}, 1\}\} = \int_0^1 \int_v^\infty f_{(R,A)}^{(0)}(u, v; \theta) dudv & \text{Cases 3, 4,} \end{cases}$$

where  $f_{(R,A)}^{(0)}(u, v; \theta)$  is the joint density of  $(T_R^{(0)}, T_A^{(0)})$  and is defined in (1.7).

The scale parameters  $b_k^{(0)}$  ( $k = R, A$ ) are derived as follows:

$$\text{Case 1: } b_k^{(0)} = \frac{1}{(-\log(1-p_k))^{1/\beta_k}}.$$

Case 2: The scale parameter  $b_R^{(0)}$  is a function of the joint density  $f_{(R,A)}^{(0)}(\cdot, \cdot; \theta)$ , and it is found as the solution of equation  $p_R = \int_0^1 \int_u^\infty f_{(R,A)}^{(0)}(u, v; \theta) dv du$ . The scale parameter  $b_A^{(0)}$  is a function of  $p_A$  and  $\beta_A$ , and it is given by  $b_A^{(0)} = \frac{1}{(-\log(1-p_A))^{1/\beta_A}}$ .

Case 3:  $b_R^{(0)} = \frac{1}{(-\log(1-p_R))^{1/\beta_R}}$  and  $b_A^{(0)}$  is a function of the joint density  $f_{(R,A)}^{(0)}(\cdot, \cdot; \theta)$  and it is found as the solution of equation  $p_A = \int_0^1 \int_v^\infty f_{(R,A)}^{(0)}(u, v; \theta) du dv$ .

Case 4: The scale parameters  $b_k^{(0)}$  are functions of the joint density  $f_{(R,A)}^{(0)}(\cdot, \cdot; \theta)$  and are found as the simultaneous solution of equations  $p_R = \int_0^1 \int_u^\infty f_{(R,A)}^{(0)}(u, v; \theta) dv du$  and  $p_A = \int_0^1 \int_v^\infty f_{(R,A)}^{(0)}(u, v; \theta) du dv$ .

The scale parameters  $b_R^{(1)}$ ,  $b_A^{(1)}$  are computed so that the assumption of proportionality of the hazards holds, that is,  $b_R^{(1)}$  and  $b_A^{(1)}$  are such that  $\text{HR}_R = \frac{\lambda_R^{(1)}(t)}{\lambda_R^{(0)}(t)}$  and  $\text{HR}_A = \frac{\lambda_A^{(1)}(t)}{\lambda_A^{(0)}(t)}$  for Cases 1 and 3; and  $\frac{\lambda_{CR}^{(1)}(t; \theta)}{\lambda_{CR}^{(0)}(t; \theta)} = \text{HR}_R$  and  $\frac{\lambda_{CA}^{(1)}(t; \theta)}{\lambda_{CA}^{(0)}(t; \theta)} = \text{HR}_A$  for Cases 2 and 4 (we refer to the original publication (Gómez and Lagakos, 2013)<sup>27</sup> for further details).

Under the above assumptions, expression (1.3) for Cases 1 and 3 becomes

$$\text{ARE}(Z_*, Z_R) = \frac{\left( \int_0^1 \log \{ \lambda_*^{(1)}(t) / \lambda_*^{(0)}(t) \} f_*^{(0)}(t) dt \right)^2}{(\log \{ \text{HR}_R \})^2 \left( \int_0^1 f_*^{(0)}(t) dt \right) \left( \int_0^1 f_R^{(0)}(t) dt \right)},$$

and for Cases 2 and 4

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

being

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

where  $f_R^{(0)}(t)$  and  $f_*^{(0)}(t)$  are the density functions of  $T_R^{(0)}$  and  $T_*^{(0)}$ , respectively;  $S_*^{(0)}(t)$  stands for the survival of  $T_*^{(0)}$ ; and  $\pi$  stands for the probability, under the null hypothesis, of being in group 1.

### 1.5.2 The ARE expression as a function of anticipatable parameter values

Gómez and Lagakos express the *ARE* function in terms of a list of parameter values that trialists might anticipate at the design stage of the trial. These parameter values are:

1. The probabilities  $p_R, p_A$  of observing the  $\mathcal{E}_R$  and the  $\mathcal{E}_A$ , respectively, in the control group,
2. the relative treatment effects given by the hazard ratios  $HR_R$  and  $HR_A$ ,
3. the Spearman correlation coefficient  $\rho$  between  $\mathcal{E}_R$  and  $\mathcal{E}_A$ ,
4. decreasing, constant or increasing hazard rates given by the shape parameters  $\beta_R$  and  $\beta_A$  for the Weibull marginal distribution for  $T_R$  and  $T_A$ , respectively, and
5. the probability  $\pi$ , under the null hypothesis, of being in group 1 (only needed for Cases 2 and 4).

The above parameters are all easily interpretable for physicians and investigators. Interpretability is important because researchers will have to decide, a priori, which are the most plausible anticipated values in a study design of a clinical trial; then, based on those chosen values, the decision to adopt or not a composite endpoint will have to be made on the *ARE* results. In cases where investigators cannot specify the exact value of some parameter, they can consider a range of values to evaluate whether the *ARE* affects their decision. The rule for deciding whether or not the composite is recommended will be based on the following:

- When  $ARE(Z_*, Z_R) > 1 \Rightarrow$  the composite endpoint should be used instead of the relevant.
- When  $ARE(Z_*, Z_R) \leq 1 \Rightarrow$  the relevant endpoint alone should be used.

## 1.6 Goals and thesis structure

The proper choice of the primary endpoint is crucial to achieve the main objectives in a randomized clinical trial. The asymptotic relative efficiency provides a method on which base the decision on the primary endpoint. Practical guidelines from the methodology can be of great help for trialists in their field of research. Furthermore, computational tools that facilitate the use of methodological developments for a specific study might significantly help clinicians in the design of the study.

The present work is organized as follows:

In chapter 2, we carry out a systematic search for the use of composite endpoints in the cardiovascular field. Based on randomized clinical trials published in 2008, the *ARE* method is applied in order to set general recommendations. This research gave rise to a publication in the journal *Circulation: Cardiovascular Quality and Outcomes* (Gómez, Gómez-Mateu and Dafni, 2014).<sup>26</sup>

In chapter 3 we prove that the usual interpretation of the Asymptotic Relative Efficiency holds when two different sets of hypotheses are set. This result implies that the *ARE* can be interpreted as the ratio of the required sample sizes when using the relevant or the composite endpoint as the primary endpoint. A publication in *SORT* (Gómez and Gómez-Mateu, 2014)<sup>24</sup> presents these developments as well as a simulation for empirically analyzing how close we are to the limiting relationship between the *ARE* and the required sample sizes for a finite number of patients.

In chapter 4 we describe *CompARE*,<sup>28</sup> a web-based platform that we have developed as a tool for making the methodology widely applicable within the scientific community. It is of great help when planning a clinical trial, since it quantifies how efficient a relevant subset of outcomes is with respect to a larger subset of outcomes. *CompARE* is a user-friendly, free tool and, although it is programmed internally in *R*, users need no knowledge of *R*, nor do they need to install it on their computers. *CompARE* is currently accessible at the following website: <https://cinna.upc.edu/compare>.

Chapter 5 provides practical solutions for assigning anticipated probabilities and hazard ratios when the relevant or the additional endpoints consist of several components. This is an important issue in practice, since investigators may know the anticipated probabilities and hazard ratios of each one of the components rather than the combined probabilities and hazard ratios. A letter to the editor in *Statistics in Medicine* (Gómez, Gómez-Mateu, 2016)<sup>23</sup> is published based on the evaluation of the *ARE* depending on the choice of different combined parameter values.

In chapter 6, we present preliminary extensions of *CompARE*. Specifically, some settings assumed by default in Chapter 4 are extended: i) we allow to assume Weibull distributions with decreasing or increasing hazard rates, different correlations and copulas other than Frank's, ii) possible values for the combined probabilities and hazard ratios are also included in the platform when the relevant or the additional endpoints consist of several components, iii) computations where both the relevant and additional endpoint include death, and iv) sample size computations for achieving a specific power at a fixed significance level when using the relevant or the composite endpoint.

Finally, in the closing chapter 7 we describe new lines of research.

The complete list of references used in the literature search in chapter 2, the *R* code to perform the computations of the *ARE*, some methodological details of chapter 5 and the publications derived from this thesis are included in the Appendix.



## THE COMPOSITE ENDPOINT IN THE CARDIOVASCULAR AREA

Composite endpoints are commonly used in randomized clinical cardiovascular trials to assess the efficacy of a new treatment. In this field, the often rare event of the relevant primary endpoint (individual or composite), such as cardiovascular death, myocardial infarction, or both, is combined with a more common secondary endpoint, such as target lesion revascularization, with the aim to increase the statistical power of the study. This increase, in the case of time-to-event endpoints, is expected to be achieved by the inclusion of component endpoints that occur with higher frequency or earlier than the main events of interest.<sup>19</sup> However, in some scenarios, adding specific components might in fact lead to loss of power to detect the true treatment differences.

The objective of the present study is to examine systematically the use of composite endpoints (CE) in cardiovascular randomized clinical trials (RCT), to illustrate the *ARE* method by means of case studies, and to use it to provide guidelines for the informed choice of the primary endpoint in the context of cardiovascular clinical trials.

**The contents of this chapter have been published in<sup>26</sup> (See Appendix):**

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

This chapter is based on the above paper after excluding the introduction because its content has been detailed in the previous chapter. The notation used here is coherent with the rest of this thesis, and differs slightly from the one used in the paper.

In this chapter we have updated the computations of the *ARE* values for case study 2 in section 2.3.2 and figure 2.5, as well as part of the writing from the published one. After the publication of the paper, we were aware of a misspecification in the *R* code when calculating the survival of the composite endpoint in control group for Case 2, which has been corrected. Luckily, the specific recommendation for the case study have not changed and also the general recommendations with the computed scenarios remains similar except that correlation is not affecting with the same strength.

**Circulation**  
Cardiovascular Quality and Outcomes



**Informed Choice of Composite End Points in Cardiovascular Trials**  
Guadalupe Gómez, Moisés Gómez-Mateu and Urania Dafni

*Circ Cardiovasc Qual Outcomes*. published online January 14, 2014;  
*Circulation: Cardiovascular Quality and Outcomes* is published by the American Heart Association, 7272  
Greenville Avenue, Dallas, TX 75231  
Copyright © 2014 American Heart Association, Inc. All rights reserved.  
Print ISSN: 1941-7705. Online ISSN: 1941-7713

## 2.1 Background

### 2.1.1 Composite endpoints in cardiovascular research

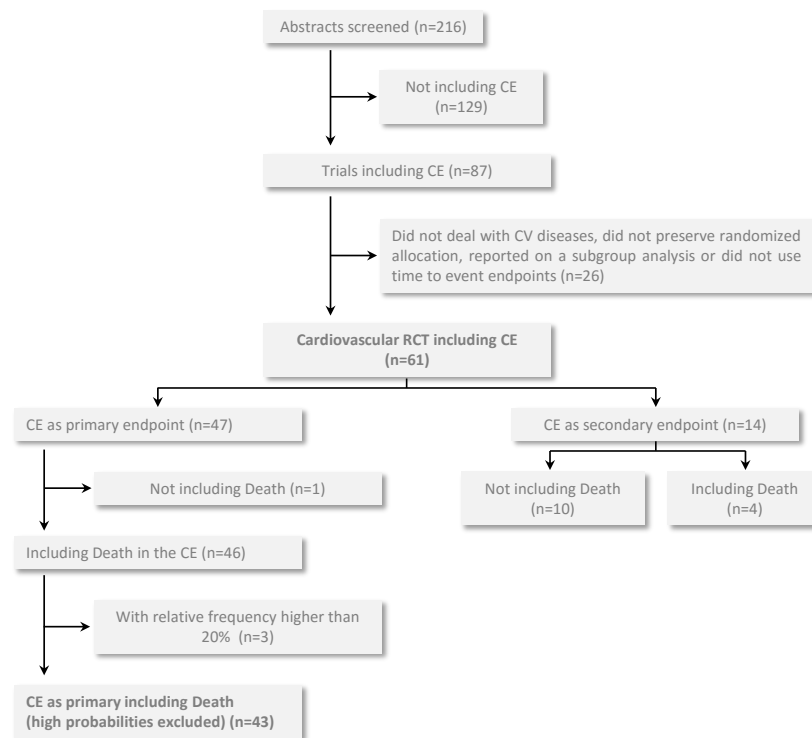
In the past 15 years, many authors have addressed the issue of using and interpreting CEs in the cardiovascular research area. In what follows we present a brief summary of relevant readings. Freemantle et al<sup>19,20</sup> examine the use of CE in major clinical trials, by means of a selection of 167 RCT (with a total of 300,276 patients), that include a primary CE incorporating all-cause mortality, assess the arguments for and against CE, and provide guidance on their applications and reporting. He acknowledges the inadequate reporting of CEs used as primary outcome measures in randomized trials, concluding that, often, the reported results apply to the individual components of the CE rather than to the overall CE.

Ferreira-González et al<sup>14-16</sup> use MEDLINE to conduct 2 systematic reviews to investigate the rationale, potential problems and solutions of using CEs. They point out that the CE, by capturing the net benefit of the intervention, could give a more appropriate reflection of the clinical spectrum of important outcomes associated with the disease being treated than would any component alone. In the conclusions, it is stated that the use of CE is often complicated by the magnitude of the effect of treatment across component endpoints and by the relative importance of the different components for the patients. The reader is referred to Huque et al<sup>35</sup> for an excellent introduction together with some key considerations for using a CE. They present as well some solutions through applications of multiple testing strategies.

## 2.2 Survey of use of composite endpoints in the cardiovascular literature

### 2.2.1 Identification of published clinical trials that used composite endpoints

We explore the use of CE in recent literature through a systematic Medline search covering the 2008 publication of RCTs in 6 high impact medical journals (Table 2.1). Medline search was restricted to randomized controlled trial and human subjects publications, including the terms coronary artery disease, valvular heart disease, arrhythmia, cardiomyopathy, congestive, heart failure, cardiovascular, or cardiovascular disease in the abstract, title, or keywords. The systematic search resulted in 216 publications. The ones that mentioned in the abstract, title, or keywords, a composite or combined endpoint, or the specific endpoints of MACE, or Net Adverse Clinical Events (NACE) were selected (87 of 216). Studies that dealt with other diseases, or looked at subgroup, or nonrandomized comparisons, or did not use time-to-event endpoints were excluded (26 of 87). A total of 61 clinical trials were considered for exploring the use of a CE (Figure 2.1). The breakdown by journal is presented in Table 2.1. The complete reference list is available in the Appendix.



**Figure 2.1:** Flow chart for systematic review of cardiovascular (CV) randomized clinical trials (RCTs). CE indicates composite endpoint.

Journal (Papers and RCT)	Total Articles	%	CE RCT	%
NEJM	46	21%	17	28%
The Lancet	36	17%	13	21%
European Heart Journal	54	25%	12	20%
Circulation	53	25%	10	16%
JAMA	24	11%	9	15%
Annals of Internal Medicine	3	1%	0	0%
<b>Total RCT</b>	<b>216</b>	<b>100%</b>	<b>61</b>	<b>100%</b>

**Table 2.1:** Summary of Medline search, for cardiovascular terms, for 2008 Publication of RCTs.

### 2.2.2 Information abstracted from each RCT

The following information was abstracted from each of the published articles: time to follow-up, sample size, components of each primary and secondary endpoint, frequency of occurrence of each endpoint (CE and components of interest), the corresponding HRs and p-values between groups compared in the trial.

### 2.2.3 Method to set the recommendations

From the information abstracted for each trial together with previously examined scenarios in Gómez and Lagakos,<sup>27</sup> we establish all the possible parameter combinations. However, because not all the combinations of frequencies (control group) and relative treatment effects ( $p_R$ ,  $HR_R$ ) or ( $p_A$ ,  $HR_A$ ) were found in the studied RCTs, we did restrict our computations to published pairs of values ( $p$ ,  $HR$ ). The *ARE* is computed for each of a total of 320 combinations to provide recommendations for the cardiovascular area trials. In all cases, computations have been done assuming that death is part of the RE, modeling the marginal laws of the times to RE and AE as Weibull, representing decreasing, constant and increasing hazard functions, combining each scenario with different degrees of dependence between times to RE and to AE and using  $HR = 0.99$  to represent relative treatment effects of no interest.

Values of  $ARE > 1$  are in favor of using the CE instead of the RE. However, because the advantage of one endpoint over the other is small in the vicinity of 1, we follow, as Gómez and Lagakos did, a general rule to use the CE instead of the RE if  $ARE > 1.1$  and to retain the RE if  $ARE \leq 1.1$ .

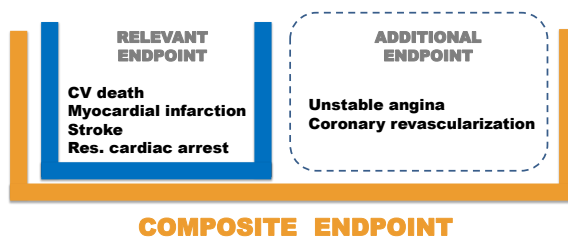
## 2.3 Case studies

Interesting cases of trials leading to a significant result for the RE whereas nonsignificant for the CE, significant for the CE driven by the effect on the RE and nonsignificant for the RE whereas significant for the CE are described next, the first two with greater detail.

### 2.3.1 Case study 1: treating patients after an acute coronary syndrome with Succinobucol

An RCT to assess the effects of the antioxidant succinobucol (AGI-1067)<sup>70</sup> on cardiovascular outcomes in patients with recent acute coronary syndrome already managed with conventional treatments, uses as PE, denoted by CE, the composite of RE (time to first occurrence of cardiovascular death, resuscitated cardiac arrest, MI, stroke), and AE (unstable angina or coronary revascularization; Figure 2.2). A total of 6144 patients having experienced an acute coronary syndrome  $\leq 1$  year before recruitment were randomized to receive succinobucol ( $n=3078$ ) or placebo ( $n=3066$ ), in addition to standard of care. A beneficial effect of succinobucol on RE was found (207 events: succinobucol versus 252 events: placebo;  $HR = 0.81$ ;  $p\text{-value}=0.029$ ). The less important but frequent outcomes (ie, hospitalization

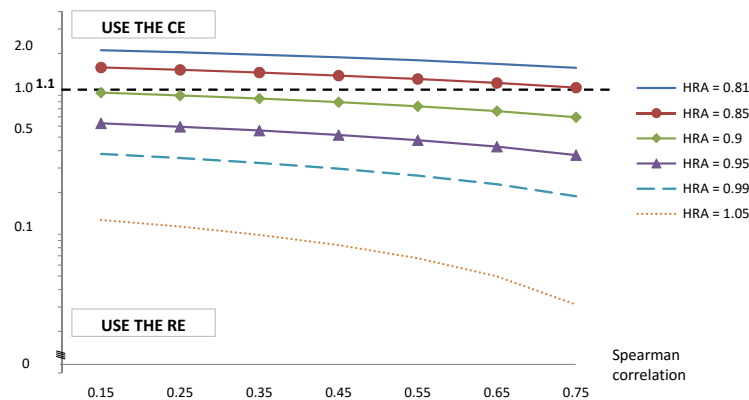
for unstable angina and coronary revascularization) were included in the primary CE. The expectation would be that by the inclusion of these outcomes, the resulting increase in the number of CE events observed would lead to an increase in study power. On the contrary, these endpoints did not differ significantly between the 2 treatment groups, and their contribution of a high relative number of events in the primary CE led to the disappearance of the statistically significant benefit of the active treatment on the important outcomes RE. Thus, the primary CE was not found to be significantly different between treatment groups (530 events: succinobucol versus 529 events: placebo). We have that the probability of observing the RE in control group is  $p_R=8.2\%$  with observed  $HR_R=0.81$ , whereas the probability of observing the AE in control group is  $p_A=10.4\%$  with  $HR_A=1.05$  (it corresponds to coronary revascularization, whereas observed HR for unstable angina is 1.10).



**Figure 2.2:** Pictorial representation of the construction of a composite endpoint as the union of the relevant endpoint and the additional endpoint based on Tardif's randomized clinical trial. CV stands for cardiovascular and Res. stands for resuscitated.

The *ARE* is explored for these parameter values. For all different shapes of the time-to-event distributions (9 combinations including increasing, constant, and decreasing hazard functions) and correlation values ranging from 0.15 to 0.75 (63 scenarios), it is found that the *ARE* is always  $<1.1$ . Following the rule of Gómez and Lagakos, the benefits of using the CE over the RE are marginal and probably too small to justify adding the AE.

The use of CE would be justified in the case that  $HR_A \leq 0.85$ , for all other parameters fixed (ie,  $p_R=8.2\%$ ;  $HR_R=0.81$ ;  $p_A=10.4\%$ ; Figure 2.3). However, if  $HR_A \geq 0.95$  not even an expected frequency of 20% for the AE would justify the use of CE. If  $HR_A=0.9$ , CE would only be justified if  $p_A \geq 20\%$ , and the association between RE and AE is weak (not shown). Thus, under these circumstances, the additional components of coronary revascularization or hospitalization for unstable angina on the primary endpoint (PE) would had only been recommended if the expected beneficial effect of succinobucol on these components would have been approximately as strong as the expected effect on cardiovascular death, resuscitated cardiac arrest, MI, or stroke (Figure 2.4).



**Figure 2.3:** Asymptotic relative efficiency (ARE) of composite versus relevant endpoint (CE and RE, respectively) for a range of Spearman correlation coefficients and different values of the hazard ratio of the additional endpoint  $HR_A$  for the parameters of case study 1 ( $p_R=0.082$ ;  $HR_R=0.81$ ;  $p_A=0.104$ ) and marginal increasing hazards.  $HR_R$  stands for the hazard ratio of the RE; and  $p_R$  and  $p_A$  indicates the probability of RE and additional endpoint in control group, respectively.

#### Recommendations from case study 1:

- Use **Composite Endpoint** when having high treatment effect on AE ( $HR_A \leq 0.85$ ).
- Use **Relevant Endpoint** when having very low treatment effect on AE ( $HR_A \geq 0.95$ ), or low treatment effect ( $HR_A = 0.9$ ) except for high frequency of AE ( $p_A \geq 20\%$ ) and very weak correlation between endpoints.

**Figure 2.4:** Summary of recommendations for case study 1 as a guide to decide between using composite endpoint (CE) or relevant endpoint (RE) as primary endpoint (PE). Values of treatment effect on RE and relative frequency of RE and AE in control group ( $HR_R=0.81$ ;  $p_R=8.2\%$  and  $p_A=10.4\%$ ) are fixed in advance and correspond to Tardif randomized clinical trial.

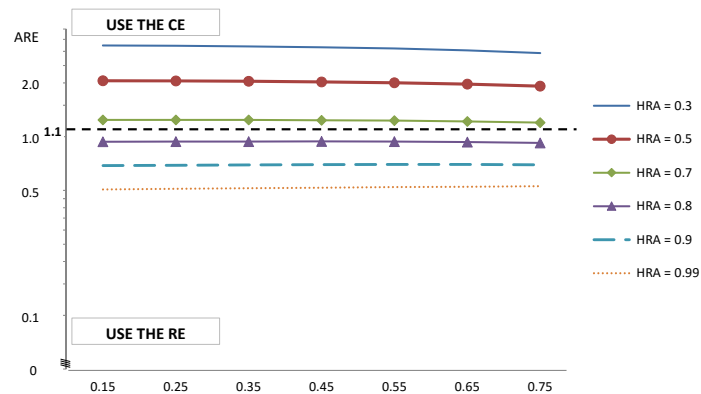
### 2.3.2 Case study 2: treating hemorrhagic complications during primary percutaneous coronary intervention in acute MI

The Harmonizing Outcomes with Revascularization and Stents in the Acute MI (HORIZONS-AMI)<sup>69</sup> study is a prospective, open-label, randomized, multicenter trial in patients with ST-segment–elevation MI presented within 12 hours after the onset of symptoms. In this study, 3602 patients were assigned to treatment with heparin plus a glycoprotein IB/IIa inhibitor ( $n=1802$ ) or the alternative treatment of bivalirudin alone ( $n=1800$ ). The interest lies on whether hemorrhagic complications are reduced, when using bivalirudin alone. Two primary 30-day endpoints were prespecified: (1) major bleeding, denoted by RE and (2) NACE, denoted by CE, a composite of major bleeding and MACE. MACE, denoted by AE, is composed, in this trial, of death, reinfarction, target vessel revascularization for ischemia and stroke. In this case, while major bleeding is the relevant event of interest, the composite CE takes into account all other additional adverse clinical events, including death. According to the results, MACE is almost identical in the 2 groups (98 versus 99 events;  $p$ -value=0.95), whereas major bleeding is statistically significantly lower in the bivalirudin-alone group (89 versus 149 events;  $p < 0.001$ ). The comparison of NACE (166 versus 218 events;  $p$ -value=0.005) between treatment groups is found statistically significant, and as mentioned by the authors, this is entirely driven by the effect on major bleeding. The risk taken by the researchers of combining the endpoint of interest with an endpoint on which treatments have no differential effect is demonstrated using this study.

The probability of observing a major bleeding event (RE) in control group, is  $p_R=8.3\%$  with  $HR_R=0.6$ , whereas the probability of observing a MACE event, AE, is  $p_A=5.5\%$  with  $HR_A=1$ . MACE is occurring with smaller frequency than the RE and in addition the treatment does not have an effect on it. Under these parameter values the *ARE* is examined, as above, for 21 scenarios, corresponding to different shapes of time-to-event distributions (including decreasing, constant, and increasing hazards) and correlation values ranging from 0.15 to 0.75. In all those cases, the *ARE* between a major bleeding event and a MACE event is  $<1.1$ , meaning that the use of the CE (NACE) is not recommended.

Other scenarios were also explored under all above combinations of distributional shapes and correlation values. First, for higher values of the probability of observing a MACE event ( $5.5\% \leq p_A \leq 80\%$ ), the same situation occurs, that is, NACE is neither recommended for any of the cases. Second, the *ARE* was also explored for larger beneficial effects on MACE ( $0.3 \leq HR_A \leq 0.9$ ), and the *ARE* value is  $>1.1$  whenever the treatment effect on the additional endpoint is high ( $HR_A \leq 0.7$ ). Figure 2.5 illustrates the *AREs* for the values of the parameters of this clinical trial ( $p_R=8.3\%$ ;  $HR_R=0.6$ ;  $p_A=5.5\%$ ) and for marginal increasing hazards. We find that the *ARE* is always  $<1.1$  for  $HR_A \geq 0.8$ , indicating that low to weak effects on MACE are not enough to prefer NACE (irrespective of the correlation). However, high or strong beneficial effects on MACE ( $HR_A \leq 0.7$ ) would advocate for the use of the CE, NACE.





**Figure 2.5:** Asymptotic relative efficiency (ARE) of composite versus relevant endpoint (CE and RE, respectively) for a range of Spearman correlation coefficients and different values of the hazard ratio of the additional endpoint  $HR_A$  for the parameters of case study 2 ( $p_R=0.083$ ;  $HR_R=0.6$ ;  $p_A=0.055$ ) and marginal increasing hazards.  $HR_R$  stands for the hazard ratio of the RE; and  $p_R$  and  $p_A$  indicates the probability of RE and additional endpoint in control group, respectively.

It is clear that the chosen PE, NACE, for the efficacy of bivalirudin alone in this study gave unexpected good results and that it was a matter of luck not to have a diluted effect in NACE because the ARE can be as low as 0.51, meaning that major bleeding as a PE can be twice as efficient as NACE.

One could wonder under which circumstances the composite NACE would have been a better, more efficient choice, and by running all the ARE computations for different values of the frequency of observing an AE, we find that for the composite NACE to be justified, at least high treatment effect on MACE ( $HR_A \leq 0.7$ ) with low probability ( $p_A = 0.055$ ) or a treatment effect  $HR_A \leq 0.8$  with high frequency ( $p_A \geq 0.4$ ) is needed.

### 2.3.3 Case Study 3: testing Fondaparinux in patients with ST-segment–elevation MI

In the clinical trial testing fondaparinux in patients with ST-segment–elevation MI,<sup>50</sup> the RE of death and the AE of myocardial reinfarction at 30 days occurred in 12.5% ( $p_R=0.125$ ) and 3.7% ( $p_A=0.037$ ) of control patients, respectively. The CE occurred in 15.1% of control patients, indicating a weak correlation between RE and AE. The corresponding HRs ( $HR_R=0.83$  and  $HR_A=0.66$ ) were both not significantly different than 1. The increased number of events for the CE and the same direction of benefit for both components led to a statistically significant HR with respect to CE of 0.80. In this trial, the use of the CE is clearly indicated by the ARE in 100% of the scenarios.

### 2.3.4 Case study 4: prevention studies

A prevention study assessed the benefit on the risk of cardiovascular disease of low-dose aspirin in the prevention of atherosclerotic events in patients with type 2 diabetes mellitus.<sup>49</sup> Composite PE was defined as fatal and nonfatal ischemic heart disease, fatal or nonfatal stroke, and peripheral arte-

rial disease. This trial could be considered an outlier because of the combination of a low frequency of fatal cardiovascular events ( $p_R=0.008$ ), yet significantly different between groups ( $HR_R=0.10$ ;  $p$ -value=0.0037). The CE occurred in 6.7% of control patients, indicating a weak correlation between RE and AE, leading to a hazard ratio for the CE ( $HR^*$ ) of 0.80 but not statistically significant.

Under these extreme conditions, the use of the RE would have been justified based on the low  $HR_R$ , whereas the use of the CE would have been justified based on the low frequency of events. The  $ARE$  points to the clear choice of the CE for anticipated strong effects of the aspirin on the nonfatal events ( $HR_A \leq 0.2$ ) and the clear choice of the RE for moderate effects ( $HR_A \geq 0.8$ ), whereas for  $HR_A$  values between 0.2 and 0.8, the CE is recommended as  $HR_A$  increases for progressively higher values of the frequency of nonfatal events. In this particular situation, the choice of the CE based on an assumption of a treatment effect at such an extreme value would be difficult to justify at the design stage although it could be taken under consideration for the next trial designed on this question.

## 2.4 Results and recommendations

A CE was used as PE for 47 of the clinical trials and as secondary for the remainder of 14 clinical trials. The frequency of use of different CEs, as well as of each individual component, for the 47 cases that CE is the PE, is presented in Table 2.2. MI and stroke were encountered as components of the CE in over half of these clinical trials (66% and 55%, respectively), Hospitalization and target vessel revascularization are AE in 30% and 13%, respectively, whereas death is encountered in all of them but 1 (46 of 47). In addition, among the 14 trials with an individual PE, in 13 of them death is either the RE (in 4) or used as an AE (in 9).

Endpoint combinations	Death	MI	Stroke	Hospitalization	TVR	N with additional endpoints	N Total (%)
1	X	X	X	...	...	8	14 (30%)
2	X	X	X	X	...	5	8 (17%)
3	X	...	...	X	...	1	6 (13%)
4	X	X	...	...	...	2	5 (11%)
5	X	...	...	...	...	5	5 (11%)
6	X	X	...	...	X	2	4 (9%)
7	X	...	X	...	...	2	2 (4%)
8	X	...	...	...	X	1	1 (2%)
9	X	...	X	...	X	1	1 (2%)
10	...	...	X	...	...	1	1 (2%)
	<b>98%</b>	<b>66%</b>	<b>55%</b>	<b>30%</b>	<b>13%</b>	<b>28</b>	<b>47</b>

**Table 2.2:** Frequencies for different combinations of endpoints for 47 RCTs with composite endpoint as primary endpoint. MI indicates myocardial infarction; and TVR, target vessel revascularization. RCT stands for randomized clinical trial.

For all the trials, including death (46 out of 47), the frequency of death was relatively low (median 4%), with the exception of 3 trials where death was frequent (>20%). The observed relative frequencies of death among the 43 low-frequency studied trials were between 0.002 and 0.15 (Table 2.3). The ob-

served relative frequencies of the AEs (MI, stroke, hospitalization, and target vessel revascularization) were between 0.002 and 0.31. Concerning the relative treatment effects, it was found that some of the component endpoints had an observed HR >1 (17 of 43). Among the clinical trials with HR <1, we have found relative treatment effects for death as small as 0.1 and as large as 0.98 and between 0.35 and 0.94 for the AEs (Table 2.3).

In the reviewed studies, specific combinations of the control group frequencies for the RE ( $p_R$ ) and AE ( $p_A$ ) with corresponding HR values emerged. The *ARE* of a CE with death as a RE adding MI, stroke, or hospitalization as AE is computed for different shapes of time-to-event distributions and a range of correlations between times to RE and AE and is described next to serve as a guide for the design of future trials.

### **Death plus MI**

For the relatively low frequency of MI (AE;  $\leq 12\%$ ) for all HR combinations found in the trials, the CE of death and MI is almost always justified based on the *ARE* except for the case where death and MI present with the same frequency and the beneficial effect on death is higher than on MI ( $HR_A > HR_R$ ).

### **Death plus Stroke**

For particularly low frequency of stroke found in the trials (0.5%), the CE of death and stroke is always justified in the cases that the beneficial effect on stroke is higher than on death ( $HR_A < HR_R$ ). The same is true for the higher frequency of stroke (12%), whereas the CE is also justified when the beneficial effect on stroke is slightly less than on death, but death presents with lower frequency.

### **Death plus Hospitalization**

The CE is justified in the cases that the HR for death is >0.8, or equal to 0.70 coupled with low frequency of death ( $p_R=3\%$ ), whereas the HR for hospitalization is <0.9. For a substantial benefit on death coupled with low frequency ( $HR_R=0.5$ ;  $p_R=6\%$ ), when the frequency of hospitalization is high ( $p_A=39\%$ ) even for a smaller benefit for hospitalization ( $HR_A=0.70$ ), the CE is justified.

The CE is not justified when, even for a substantial benefit on death ( $HR_R \approx 0.5$ ), low frequency of death ( $p_R \approx 6\%$ ), and high frequency of hospitalization ( $p_A \approx 39\%$ ), the benefit for hospitalization is small ( $HR_A > 0.90$ ). The CE is neither justified when  $HR_A > HR_R$  provided that the frequency of death is higher ( $p_R=12\%$ ).

#### **2.4.1 Death from cardiovascular or death from any cause as the individual primary or co-PE**

In only 4 trials, death from cardiovascular or death from any cause was used as the individual primary or co-PE.<sup>7,30,34,60</sup> The frequency of cardiovascular death or any death in 2 of the trials<sup>7,34</sup> on patients with New York Heart Association class II–IV Chronic Heart Failure, or Atrial Fibrillation and New York Heart Association class II or IV heart failure, was 25% and 29%, respectively. In such cases of high death frequency, the use of the CE is justified only when the anticipated treatment benefit for the AE is similar or higher than the one for survival. Such is the case in the trial exploring the effect of n-3

polyunsaturated fatty acids in patients with chronic heart failure,<sup>7</sup> where the use of the CE of death and admission to hospital for cardiovascular reasons as co-PE would be fully supported by the *ARE*.

## 2.4.2 Recommendations for cardiovascular clinical trials

We present recommendations for future design choice between RE and CE for cardiovascular clinical trials that use CEs as an option for the PE, include death as the RE, and add other nonfatal endpoints, such as MI, hospitalization. We discuss the recommendations in terms of the values of the anticipated hazard ratios  $HR_R$  and  $HR_A$ , and, when needed, in terms of the anticipated probabilities of occurrence  $p_R$  and  $p_A$ . These guidelines have been based on the scenarios explored by Gómez and Lagakos and on the 43 clinical trials of the 47 (Table 2.2) having death (observed control group frequency  $\leq 15\%$ ) as RE and stroke, MI, hospitalization, and target vessel revascularization as AE. Table 2.3 shows the observed relative frequencies and relative treatment effects of death and the AEs, and Table 2.4 the possible pairs  $(p, HR)$  for RE and AE, after excluding pairs with  $HR \geq 1$  (17 of 43).

Keeping in mind that the specific decision for a given trial has to be based on a thorough study as has been shown in the case studies and the Results section of this chapter, a set of recommendations on whether to use the RE or the CE is outlined below (Figures 2.6 and 2.7):

- $HR_A < HR_R$ : the relative treatment effect is greater on the AE than on the RE  $\Rightarrow$  CE should always be used.
- $HR_A = HR_R$ : RE and AE have approximately the same relative treatment effect  $\Rightarrow$  CE should almost always be used. Only in those cases where the anticipated probability for AE has a low frequency ( $p_A \leq 0.06$ ) and the frequency for RE is between 2 and 5 times the frequency of the other endpoints ( $2 < p_R/p_A < 5$ ), RE could be a better choice.
- $HR_A = HR_R + 0.1$ : AE has a slightly smaller effect on treatment than RE  $\Rightarrow$  RE should always be used if  $p_R/p_A \geq 3$  and CE should always be used if  $p_R/p_A \leq 0.25$ . Whenever  $0.25 < p_R/p_A < 3$  the decision will depend on the anticipated values of the relative treatment effect, the frequency of observation of either endpoint along with its correlation and to a lesser extent on the shape of the marginal density.
- $HR_A = HR_R + 0.2$ : AE has a smaller effect on treatment than RE  $\Rightarrow$  RE should almost always be used except when the relative frequency of the AE is extremely higher than that of the RE ( $p_R/p_A \leq 0.06$ ).
- $HR_A \geq HR_R + 0.3$ : AE has a much smaller effect on treatment than RE  $\Rightarrow$  RE should always be used.
- $HR_A$  close to 1 and  $p_A \leq 0.005 \Rightarrow$  RE should always be used.

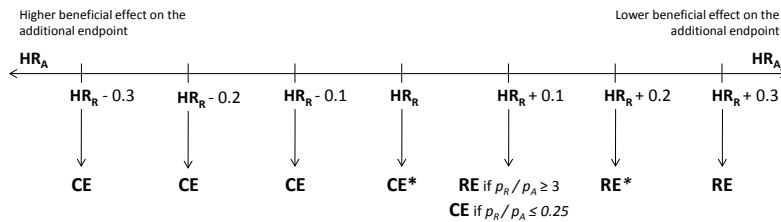
One has also to keep in mind that the association between time to RE and time to AE could play an important role (*ARE* decreases when the correlation between the 2 endpoints increases) and that decisions based on hazard plots as the ones in Figures 2.3 and 2.5 are recommended (*ARE* decreases when the relative effect of treatment on the AE is smaller). Furthermore, the recommendations are to be taken cautiously because infrequent events ( $p$  in the order of 0.005), frequencies of death with order of magnitude larger than the frequency of AE ( $p_R/p_A > 12$ ), and unlikely frequent endpoints ( $p > 0.35$ ) could reverse the direction of the recommendation.

Endpoint	Relative frequency			Hazard Ratio		
	Min.	Median	Max.	Min.	Median	Max.
Death	0.2%	3.8%	15%	0.1	0.83	0.98
Myocardial Infarction	0.2%	3.7%	11.3%	0.35	0.78	0.92
Stroke	0.4%	2.2%	4.7%	0.52	0.83	0.89
Hospitalizations	0.3%	3.6%	31%	0.59	0.75	0.94
TVR	0.7%	7.3%	16.2%	0.79	0.79	0.83

**Table 2.3:** Summary of observed relative frequencies and relative treatment effect among clinical trials with observed frequency of death < 20%. Hazard ratios restricted to clinical trials with  $HR < 1$ . TVR indicates target vessel revascularization.

HR	Probability in control group									
	0.005	0.03	0.06	0.09	0.12	0.15	0.18	0.27	0.39	0.48
<b>0.3</b>	AE	...	...	...	...	...	...	...	...	...
<b>0.4</b>	RE	...	...	...	...	...	...	...	...	...
<b>0.5</b>	AE	...	RE	...	...	...	...	...	...	...
<b>0.6</b>	AE	AE	...	...	...	...	...	...	...	...
<b>0.7</b>	...	RE	RE	RE	...	...	...	...	AE	...
<b>0.8</b>	AE	RE	RE	...	RE	RE	...	...	...	...
<b>0.9</b>	...	AE	AE	RE	RE	...	AE	AE	...	AE
<b>0.99</b>	RE	RE	RE	...	RE	...	...	...	...	...
	AE	AE	AE	...	AE	...	...	...	...	...

**Table 2.4:** Chosen pairs of values ( $p$ , HR) for RE (Death), and AE (Stroke, Myocardial infarction, Hospitalization, and Target Vessel Revascularization) used for the recommendations.  $p$  stands for the anticipated probability of observing the event in control group, and HR the corresponding hazard ratio. Clinical trials with  $HR > 1$  are excluded. AE indicates additional endpoint; and RE, relevant endpoint.



**Figure 2.6:** The horizontal axis represents the values of the hazard ratio  $HR_A$  of additional endpoint (AE) as a function of the  $HR_R$  of relevant endpoint (RE). Each tick summarizes several scenarios, corresponding to different shapes of the marginal hazards and different degree dependences between RE and AE. For each tick, we indicate whether it is advisable to adopt composite endpoint (CE) in preference to RE. See explanation in text for scenarios with  $CE^*$  and  $RE^*$ .

**Use Composite Endpoint when:**

- Treatment effect on AE is higher than on RE.
- Same treatment effect between endpoints except for low frequency of AE ( $p_A \leq 0.06$ ) and ratios between frequencies ( $p_R/p_A$ ) between 2 and 5.
- Slightly smaller treatment effect on AE and small ratio of frequencies ( $p_R/p_A \leq 0.25$ ).

**Use Relevant Endpoint when:**

- Slightly small treatment effect on AE and high ratio of frequencies ( $p_R/p_A \geq 3$ ).
- Smaller treatment effect on AE ( $HR_A = HR_R + 0.2$ ) except for small ratios between frequencies ( $p_R/p_A \leq 0.06$ ).
- Much smaller treatment effect on AE ( $HR_A \geq HR_R + 0.3$ ).
- Very small treatment effect on AE and low frequency on treatment ( $HR_A \approx 1$  and  $p_A \leq 0.005$ ).

**Figure 2.7:** Summary of recommendations as a general guide for using composite or relevant endpoints (RE) as primary endpoint in cardiovascular clinical trials. AE indicates additional endpoint; and HR, hazard ratio.

## 2.5 Discussion

The use of composite PE in cardiovascular randomized trials has been addressed by many authors who have discussed, among other issues, the suitability of components that are clinically less important and the difficulties in interpreting results. Our study helps the trialist, in the design of a future trial, to choose in an objective manner between candidates of PE, by computing the *ARE* based on the anticipated values of the control group frequency and HR of each candidate endpoint.

It is clear that in the cardiovascular context, the CEs under consideration overwhelmingly include a terminal event either as a RE or as an AE. This chapter explores under which circumstances adding other endpoints to a RE of death would result in a more efficient choice. It is clear from our results that, contrary to a common belief, adding a frequent event to a RE of death does not always help and, indeed, may even prove harmful. The fact that the CE increases the number of events, does not mean, even in the case of a common event rate and similar magnitude of the treatment effects, that the required sample size of a trial is reduced because, depending on the strength of the association between RE and AE, the *ARE* is not necessarily  $>1$ .

It is important to point out that the *ARE* method is intended for the planning phase of the RCT. The reader should be aware of the presence of competing risks and how the analysis should appropriately take care of this issue. Chi<sup>6</sup> describes how to properly analyze a RCT based on CE. They recommend to use all-cause mortality instead of cause-specific mortality to prevent from informative censoring, and although not strictly necessary if the CE is valid, to analyze separately the individual components, and to gain a more accurate assessment and interpretation of the clinical benefits and risks involved. They propose 2 basic formats for the presentation of trial data and the results of the analysis.

Finally, although the *ARE* method has been developed with a RCT in mind, well-planned observational studies, viewed as conditionally randomized experiments could take advantage of an appropriately adjusted version of the *ARE* method. Recommendations about reporting completely and accurately an observational study have been developed by the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) initiative.<sup>78</sup> The explanation of how the study size was arrived at is among the requirements included in the STROBE Statement Checklist.<sup>77</sup> Admittedly, the importance of sample size determination in observational studies depends on the context. When planning a new study, formal, a priori calculation of sample size is useful, especially for studies that will gather new data and will be planned for the purpose to overcome potential problems with previous reports. There is even a call for registration of observational studies on a World Health Organization-compliant registry before they begin to lend greater credibility to the study findings.<sup>12</sup> In the case of large, hypothesis-driven cohort studies, there is no doubt that a solid protocol, including sample size and power justification, is required, and in that context, the *ARE* method is as useful for the informed choice of the endpoint as for any well-designed RCT.

As a conclusion, if a well-defined experiment is conducted and if the censoring patterns of both groups can be considered similar, the *ARE* method could be a valid option to discriminate between a RE and a CE.





## RELATIONSHIP BETWEEN THE ARE AND SAMPLE SIZES

The purpose of this chapter is to prove that the usual interpretation of the asymptotic relative efficiency (*ARE*), as the ratio of sample sizes,  $n$  and  $n_*$ , needed to attain the same power for a given significance level, still holds even though two different sets of hypotheses ( $H_0$  versus  $H_a$  and  $H_0^*$  versus  $H_a^*$ ) are compared, where  $H_0, H_0^*$  and  $H_a, H_a^*$  are the null and the alternative hypothesis of no treatment effect evaluated on  $\mathcal{E}_R$  and on  $\mathcal{E}_*$ , respectively.

To clarify the purpose of our investigation consider the following. If we were to test  $H_0$  versus  $H_a$  with two different test statistics  $S_n$  and  $T_m$ , Pitman's relative efficiency would be defined as the ratio  $m/n$ , where  $n$  and  $m$  are the required sample sizes for  $S_n$  and  $T_m$ , respectively, to attain the same power for a given significance level. Furthermore, if both  $S_n$  and  $T_m$  are asymptotically normal with unit variance and means  $\mu_S$  and  $\mu_T$ , it can be proved that Pitman's *ARE* corresponds to the squared of the ratio of the noncentrality parameters, that is  $(\mu_S/\mu_T)^2$ . Gómez and Lagakos' method compares the logrank statistics  $Z$  and  $Z_*$ , derived for two different set of hypotheses  $H_0$  versus  $H_a$  and  $H_0^*$  versus  $H_a^*$  and do so using, as definition of the *ARE*, the ratio  $(\mu_*/\mu)^2$  where  $\mu$  and  $\mu_*$  are, respectively, the asymptotic means of  $Z$  and  $Z_*$ , under alternative contiguous hypotheses to  $H_0$  and  $H_0^*$ .

We carry out a simulation to study under which conditions and for finite sample sizes, the relationship  $ARE(Z_*, Z) = (\mu_*/\mu)^2 = n/n_*$  holds where  $n$  and  $n_*$  are the needed sample sizes for  $Z$  and  $Z_*$ , respectively, to attain the same power for a given significance level.

**The contents of this chapter have been published in<sup>24</sup> (see Appendix):**

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

This chapter reproduces the paper after excluding the introduction because its content has been detailed in previous chapters. The notation used here is coherent with the rest of this thesis, and differs slightly from the one used in the paper.

SORT 38 (1) January-June 2014, 73-88

## The asymptotic relative efficiency and the ratio of sample sizes when testing two different null hypotheses

Guadalupe Gómez\*,<sup>1</sup> and Moisés Gómez-Mateu<sup>1</sup>

---

### Abstract

Composite endpoints, consisting of the union of two or more outcomes, are often used as the primary endpoint in time-to-event randomized clinical trials. Previously, Gómez and Lagakos provided a method to guide the decision between using a composite endpoint instead of one of its components when testing the effect of a treatment in a randomized clinical trial. Consider the problem of testing the null hypotheses of no treatment effect by means of either the single component or the composite endpoint. In this paper we prove that the usual interpretation of the asymptotic relative efficiency as the reciprocal ratio of the sample sizes required for two test procedures, for the same null and alternative hypothesis, and attaining the same power at the same significance level, can be extended to the test procedures considered here for two different null and alternative hypotheses. A simulation to study the relationship between asymptotic relative efficiency and finite sample sizes is carried out.

---

MSC: 62N03, 62P10

*Keywords:* Asymptotic relative efficiency, composite endpoint, logrank test, sample size, simulation, survival analysis.

### 3.1 Notation, the logrank test and the asymptotic relative efficiency

#### 3.1.1 The logrank tests for the relevant and for the composite endpoints

Assume that we have a two-arm study involving random assignment to an active ( $X = 1$ ) or control treatment ( $X = 0$ ) aiming to prove the efficacy of the new active treatment. The effect of treatment is to be evaluated on the time  $T_R^{(j)}$  to a relevant event  $\mathcal{E}_R$ , where the superscript  $j$  indicates the treatment group ( $j = 0$  for the control group and  $j = 1$  for the treatment group). Let  $\lambda_R^{(j)}(t)$  denote the hazard function of  $T_R^{(j)}$  ( $j = 0, 1$ ). The null hypothesis of no effect is given by  $H_0: \text{HR}_R(t) = \lambda_R^{(1)}(t)/\lambda_R^{(0)}(t) = 1$  and the alternative that the new treatment improves survival by  $H_a: \text{HR}_R(t) < 1$ . The logrank test  $Z$  is used to test that the new treatment improves survival.

Assume now that an additional endpoint  $\mathcal{E}_A$  is considered as component of the primary endpoint and the composite endpoint  $\mathcal{E}_* = \mathcal{E}_R \cup \mathcal{E}_A$  is to be used, instead, to prove the efficacy of the new treatment. The effect of treatment would then be evaluated on the time  $T_*^{(j)}$  to  $\mathcal{E}_*$  where  $T_*^{(j)} = \min\{T_R^{(j)}, T_A^{(j)}\}$  and  $T_A^{(j)}$  stands for the time to  $\mathcal{E}_A$  ( $j = 0, 1$ ). Let  $\lambda_A^{(j)}(t)$  and  $\lambda_*^{(j)}(t)$  denote, respectively, the hazard functions of  $T_A^{(j)}$  and  $T_*^{(j)}$  ( $j = 0, 1$ ). The treatment effect on  $\mathcal{E}_*$  would then be tested with the logrank test  $Z_*$  to compare  $H_0^*: \text{HR}_*(t) = \lambda_*^{(1)}(t)/\lambda_*^{(0)}(t) = 1$  versus  $H_a^*: \text{HR}_*(t) < 1$ .

We assume that the additional endpoint does not include a terminating event, which corresponds to Case 1 when neither the relevant nor the additional endpoint includes a terminating event, and Case 3, when the relevant endpoint includes a terminating event.

Schoenfeld (1981)<sup>65</sup> studies the asymptotic behaviour of the logrank statistic and proves that under the null hypothesis of no treatment difference, the logrank is asymptotically  $N(0, 1)$  and, under a sequence of alternatives contiguous to the null, the logrank is asymptotically normal with unit variance and finite mean. Gómez and Lagakos apply Schoenfeld's results and proceed as follows. They consider  $\lambda_R^{(0)}(t)$  as fixed and define a sequence of alternatives  $H_{a,n}$  consisting of instantaneous hazard functions close enough to  $\lambda_R^{(0)}(t)$ , for instance taking  $\lambda_{R,n}^{(1)}(t) = \lambda_R^{(0)}(t)e^{g(t)/\sqrt{n}}$  for some  $g(t)$  function. These sequence of alternatives, formulated equivalently as  $\text{HR}_{R,n}(t) = e^{g(t)/\sqrt{n}}$ , include proportional hazard alternatives, i.e, taking  $g(t) = \beta$  for a fixed real value  $\beta$ . Logrank  $Z$  is asymptotically  $N(0, 1)$  under the null hypothesis of no treatment difference ( $H_0: \text{HR}_R(t) = 1$ ) and asymptotically normal with unit variance and mean  $\mu$  given in equation (3.1) under the sequence of alternatives  $H_{a,n}: \text{HR}_{R,n}(t) = e^{g(t)/\sqrt{n}} < 1$ . Analogously, fix  $\lambda_*^{(0)}(t)$  and define  $H_0^*: \text{HR}_*(t) = 1$  and the sequence of alternatives  $H_{a,n}^*: \text{HR}_{*,n}(t) = e^{g_*(t)/\sqrt{n}} < 1$  for a given function  $g_*(t)$ . It follows that  $Z^*$  is asymptotically  $N(0, 1)$  under  $H_0^*$ , and asymptotically normal with unit variance and mean  $\mu_*$  given in equation (3.2) under the sequence  $H_{a,n}^*$ . The asymptotic means of  $Z$  and  $Z^*$  are given by

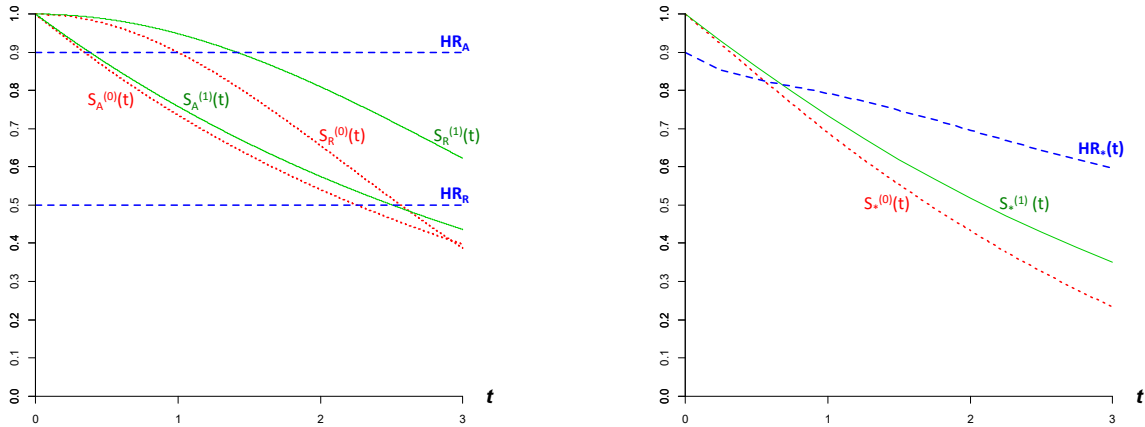
$$\mu = \frac{\int_0^\infty g(t)p(t)[1-p(t)]\Pr_{H_0}\{U \geq t\}\lambda_R^{(0)}(t)dt}{\sqrt{\int_0^\infty p(t)[1-p(t)]\Pr_{H_0}\{U \geq t\}\lambda_R^{(0)}(t)dt}}, \quad (3.1)$$

$$\mu_* = \frac{\int_0^\infty g_*(t)p_*(t)[1-p_*(t)]\Pr_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)dt}{\sqrt{\int_0^\infty p_*(t)[1-p_*(t)]\Pr_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)dt}}, \quad (3.2)$$

where  $U = \min\{T_R, \tau\}$  (in Cases 1 and 3) and  $U_* = \min\{T_*, \tau\}$  denote the observed outcome;  $\tau$  denotes the censoring time;  $p(t) = \Pr_{H_0}\{X = 1|U \geq t\}$  and  $p_*(t) = \Pr_{H_0^*}\{X = 1|U_* \geq t\}$  are the null probabilities that someone at risk at time  $t$  is in treatment group 1;  $\Pr_{H_0}\{U \geq t\}$  and  $\Pr_{H_0^*}\{U_* \geq t\}$  are the null probabilities that someone is still at risk at time  $t$  and  $\Pr_{H_0}\{U \geq t\}\lambda_R^{(0)}(t)$  and  $\Pr_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)$  correspond to the probabilities, under the null hypothesis, of observing events  $\mathcal{E}_R$  and  $\mathcal{E}_*$ , respectively, by time  $t$ .

### 3.1.2 Asymptotic Relative Efficiency

Efficiency calculations throughout this chapter will assume that end-of-study censoring at time  $\tau$  ( $\tau = 1$  without loss of generality) is the only non-informative censoring cause for both groups; this assumption indirectly implies that the censoring mechanism is the same for both groups. It is as well assumed that the hazard functions  $\lambda_R^{(j)}(t)$  and  $\lambda_A^{(j)}(t)$  ( $j = 0, 1$ ) are proportional, that is,  $\text{HR}_R(t) = \text{HR}_R$  and  $\text{HR}_A(t) = \text{HR}_A$ , for all  $t$ , where  $\text{HR}_R(t) = \lambda_R^{(1)}(t)/\lambda_R^{(0)}(t)$  and  $\text{HR}_A(t) = \lambda_A^{(1)}(t)/\lambda_A^{(0)}(t)$  are the hazard ratios between  $T_R^{(0)}$  and  $T_R^{(1)}$  and between  $T_A^{(0)}$  and  $T_A^{(1)}$ , respectively. Note that although we are assuming that the hazard functions  $\lambda_R^{(j)}(t)$  and  $\lambda_A^{(j)}(t)$  ( $j = 0, 1$ ) are proportional, this does not imply the proportionality of hazards  $\lambda_*^{(0)}(t)$  and  $\lambda_*^{(1)}(t)$  for the composite endpoint  $T_*$  (see Figure 3.1).



**Figure 3.1:** Survival and hazard ratio for the relevant endpoint (RE),  $T_R$ , for the additional endpoint (AE),  $T_A$  and for the composite endpoint (CE),  $T_* = \min\{T_R, T_A\}$ .  $T_R \sim$  Weibull with shape parameter  $\beta_R = 2$  (increasing hazard) for treatment groups 0 and 1 and  $T_A \sim$  Weibull with shape parameter  $\beta_A = 1$  (constant hazard) for treatment groups 0 and 1. Scale parameters for  $T_R$  and  $T_A$  have been calculated such that  $\Pr\{T_R \text{ observed in group } 0\} = 0.1$ ,  $\Pr\{T_A \text{ observed in group } 0\} = 0.25$ ,  $\text{HR}_R = 0.5$ ,  $\text{HR}_A = 0.9$  and Spearman's  $\rho(T_R, T_A) = 0.45$  assuming Frank's copula between  $T_R$  and  $T_A$ . Considering the RE as a terminating event (case 3), in this setting  $\text{ARE}(Z_*, Z) = 0.21$ .

To assess the difference in efficiency between using logrank test  $Z$ , based on the relevant endpoint  $\mathcal{E}_R$ , and logrank test  $Z_*$ , based on the composite endpoint  $\mathcal{E}_*$ , Gómez and Lagakos<sup>27</sup> base their strategy on the behaviour of the asymptotic relative efficiency (ARE) of  $Z_*$  versus  $Z$ . The ARE is a measure of the relative power of two tests that can be interpreted, when the two tests are for the same null and alternative hypothesis, as the ratio of the required sample sizes to detect a specific treatment effect

to attain the same power for a given significance level (Lehmann and Romano, 2005).<sup>41</sup> In this case, a value of  $ARE = 0.6$  would mean that we only need 60% as many cases to reach a given power if we use  $\mathcal{E}_R$  as we would need if we used  $\mathcal{E}_*$ . Whenever the tests under consideration,  $Z$  and  $Z_*$ , are asymptotically  $N(0,1)$  under  $H_0$  and  $H_0^*$ , respectively, and asymptotically normal with variance 1 under a sequence of contiguous alternatives to the null hypothesis, a different definition for Pitman's relative efficiency as the square of the ratio of the non-centrality parameters  $\mu$  and  $\mu_*$  is appropriate

$$ARE(Z_*, Z) = \left( \frac{\mu_*}{\mu} \right)^2, \quad (3.3)$$

where  $\mu$  and  $\mu_*$  are to be replaced by expressions (3.1) and (3.2).

Before providing the expression that is being used to evaluate the  $ARE$ , and for the sake of clarity, we enumerate the assumptions that have been taken into account:

- End-of-study censoring at time  $\tau$  is the only non-informative censoring cause for both groups.
- The additional endpoint does not include a terminating event.
- The hazard ratios between  $T_R^{(0)}$  and  $T_R^{(1)}$  and between  $T_A^{(0)}$  and  $T_A^{(1)}$  are proportional, that is,  $HR_R(t) = \lambda_R^{(1)}(t)/\lambda_R^{(0)}(t) = HR_R$  and  $HR_A(t) = \lambda_A^{(1)}(t)/\lambda_A^{(0)}(t) = HR_A$  for all  $t$ .
- Effect of treatment on  $\mathcal{E}_R$  is tested establishing  $H_0: HR_R = 1$  versus a sequence of alternatives  $H_{a,n}: \lambda_{R,n}^{(1)}(t) = \lambda_R^{(0)}(t)e^{g(t)/\sqrt{n}}$  for some  $g(t)$  function. Note that  $g(t)/\sqrt{n} = \log\{\lambda_{R,n}^{(1)}(t)/\lambda_R^{(0)}(t)\}$ .
- Effect of treatment on  $\mathcal{E}_*$  is tested establishing  $H_0^*: HR_*(t) = 1$  versus a sequence of alternatives  $H_{a,n}^*: HR_{*,n}(t) = e^{g_*(t)/\sqrt{n}} < 1$  for a given function  $g_*(t)$ . Note that  $g_*(t)/\sqrt{n} = \log\{HR_{*,n}(t)\}$ .

Under the above assumptions, expression (3.3) becomes

$$ARE(Z_*, Z) = \frac{\left( \int_0^1 \log\{\lambda_*^{(1)}(t)/\lambda_*^{(0)}(t)\} f_*^{(0)}(t) dt \right)^2}{(\log\{HR_R\})^2 \left( \int_0^1 f_*^{(0)}(t) dt \right) \left( \int_0^1 f_R^{(0)}(t) dt \right)}, \quad (3.4)$$

where  $f_R^{(0)}(t)$  and  $f_*^{(0)}(t)$  are the density functions of  $T_R^{(0)}$  and  $T_*^{(0)}$ , respectively.

**Remark:** The density function  $f_*^{(0)}(t)$  is the density of the  $T_*^{(0)} = \min\{T_R^{(0)}, T_A^{(0)}\}$ , computed from the joint density between  $T_R^{(0)}$  and  $T_A^{(0)}$ , which itself is built from the marginals of  $T_R^{(0)}$  and  $T_A^{(0)}$  by means of a bivariate copula.

### 3.2 Relationship between ARE and sample sizes

We start establishing that if the hazard ratios for  $T_R^{(j)}$  ( $j = 0, 1$ ) and for  $T_A^{(j)}$  ( $j = 0, 1$ ) approach the unity as  $n$  gets large, so does the hazard ratio of the minimum  $T_*^{(j)}$  between  $T_R^{(j)}$  and  $T_A^{(j)}$  ( $j = 0, 1$ ).

**Lemma 1** *Given two sequences of hazard ratios  $\{\text{HR}_{R,n}(t) = \lambda_{R,n}^{(1)}(t)/\lambda_R^{(0)}(t)\}$  and  $\{\text{HR}_{A,n}(t) = \lambda_{A,n}^{(1)}(t)/\lambda_A^{(0)}(t)\}$ , both converging uniformly to 1 as  $n \rightarrow \infty$ , the sequence corresponding to the hazard ratio of  $T_*^{(j)} = \min\{T_R^{(j)}, T_A^{(j)}\}$ , namely  $\{\text{HR}_{*,n}(t) = \lambda_{*,n}^{(1)}(t)/\lambda_*^{(0)}(t)\}$ , tends to 1 as  $n \rightarrow \infty$ . In particular, this lemma holds whenever  $\log(\lambda_{k,n}^{(1)}(t)/\lambda_k^{(0)}(t)) = O(n^{-1/2})$ , which in turn, is true if  $\log(\lambda_{k,n}^{(1)}(t)/\lambda_k^{(0)}(t)) = g_k(t)/\sqrt{n}$ , for any bounded real function  $g_k(t)$  ( $k = R, A$ ).*

**Proof 1** *It follows immediately that for fixed  $t$ ,  $\lim_{n \rightarrow \infty} \lambda_{R,n}^{(1)}(t) = \lambda_R^{(0)}(t)$  and  $\lim_{n \rightarrow \infty} \lambda_{A,n}^{(1)}(t) = \lambda_A^{(0)}(t)$ . Furthermore, it follows that the corresponding densities and survival functions  $f_{R,n}^{(1)}(t)$ ,  $f_{A,n}^{(1)}(t)$ ,  $S_{R,n}^{(1)}(t)$  and  $S_{A,n}^{(1)}(t)$ , converge to  $f_R^{(0)}(t)$ ,  $f_A^{(0)}(t)$ ,  $S_R^{(0)}(t)$  and  $S_A^{(0)}(t)$ , respectively. Taking into account that the survival function of the minimum,  $S_{*,n}^{(1)}(t)$  is expressed in terms of the marginal survival functions  $S_{R,n}^{(1)}(t)$  and  $S_{A,n}^{(1)}(t)$  of  $T_R^{(1)}$  and  $T_A^{(1)}$  via a copula  $C$ , that is,*

*$S_{*,n}^{(1)}(t) = C(S_{R,n}^{(1)}(t), S_{A,n}^{(1)}(t))$ , it remains to prove that  $\lim_{n \rightarrow \infty} S_{*,n}^{(1)}(t) = S_*^{(0)}(t)$ . This result will imply that*

*$\lim_{n \rightarrow \infty} f_{*,n}^{(1)}(t) = f_*^{(0)}(t)$ ,  $\lim_{n \rightarrow \infty} \lambda_{*,n}^{(1)}(t) = \lambda_*^{(0)}(t)$  and hence the sequence  $\text{HR}_{*,n}(t) \rightarrow 1$  as  $n \rightarrow \infty$ , as we wanted to prove.*

*The convergence of  $S_{*,n}^{(1)}(t)$  to  $S_*^{(0)}(t)$  is guaranteed by the convergence of  $S_{R,n}^{(1)}(t)$  and  $S_{A,n}^{(1)}(t)$  to  $S_R^{(0)}(t)$  and  $S_A^{(0)}(t)$ , respectively, together with the fact that bivariate copulas  $C$  are bivariate distribution functions with uniform marginals. The reader is addressed to Lindner and Szimayer (2005)<sup>42</sup> for the corresponding technical proofs.*

**Proposition 1** *Consider two test procedures  $\phi_n$  and  $\phi_n^*$  to test  $H_0 : \text{HR}_R(t) = 1$  against  $H_{a,n} : \text{HR}_{R,n}(t) < 1$  and  $H_0^* : \text{HR}_*(t) = 1$  against  $H_{a,n}^* : \text{HR}_{*,n}(t) < 1$ , respectively. Let  $n$  and  $n_*$  be the sample sizes required for  $\phi_n$  and  $\phi_n^*$ , respectively, to have power at least  $\Pi$  at level  $\alpha$ . Assume the sequences  $\phi = \{\phi_n\}$  and  $\phi^* = \{\phi_n^*\}$  are based on the logrank statistics  $Z$  and  $Z^*$ , respectively, converging, to Normal  $(\mu, 1)$  and Normal  $(\mu_*, 1)$  with  $\mu$  and  $\mu_*$  given in (3.1) and (3.2), under sequences of local alternatives  $\text{HR}_{k,n}(t)$  ( $k = R, A$ ) converging uniformly to 1 as  $n \rightarrow \infty$ . Given  $0 < \alpha < \Pi < 1$ ,*

$$\lim_{\substack{\text{HR}_{R,n}(t) \rightarrow 1 \\ \text{HR}_{A,n}(t) \rightarrow 1}} \frac{n}{n_*} = \text{ARE}(Z_*, Z).$$

*The usual interpretation of the ARE as the reciprocal ratio of the sample sizes holds even when two different sets of hypotheses ( $H_0$  versus  $H_{a,n}$  and  $H_0^*$  versus  $H_{a,n}^*$ ) are tested. As a consequence of this proposition, the interpretation of the ARE is the following. If  $\text{ARE}(Z_*, Z) = 0.7$ , then, asymptotically, we only need 70% as many cases to attain a given power if we use  $Z$  as we would need if we used  $Z_*$ .*

**Proof 2** *By Lemma 1, uniform convergence to 1 of  $\{\text{HR}_{R,n}(t)\}$  and  $\{\text{HR}_{A,n}(t)\}$  imply that  $\lim \text{HR}_{*,n}(t) \rightarrow 1$ . Under the sequence of contiguous alternatives to the null,  $H_{a,n} : \{\text{HR}_{R,n}(t) = \lambda_{R,n}^{(1)}(t)/\lambda_R^{(0)}(t)\} \rightarrow 1$*

and  $H_{a,n}^* : \{\text{HR}_{*,n}(t) = \lambda_{*,n}^{(1)}(t)/\lambda_*^{(0)}(t)\} \rightarrow 1$ , both  $Z$  and  $Z^*$  are asymptotically  $N(\mu, 1)$  and  $N(\mu_*, 1)$ , respectively. The power function for a one-sided test with size  $\alpha$  is therefore given, respectively, by

$$\begin{aligned}\Pi_R &= \lim_{n \rightarrow \infty} \text{Prob}\{Z < z_{1-\alpha} | H_{a,n}\} = 1 - \Phi(-z_{1-\alpha} + \mu) \\ \Pi_* &= \lim_{n \rightarrow \infty} \text{Prob}\{Z_* < z_{1-\alpha} | H_{a,n}^*\} = 1 - \Phi(-z_{1-\alpha} + \mu_*)\end{aligned}\quad (3.5)$$

where  $\Phi$  is the distribution function of the standard normal and  $z_{1-\alpha}$  is the standard normal quantile corresponding to the left tail probability  $\alpha$ . It immediately follows that  $\Pi_R = \Pi_*$  is equivalent to  $\mu = \mu_*$ , given by (3.1) and (3.2). Equivalently

$$\left(\frac{\mu_*}{\mu}\right)^2 = 1 \iff \left(\frac{\frac{\int_0^\infty g(t)p(t)[1-p(t)]\text{Pr}_{H_0}\{U \geq t\}\lambda_R^{(0)}(t)dt}{\sqrt{\int_0^\infty p(t)[1-p(t)]\text{Pr}_{H_0}\{U \geq t\}\lambda_R^{(0)}(t)dt}}}{\frac{\int_0^\infty g_*(t)p_*(t)[1-p_*(t)]\text{Pr}_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)dt}{\sqrt{\int_0^\infty p_*(t)[1-p_*(t)]\text{Pr}_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)dt}}}\right)^2 = 1. \quad (3.6)$$

Since

$$p(t) = \frac{\text{Pr}_{H_0}\{U \geq t | X = 1\}\pi}{\text{Pr}_{H_0}\{U \geq t\}} = \frac{\text{Pr}_{H_0}\{U^{(j)} \geq t\}\pi}{\text{Pr}_{H_0}\{U \geq t\}}$$

where  $\pi = \text{Pr}_{H_0}\{X = 1\}$ , we have

$$p(t)(1-p(t))\text{Pr}_{H_0}\{U \geq t\} = \frac{\text{Pr}_{H_0}\{U^{(1)} \geq t\}\pi\text{Pr}_{H_0}\{U^{(0)} \geq t\}(1-\pi)}{\text{Pr}_{H_0}\{U^{(0)} \geq t\}(1-\pi) + \text{Pr}_{H_0}\{U^{(1)} \geq t\}\pi}.$$

Based in the stated assumptions, because  $T_R^{(j)}$  is right-censored by the end-of-study at time  $\tau$ , and under the null hypothesis of no effect ( $S_R^{(0)}(t) = S_R^{(1)}(t)$ ), we have  $\text{Pr}_{H_0}\{U^{(j)} \geq t\} = S_R^{(0)}(t)\mathbf{1}\{[0, 1]\}(t)$ , for  $j = 0, 1$ . Replacing in (3.1), the noncentrality parameter  $\mu$  becomes

$$\mu = \frac{\sqrt{\pi(1-\pi)} \int_0^1 g(t)S_R^{(0)}(t)\lambda_R^{(0)}(t)dt}{\sqrt{\int_0^1 S_R^{(0)}(t)\lambda_R^{(0)}(t)dt}} = \frac{\sqrt{\pi(1-\pi)} \int_0^1 g(t)f_R^{(0)}(t)dt}{\sqrt{\int_0^1 f_R^{(0)}(t)dt}}$$

where  $f_R^{(0)}(t)$  is the marginal density function for  $T_R^{(0)}$ . Analogously, it can be seen that

$$\mu_* = \frac{\sqrt{\pi(1-\pi)} \int_0^1 g_*(t)f_*^{(0)}(t)dt}{\sqrt{\int_0^1 f_*^{(0)}(t)dt}}$$

where  $f_*^{(0)}(t)$  is the density function for  $T_*^{(0)}$ . The reader is addressed to the online supporting material of Gómez and Lagakos paper for other technical details.

If we would replace  $g(t)$  and  $g_*(t)$  by  $\sqrt{n} \log\left(\frac{\lambda_{R,n}^{(1)}(t)}{\lambda_R^{(0)}(t)}\right) = \sqrt{n} \log(\text{HR}_R)$  and  $\sqrt{n_*} \log\left(\frac{\lambda_{*,n}^{(1)}(t)}{\lambda_*^{(0)}(t)}\right)$ , respectively, equality (3.6), after cancelling  $\pi(1-\pi)$ , becomes equal to

$$\lim_{\substack{\text{HR}_{R,n}(t) \rightarrow 1 \\ \text{HR}_{A,n}(t) \rightarrow 1}} \frac{\sqrt{n_*}}{\sqrt{n}} \frac{\frac{\int_0^1 \log\{\lambda_*^{(1)}(t)/\lambda_*^{(0)}(t)\} f_*^{(0)}(t) dt}{\sqrt{\int_0^1 f_*^{(0)}(t) dt}}}{\log(\text{HR}_R) \sqrt{\int_0^1 f_R^{(0)}(t) dt}} = 1$$

which in turn is equivalent to

$$\lim_{\substack{\text{HR}_{R,n}(t) \rightarrow 1 \\ \text{HR}_{A,n}(t) \rightarrow 1}} \frac{n}{n_*} = \frac{\left( \int_0^1 \log \{ \lambda_*^{(1)}(t) / \lambda_*^{(0)}(t) \} f_*^{(0)}(t) dt \right)^2}{(\log(\text{HR}_R))^2 \left( \int_0^1 f_*^{(0)}(t) dt \right) \left( \int_0^1 f_R^{(0)}(t) dt \right)} \quad (3.7)$$

and it follows that  $\text{ARE}(Z_*, Z) = \lim_{\substack{\text{HR}_{R,n}(t) \rightarrow 1 \\ \text{HR}_{A,n}(t) \rightarrow 1}} \frac{n}{n_*}$ , as we wanted to prove.

Note that (3.7) implies

$$\frac{\left( \int_0^1 \log \{ \lambda_*^{(1)}(t) / \lambda_*^{(0)}(t) \} f_*^{(0)}(t) dt \right)^2}{(\log(\text{HR}_R))^2 \left( \int_0^1 f_*^{(0)}(t) dt \right)^2} = \lim_{\substack{\text{HR}_{R,n}(t) \rightarrow 1 \\ \text{HR}_{A,n}(t) \rightarrow 1}} \frac{n \left( \int_0^1 f_R^{(0)}(t) dt \right)}{n_* \left( \int_0^1 f_*^{(0)}(t) dt \right)} \approx \frac{\text{expected number } \mathcal{E}_R}{\text{expected number } \mathcal{E}_*}$$

and whenever  $\lambda_*^{(1)}(t) / \lambda_*^{(0)}(t)$  is approximately constant and equal to  $\text{HR}_*$ , we would have

$$\frac{\left( \frac{1}{\log(\text{HR}_R)} \right)^2}{\left( \frac{1}{\log(\text{HR}_*)} \right)^2} = \lim_{\substack{\text{HR}_{R,n}(t) \rightarrow 1 \\ \text{HR}_{A,n}(t) \rightarrow 1}} \frac{n \left( \int_0^1 f_R^{(0)}(t) dt \right)}{n_* \left( \int_0^1 f_*^{(0)}(t) dt \right)} \approx \frac{\text{expected number } \mathcal{E}_R}{\text{expected number } \mathcal{E}_*}.$$



### 3.3 Simulation

#### 3.3.1 Simulation

Our next aim is to simulate data to empirically check how close we are to the limiting relationship  $n/n_* = ARE(Z_*, Z)$  when  $\Pi_R = \Pi_*$  for different finite sample sizes. To conduct the simulations we will assume, as Gómez and Lagakos did, that  $T_R^{(j)}$  and  $T_A^{(j)}$  follow Weibull distributions. Weibull distributions are chosen for their wide use in the field of survival analysis due to its flexibility, allowing decreasing, constant and increasing hazard rates. The corresponding shape and scale parameters are denoted by  $\beta_k$  and  $b_k^{(j)}$  ( $j = 0, 1, k = R, A$ ) (shape parameters for both groups are taken equal so that the assumption of the proportionality of the hazard ratios holds). To establish the bivariate distribution of  $(T_R^{(j)}, T_A^{(j)})$ , we consider Frank's Archimedean survival copula, again as Gómez and Lagakos did. Other choices of copulas would be possible, although main conclusions and recommendations will not differ (Plana-Ripoll and Gómez, 2015).<sup>51</sup> Frank's copula depends on the association parameter  $\theta^{(j)}$  between  $T_R^{(j)}$  and  $T_A^{(j)}$ . We assume equal association parameter  $\theta = \theta^{(0)} = \theta^{(1)}$  for both groups 0 and 1, which is biunivocally related to Spearman's rank correlation  $\rho$ . Different scenarios will be simulated according to several choices of  $(\beta_R, \beta_A, p_R^{(0)}, p_A^{(0)}, HR_R, HR_A, \rho)$ , where  $p_R^{(0)}$  and  $p_A^{(0)}$  are the probability of observing events  $\mathcal{E}_R$  and  $\mathcal{E}_A$ , respectively, for treatment group 0, and  $HR_R, HR_A$  are the relative treatment hazard ratios of  $T_k^{(1)}$  versus  $T_k^{(0)}$  ( $k = R, A$ , respectively).

Given a set of values for  $(\beta_R, \beta_A, p_R^{(0)}, p_A^{(0)}, HR_R, HR_A, \rho)$ , for a given power  $\Pi$  and a significance level  $\alpha$ , the simulation steps are the following:

1. **Computations for the relevant endpoint  $\mathcal{E}_R$ .** The scale parameters  $b_R^{(0)}$  and  $b_R^{(1)}$  and the probability  $p_R^{(1)}$  of observing the relevant endpoint in group 1 are derived as:

$$\begin{aligned} b_R^{(0)} &= \frac{1}{(-\log(1-p_R^{(0)}))^{1/\beta_R}} \\ b_R^{(1)} &= \frac{b_R^{(0)}}{HR_R^{(1/\beta_R)}} \\ p_R^{(1)} &= 1 - e^{-(1/b_R^{(1)})^{\beta_R}} \end{aligned}$$

2. **Computations for the additional endpoint  $\mathcal{E}_A$ .** The scale parameter  $b_A^{(0)}$  is derived as:

$$b_A^{(0)} = \begin{cases} \frac{1}{(-\log(1-p_A^{(0)}))^{1/\beta_A}} & \text{for Case 1} \\ * & \text{for Case 3} \end{cases}$$

\* For Case 3,  $b_A^{(0)}$  is found as the solution of equation  $p_A^{(0)} = \int_0^1 \int_v^\infty f_{(1,2)}^{(0)}(u, v; \rho) du dv$ , where  $f_{(1,2)}^{(0)}(\cdot, \cdot; \rho)$  is the joint density between  $T_R^{(0)}$  and  $T_A^{(0)}$ .

### 3. Computation of sample sizes $n$ and $n_*$

a) Compute  $n$  (per group) following Freedman (1982)<sup>18</sup> formulas as follows

$$n = \frac{E}{p_R^{(0)} + p_R^{(1)}}, \quad (3.8)$$

where

$$E = \frac{(HR_R + 1)^2 (z_{1-\alpha} + z_{\Pi})^2}{(HR_R - 1)^2}. \quad (3.9)$$

b) Compute  $ARE(Z_*, Z)$  based on  $(\beta_R, \beta_A, p_R^{(0)}, p_A^{(0)}, HR_R, HR_A, \rho)$ .

c) Compute  $n_* = n / ARE(Z_*, Z)$ .

d) Compute  $N = \max\{n, n_*\}$ .

### 4. Simulation of $T_R^{(0)}, T_R^{(1)}, T_A^{(0)}, T_A^{(1)}, T_*^{(0)}, T_*^{(1)}$

Simulate 1000 samples of size  $N$  for the 4 endpoints  $T_k^{(j)}$  from Weibull  $(b_k^{(j)}, \beta_k)$  ( $j = 0, 1, k = R, A$ ). Compute  $T_*^{(j)} = \min\{T_R^{(j)}, T_A^{(j)}\}$ .

### 5. Computation of empirical powers $\hat{\Pi}_R$ and $\hat{\Pi}_*$

For each sample of size  $n$  ( $n_*$ ), compute the logrank statistic  $Z$  ( $Z_*$ ) to compare the treatment effect between  $T_R^{(0)}$  and  $T_R^{(1)}$  ( $T_*^{(0)}$  and  $T_*^{(1)}$ ). For a given significance level  $\alpha$ , the rejection region comprises all observed  $Z$  ( $Z_*$ ) such that  $Z < z_{1-\alpha}$  ( $Z_* < z_{1-\alpha}$ ) where  $z_{1-\alpha}$  is the standard normal quantile corresponding to the left tail probability  $\alpha$ . The empirical powers, denoted by  $\hat{\Pi}_R$  ( $\hat{\Pi}_*$ ), are calculated as the proportion of samples for which  $Z < z_{1-\alpha}$  ( $Z_* < z_{1-\alpha}$ ).

We note here that whenever  $n_* < n$ , we only use, for each sample, the first  $n_*$  simulated values to compute  $\hat{\Pi}_*$ , while when  $n < n_*$ , we only use the first  $n$  simulated values to compute  $\hat{\Pi}_R$ .

### 6. Comparison between $\hat{\Pi}_R$ and $\hat{\Pi}_*$

For each scenario  $(\beta_R, \beta_A, p_R^{(0)}, p_A^{(0)}, HR_R, HR_A, \rho)$ , we compare the differences between the two empirical powers  $\hat{\Pi}_R$  and  $\hat{\Pi}_*$  obtained from the 1000 simulations.

### 3.3.2 Results

We set  $\Pi = 0.9$  and  $\alpha = 0.05$  (other values would not provide additional information). We choose meaningful values for  $(\beta_R, \beta_A, p_R^{(0)}, p_A^{(0)}, HR_R, HR_A, \rho)$ , based on those arising in cardiovascular clinical trials (Gómez, Gómez-Mateu, Dafni, 2014)<sup>26</sup> (see Table 3.1). We restrict our simulation study to 624 scenarios corresponding to  $ARE(Z_*, Z) \leq 10$  and sample sizes smaller than 1100 patients per group. These scenarios yield  $ARE(Z_*, Z)$  values between 0.20 and 9.93; sample sizes,  $n$ , for the relevant endpoint between 142 and 1081; and,  $n_*$ , for the composite endpoint between 53 and 1077 (see Table 3.2).

Parameters						
$\beta_R = \beta_A$ $(p_R^{(0)}, p_A^{(0)})$	0.5	1	2			
	(0.05, 0.01) (0.15, 0.01)	(0.05, 0.15) (0.15, 0.15)	(0.05, 0.35) (0.15, 0.35)	(0.1, 0.01) (0.35, 0.01)	(0.1, 0.15) (0.35, 0.15)	(0.1, 0.35) (0.35, 0.35)
$\rho$ $(HR_R, HR_A)$	0.15	0.45	0.75			
	(0.5, 0.3) (0.7, 0.3)	(0.5, 0.7) (0.7, 0.7)	(0.5, 0.9) (0.7, 0.9)	(0.6, 0.3) (0.8, 0.3)	(0.6, 0.7) (0.8, 0.7)	(0.6, 0.9) (0.6, 0.9)
Total number of cases	624					

**Table 3.1:** Values of parameters  $\beta_R, \beta_A, p_R^{(0)}, p_A^{(0)}, HR_R, HR_A$  and  $\rho$  used for the simulations. There are 624 different configurations, excluding those yielding sample sizes larger than 1100 and  $ARE(Z_*, Z) > 10$ .

	min	median	max
$n$	142	509	1081
$n_*$	53	398	1077
$ARE(Z_*, Z)$	0.2	1.04	9.93

**Table 3.2:** Computed values of  $n, n_*$  and  $ARE(Z_*, Z)$  in step 3 of the simulation based on the parameter values given in Table 3.1.

The empirical powers  $\hat{\Pi}_R$  in our simulation study resulted in powers between 0.87 and 0.94, with a median of 0.91. A slightly higher median was found for scenarios with low hazard ratios. This finding is acknowledged as well by Freedman (1982)<sup>18</sup>.

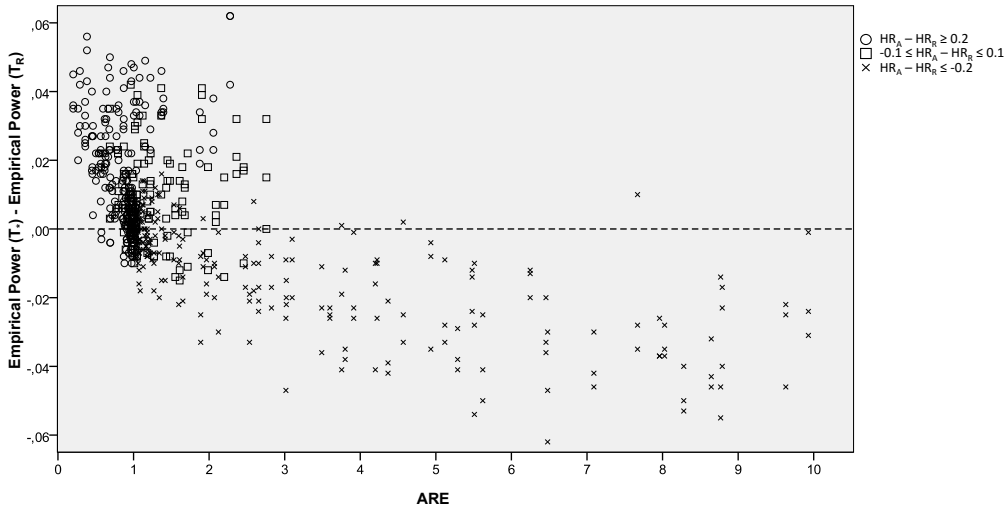
Table 3.3 provides the percentiles for the absolute value differences between  $\hat{\Pi}_*$  and  $\hat{\Pi}_R$ . We observe that in 75% of the cases the difference is smaller than 2.3%, and among cases with  $ARE$  as large as 3 the difference shrinks to 1.9%. There are, however, few instances, where this difference can be as large as 6%, and they deserve a closer look.

Figure 3.2 plots the differences  $\hat{\Pi}_* - \hat{\Pi}_R$  as a function of the  $ARE(Z_*, Z)$  values. The behaviour is remarkably different when  $ARE(Z_*, Z) \leq 3$  or  $ARE(Z_*, Z) > 3$ . Whenever  $ARE(Z_*, Z) \leq 3$ ,  $\hat{\Pi}_*$  fluctuates around  $\hat{\Pi}_R$ , within a range of 4%. However, when  $ARE(Z_*, Z) > 3$ , corresponding mostly to scenarios where treatment has a stronger effect on the additional endpoint than on the relevant endpoint ( $HR_A \leq HR_R - 0.2$ ) and the anticipated number of events in the control group is larger for the additional endpoint than for the relevant ( $p_A^{(0)} \geq p_R^{(0)}$ ), the empirical power  $\hat{\Pi}_*$  of the logrank test based on the CE

	min	$w_{0.1}$	$w_{0.25}$	$w_{0.5}$	$w_{0.75}$	$w_{0.9}$	max
For all ARE	0	0.002	0.004	0.010	0.023	0.036	0.062
$ARE(Z_*, Z) \leq 3$	0	0.002	0.004	0.008	0.019	0.033	0.062
$ARE(Z_*, Z) > 3$	0.001	0.009	0.016	0.026	0.038	0.046	0.062

**Table 3.3:** Percentiles of  $|\hat{\Pi}_* - \hat{\Pi}_R|$  as a function of ARE values, where  $w_i$  indicates the corresponding percentile.

never achieves the same power as the logrank test for the relevant endpoint would get. In these cases the interpretation of the  $ARE(Z_*, Z)$  as the ratio of the sample sizes,  $n/n_*$ , is not as straightforward. Nevertheless, this does not mean that the recommendation of using the CE does not have to be followed since larger values for  $n_*$  needed to attain the same power as  $n$  does, would reduce the ARE value but not as much as to cross the “1” border that would imply to use the relevant endpoint instead of the CE.



**Figure 3.2:** Differences between empirical powers  $\hat{\Pi}_* - \hat{\Pi}_R$  as function of  $ARE(Z_*, Z)$  and in terms of  $HR_A - HR_R$ .

If we analyze the differences between  $\hat{\Pi}_*$  and  $\hat{\Pi}_R$  as a function of the differences between the two hazard ratios ( $HR_A - HR_R$ ), we observe that when the two hazard ratios are very close, the two empirical powers are as well very close. Whenever  $HR_A - HR_R \leq 0.2$ , not only  $ARE(Z_*, Z)$  values tend to be higher, but also  $\hat{\Pi}_* < \hat{\Pi}_R$ . (see Figure 3.2).

Taking into account that absolute differences between powers smaller than 5% could be considered irrelevant, we conclude that the asymptotic relationship  $ARE(Z_*, Z) = n/n_*$  is valid in the majority of scenarios.

All computations in this chapter have been implemented in R and are available under request to either author.

### 3.4 Discussion

Pitman's relative efficiency is defined as the limiting ratio of sample sizes to give the same asymptotic power under sequences of local alternatives. Given two asymptotically standard normal tests  $S_n$  and  $T_m$  under the same null and alternative hypotheses, the alternative definition  $ARE = (\mu_S/\mu_T)^2$  can be used because the equality of the powers holds if  $\frac{m}{n} = (\frac{\mu_S}{\mu_T})^2$ .

Gómez and Lagakos' method uses the alternative definition of  $ARE$  to develop all the computations for the two corresponding logrank tests. Our goal has been to check that the relationship between  $(\mu_S/\mu_T)^2$  and the ratio of sample sizes still held when the two hypotheses under test were not the same ( $H_0$  versus  $H_a$  and  $H_0^*$  versus  $H_a^*$ ).

It is important to keep in mind that these two hypotheses tests are by no means equivalent. For instance, to check whether treatment has a beneficial effect, we might use  $\mathcal{E}_R$  or we might add endpoint  $\mathcal{E}_A$  and use  $\mathcal{E}_*$ . As it is shown in Gómez (2011),<sup>21</sup> even if we assume that the times to  $\mathcal{E}_R$  and to  $\mathcal{E}_A$  are independent, a beneficial effect on  $\mathcal{E}_*$  can occur simultaneously with a beneficial effect on  $\mathcal{E}_R$  and a harmful effect on  $\mathcal{E}_A$  and not finding a beneficial effect on the composite event  $\mathcal{E}_*$  is no guarantee of not having some effect on the individual events  $\mathcal{E}_R$  or  $\mathcal{E}_A$ .

The main result of this paper proves that  $ARE(Z_*, Z)$  coincides with  $n/n_*$ , being  $n$  and  $n_*$  the sample sizes needed to detect specific alternatives  $HR_R$  and  $HR_*$  to attain power  $\Pi$  and for the same significance level  $\alpha$ . Therefore, we can use and interpret  $ARE$  in its usual way.

The simulation study has been conducted in such a way that for fixed values  $n$  and  $ARE(Z_*, Z)$ , the sample size  $n_*$  is calculated as  $n_* = n/ARE(Z_*, Z)$ . Hence, an approximate equality of the empirical powers  $\hat{\Pi}_R$ , of logrank test  $Z$  for  $H_0$  versus  $H_{a,n}$ , and of  $\hat{\Pi}_*$  of logrank test  $Z_*$  for  $H_0^*$  versus  $H_{a,n}^*$ , indicates that the relationship  $ARE(Z_*, Z) = n/n_*$  holds. Main results from our simulations show that the absolute differences between  $\hat{\Pi}_R$  and  $\hat{\Pi}_*$  are most of the times less than 2.5%, hence the usual interpretation between  $(n, n_*)$  and  $ARE(Z_*, Z)$  holds for finite sample sizes.

For those scenarios under which  $ARE(Z_*, Z) > 3$ , we observe that the empirical power of the test based on  $\mathcal{E}_*$  never achieves the empirical power that the logrank test based on  $\mathcal{E}_R$  would get. Consequently, larger values of  $n_*$  would be needed to attain the same power as  $n$  does. In these instances, even though the relationship  $ARE(Z_*, Z) = n/n_*$  is not necessarily true, the recommendation to use the composite endpoint  $\mathcal{E}_*$  instead of the relevant endpoint  $\mathcal{E}_R$  will still be valid because very rarely a value of  $ARE(Z_*, Z) > 3$  would go down to less than 1. However, caution will be needed if one wants to use the relationship  $ARE(Z_*, Z) = n/n_*$  to compute the required sample size  $n_*$  if  $ARE(Z_*, Z) > 3$ . In these cases, a different formulation should be seek.

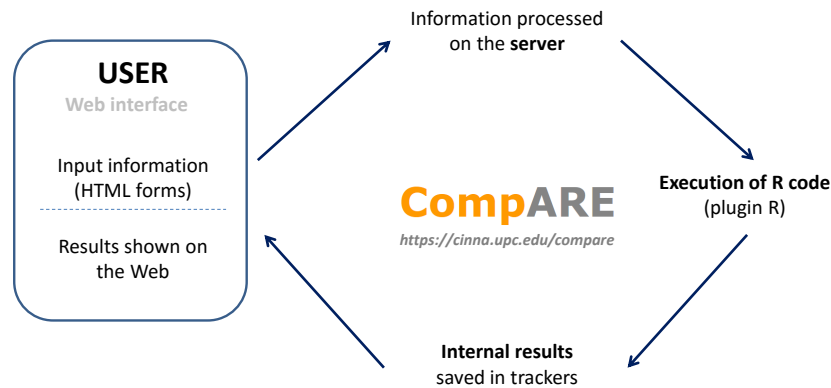


## ***CompARE*. AN ON-LINE PLATFORM AS A DECISION TOOL FOR INVESTIGATORS**

*CompARE* is an online web-based platform that provides efficient measures for discerning between possible time-to-event endpoints when evaluating the efficacy of a treatment. In the design phase of a clinical trial, clinicians, biostatisticians, and other members of the team can use *CompARE* to study the performance of different endpoints in a variety of scenarios. *CompARE* has an intuitive interface and it is a convenient tool for a better informed decision on the primary endpoint.

Users introduce the information needed in *CompARE* by means of intuitive web-page forms, such as the list of candidate endpoints, together with the anticipated parameter values. The information, which is saved in *trackers*, is used to run the code written in *R*, statistical software that is widely used in the field of statistics (see the schematic in Figure 4.1). It is not necessary to install *R* to run *CompARE*, nor is knowledge of *R* required. Furthermore, all the inputs and outputs are presented in HTML format and are compatible with any web browser.

In this chapter we detail the software we used to build *CompARE* and the basic features of this platform, describing step by step how to access, register, run it and get the results. We illustrate one of the capabilities of *CompARE* using the *ARE* method to quantify how efficient is a relevant subset of outcomes with respect to a larger subset. Results from different scenarios, depending on the parameter values, are shown immediately by means of tables and plots such as survival and hazard ratio curves. Moreover, conclusions and recommendations are provided in written form as an aid. A complete online user's guide is also available. *CompARE* is currently accessible as a beta version at the following website: <https://cinna.upc.edu/compare>.



**Figure 4.1**

In *Chapter 6*, we extend *CompARE* to develop different advanced options. Among other features, we detail how to specify different marginal distribution laws, copulas and correlations between times. Sample size calculations based on these parameters are calculated as well. *CompARE* is also extended to accommodate the computation of combined probabilities and combined hazard ratios, based on the marginal components. For non-proportional hazards, *CompARE* visualizes how the hazard ratios of the components depart from constancy by depicting the corresponding values during the follow-up period for different scenarios.



## 4.1 Software used to build *CompARE*

The web-based platform *CompARE* was built using the free software Tightly Integrated Knowledge Infrastructure (Tiki Wiki CMS/Gouppware).<sup>79</sup> Free software is widely used and is commonly developed with volunteer computer programmers, guaranteeing that every user has equal rights of access. It allows any programmer to study the source code, modify it, and share it (FSF).<sup>17</sup> Moreover, online graphic interfaces based on free software leads to collaboration synergies between partners from different areas such as informatics, biology or statistics.

We chose Tiki because the majority of other interface web programs that include the use of *R* routines present problems in the short or medium term (de Pedro and Sánchez, 2010).<sup>11</sup> Some of the reasons might be because either they do not work with free software, they are not updated or they are too complex to be used by most professionals who are not involved in web computing. Moreover, Tiki is safe and updated periodically by their community members, who add new features, fix bugs and patch security holes. It is constantly maintained under the license LGPL (Lesser General Public License). Repositories are used for the version control system.

Other remarkable features of Tiki are the use of standard codes and its flexibility. Web standard codes such as HTML, PHP and javascript are used. It allows to design entry forms on the web through the use of *trackers* that save the data collected in *items*. Due to its flexibility, these parameter values can be used to execute other applications through the use of *plugins* (see Sapir (2010) for more details<sup>63</sup>). By means of the *pluginR*,<sup>71</sup> developed by de Pedro and Sánchez,<sup>11</sup> it is possible to execute the *ARE* method in *R* in *CompARE*.

The Tiki software is installed in a virtual machine at the Universitat Politècnica de Catalunya. It runs under a Linux server Ubuntu 12.04 LTS with a total of 32GB of RAM, 2 Quad Core Processor and 2x1TB of disk space.

Initially, before designing *CompARE*, the code to execute the *ARE* method was written in MAPLE by Gómez G. The code was programmed to run scenarios where both the relevant endpoint (RE) and the additional endpoint (AE) does not include a terminating event (Case 1) or either the relevant or additional endpoint includes a terminating event (Cases 2 and 3). We adapted the code in *R* because it is free software and because of its the great capabilities and flexibility. Moreover, it is compatible with the software used in *CompARE*. We also extended the *R* code to Case 4, where both the RE and the AE includes a terminating event, and it is described in *Chapter 6*.

## 4.2 *CompARE* step by step

### Access and registration

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

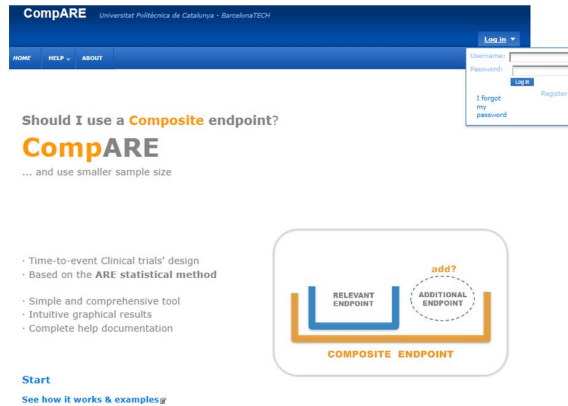


Figure 4.2

Only a quick registration is needed from the main web page. The system asks you for a username and a password, which will be used to enter the application under your own session (see Figure 4.3). For security reasons, an e-mail is required. You will have to accept the registration from your own e-mail. In order to avoid spam registrations, you need to correctly introduce a captcha code (Completely Automated Public Turing test to tell Computers and Humans Apart). The web administrator accepts the registration as a final step.

**Registration**

Register as a new user

Username: \*

Password: \*

Repeat password: \*

Email: \*

*A valid email is mandatory to register*

Institution \*

Country



Enter the code you see above \*  [Try another code](#)

*Fields marked with an \* are mandatory.*

Figure 4.3

**Running *CompARE*. Information about all the candidate endpoints**

Once you are logged in to *CompARE*, you can get to the input grid of information by clicking on the "Start" button from the main web page (see Figure 4.4). Then, enter the information about the endpoints you might plan to use as the primary endpoint in your randomized clinical trial. Place the cursor over each header as an aid for getting a quick definition of each concept:

1. In the first column, write the name of each endpoint. Indicate whether each endpoint is terminating (i.e., when the occurrence of it precludes the observation of other endpoints, as for example Death).
2. Specify the expected probabilities of observing the event in the control group during the follow-up period. By default, when the relevant or the additional endpoints consist of several components, *CompARE* will use their maximum probability to calculate the *ARE* values. In *Chapter 5*, we discuss how to assign the specific value or range of plausible values for this combined probability by means of the marginal information of each component.
3. Indicate the anticipated treatment effect in terms of the hazard ratios between groups (constant hazard ratios are assumed). By default, when the relevant or the additional endpoints consist of several components, *CompARE* will use the average hazard ratio to calculate the *ARE* values. As for the combined probability, in *Chapter 5* we discuss how to assign the specific value or range of plausible values of this combined hazard ratio by means of the marginal information of each component.
4. Select whether the endpoint is a component of the relevant or the additional endpoint.
5. In the last column, select those candidates that form the composite endpoint. Note: unselected endpoints will be excluded from the current analysis.
6. At the bottom, click on "Remove executions history" to delete previous analyses you may have done before.
7. When ready, click on the "Run" button to execute the process.

## Information about all the candidate endpoints for your trial

(You can modify the parameter values and run it again)

Candidate endpoint $E$	Terminating? (click if yes)	Probability of observing $E$ in control group	Hazard Ratio	Type of endpoint	Definition of the Composite
Cardiovascular Death	<input checked="" type="checkbox"/>	0.05	0.7	Relevant component	<input checked="" type="checkbox"/>
Myocardial Infarction	<input type="checkbox"/>	0.1	0.6	Relevant component	<input checked="" type="checkbox"/>
Stroke	<input type="checkbox"/>	0.2	0.6	Additional component	<input checked="" type="checkbox"/>
Hospitalization	<input type="checkbox"/>	0.35	0.8	Additional component	<input checked="" type="checkbox"/>
	<input type="checkbox"/>				<input type="checkbox"/>

Add Rows? 

Remove executions history?

Run

**Figure 4.4**

By default, exponential distributions and the Frank copula relationship between marginal times with moderate correlations is assumed. In *Chapter 6*, among other developments, we extend *CompARE* in order to allow users to modify these assumptions.

As an example, in Figure (4.4) we consider the union of Cardiovascular Death + Myocardial Infarction as the relevant endpoint (RE), and the union of Stroke + Hospitalization as the additional endpoint (AE). The union of the RE and the AE forms the composite endpoint (CE).

## Results

Once the program executes the computations, an output screen that is divided into four tags appears at the top:

- **Results** (see Figure 4.5): In this tag, a table specifies the parameter information set by the user together with the exact value of ARE. When  $ARE > 1$ , the composite endpoint should be used. That is, a smaller sample size would be required, using the composite endpoint as primary rather than the relevant endpoint (Gómez and Gómez-Mateu, 2014)<sup>24</sup>. Otherwise, the relevant endpoint will be advisable to use as primary endpoint. A paragraph below the table shows a detailed recommendation written in text.

Results
Other scenarios
Graphical outputs
Recorded results
No tabs

### Numerical Results

Specific ARE value from your candidates information and Recommendation

Probability RE (Control group)	Probability AE (Control group)	Hazard Ratio RE	Hazard Ratio AE	Distribution RE	Distribution AE	Correlation	ARE	Recommendation
0.07	0.25	0.62	0.7	Increasing Hazard Rate	Constant Hazard Rate (exponential)	0.7	2.1	Use CE

**OUTPUT\_PARAGRAPH**

For a relative frequency of 7 % of observing Cardiovascular Death + Myocardial Infarction (Relevant endpoint) and a relative frequency of 25 % of observing Hospitalization + Stroke (Additional endpoint); a treatment effect given by a Hazard Ratio of 0.62 and 0.7 for the Relevant and the Additional endpoint respectively; and a strong correlation between endpoint times, the Asymptotic Relative Efficiency is higher than 1 ( 2.1 ). Hence, the use of the Composite consisting of Cardiovascular Death + Myocardial Infarction + Hospitalization + Stroke is recommended for the primary endpoint in your randomized clinical trial. That is, for a fixed significance level, the number of required patients you would need to attain the same power is 2.1 times if you use Cardiovascular Death + Myocardial Infarction instead of Cardiovascular Death + Myocardial Infarction + Hospitalization + Stroke .

Note that regardless of the correlation values, the decision about using the composite or not remains the same (see graphical outputs tab).

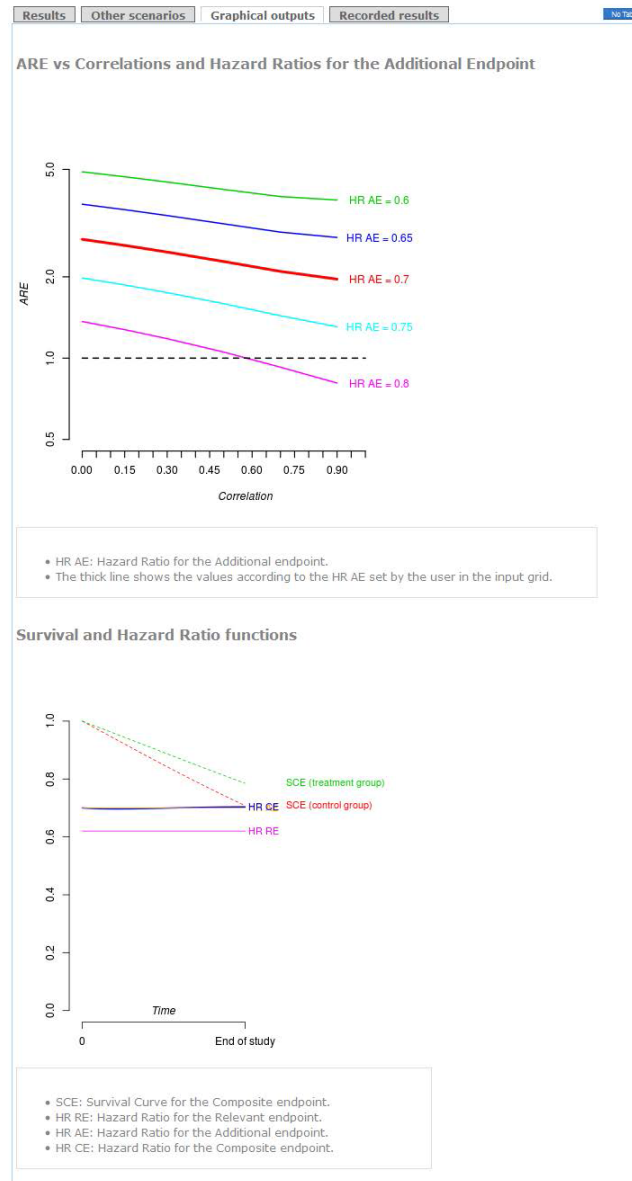
**Figure 4.5**

- **Other scenarios** (see Figure 4.6): Several scenarios depending on different correlations and hazard ratios for the additional endpoint are detailed in a table.

Results		Other scenarios		Graphical outputs		Recorded results		No tabs	
<b>Results</b>									
<b>ARE results depending on different correlation values and Hazard Ratios for the Additional Endpoint:</b>									
(See also graphical outputs)									
Fixed parameters:			Hazard Ratio AE	Correlation	ARE	Recommendation			
Probability RE (Control group)	0.07		0.6	0	4.9	Use CE			
Probability AE (Control group)	0.25		0.6	0.15	4.7	Use CE			
Hazard Ratio RE	0.62		0.6	0.3	4.49	Use CE			
Distribution RE	Increasing Hazard Rate		0.6	0.5	4.21	Use CE			
Distribution AE	Constant Hazard Rate (exponential)		0.6	0.7	3.97	Use CE			
			0.6	0.9	3.85	Use CE			
			0.65	0	3.72	Use CE			
			0.65	0.15	3.55	Use CE			
			0.65	0.3	3.37	Use CE			
			0.65	0.5	3.14	Use CE			
			0.65	0.7	2.93	Use CE			
			0.65	0.9	2.8	Use CE			
			0.75	0.7	1.44	Use CE			
			0.75	0.9	1.31	Use CE			
			0.8	0	1.37	Use CE			
			0.8	0.15	1.27	Use CE			
			0.8	0.3	1.18	Use CE			
			0.8	0.5	1.05	Use CE			
			0.8	0.7	0.93	Use RE			
			0.8	0.9	0.81	Use RE			

**Figure 4.6**

- **Graphical outputs** (see Figure 4.7): The results from the previous tag are shown graphically here. Also, a plot with survival distributions and hazard ratios is shown at the bottom. The "end of study" time point corresponds to the follow-up period of your clinical trial.



**Figure 4.7**

- **Recorded results** (see Figure 4.8): In this tag, the user can see the history of results performed in previous analyses.

Execution	Relevant	endpoint	components	Additional		endpoint	components
1	Cardiovascular Death	Myocardial Infarction		Hospitalization	Stroke		

**Figure 4.8**

In our example, the use of the composite endpoint is clearly advisable, since the *ARE* is higher than 1 ( $ARE = 2.1$ ). That is, by considering moderate correlations between endpoint times, we would need less than half of the sample size when using the CE as primary instead of the RE. As we see in Figure (4.7, top), the decision remains the same irrespective of the correlation. However, note that if the expected hazard ratio for the additional endpoint was 0.8 instead of 0.7 (i.e., a smaller expected effect of the AE), the use of the relevant endpoint would be recommended when having a strong correlation between marginal times.



### 4.3 Summary

*CompARE* provides investigators a free online tool that can be used at the design stage of randomized clinical trials. One of the capabilities of *CompARE* is to help in deciding whether the primary endpoint should consist of a composite of several components, or of a smaller subset that forms the relevant endpoint. In this chapter we have shown some illustrations and the basic functionality of the platform.

*CompARE* is permanently extended with new functionalities and features, and we are constantly improving it thanks to the feedback of colleagues from different national and international universities, institutions and companies. In *Chapter 6* we detail several improvements, such as the possibility of assuming other distribution laws, copulas or correlations. Sample size calculations will also be tackled.

An important issue to address is the assignment of the combined probability and hazard ratio when having several components in the relevant or the additional endpoint, as seen in the example in *Figure 4.4*. By default, *CompARE* assigns the maximum value of the marginal probabilities and the average of the marginal hazard ratios. The assignment of these combined outcomes is of paramount importance, because it might affect the decision regarding the primary endpoint and also regarding sample size calculations. In the next chapter, we develop practical solutions for calculating a specific value or assigning a plausible range of values for the combined probability and combined hazard ratio. We apply this development in *CompARE* and describe it in *Chapter 6*.



## PRACTICAL ISSUES TO ASSIGN COMBINED PROBABILITIES AND HAZARD RATIOS

When the relevant and/or the additional endpoint consist of a composite of several components, investigators may anticipate the probabilities and hazard ratios of each one of the components instead of the combined probabilities and hazard ratios. For example, consider the relevant endpoint as a composite of stroke and myocardial infarction. While the marginal parameter values for stroke and myocardial infarction might be anticipated by trialists (ex. the LIFE trial<sup>10</sup>), the specific parameter values for the union of stroke+myocardial infarction could be more difficult to specify (see Figure 5.1).

Since the computation of the *ARE* needs specific values for the probability (in the control group) and hazard ratio of both the relevant and additional endpoints, we will discuss in this chapter how to derive these quantities based on the marginal values. We propose solutions to assign combined probabilities and combined hazard ratios from any pair of endpoints  $\mathcal{E}_1$  and  $\mathcal{E}_2$ . These two endpoints might refer either to the marginal components of the relevant endpoint or the additional endpoint.

In this chapter, we will use the term combined endpoint instead of composite endpoint in order to avoid confusion with the definition used by Gómez and Lagakos of  $\mathcal{E}_*$  as the union of the relevant endpoint  $\mathcal{E}_R$  and the additional endpoint  $\mathcal{E}_A$ . We restrict to combined endpoints with only two components although some of the results of this chapter could be extended to more than two components. The development of the practical implementation has been incorporated in *CompARE* and is described in chapter 6.

**A letter to the editor in *Statistics in Medicine* is published based on the evaluation of the ARE depending on the choice of different combined parameter values<sup>23</sup> (See Appendix) :**

*Gómez G, Gómez-Mateu M. Comments on "Use of composite endpoints in clinical trials" by Abdul J. Sankoh, Haihong Li and Ralph B. D'Agostino, Sr (2016). *Statistics in Medicine*, 35, 317–318.*

**Letter to the Editor**

---

(wileyonlinelibrary.com) DOI: 10.1002/sim.6483

**Statistics  
in Medicine**

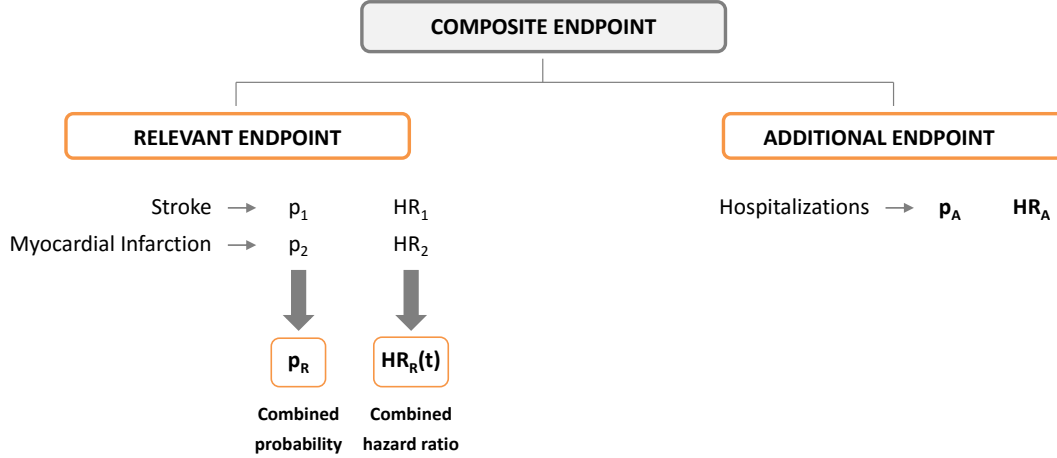
Published online in Wiley Online Library

**Comments on ‘Use of composite endpoints  
in clinical trials’ by Abdul J. Sankoh,  
Haihong Li and Ralph B. D’Agostino, Sr**

*From: Guadalupe Gómez  
Moisés Gómez-Mateu  
Department of Statistics and Operations Research,  
Universitat Politècnica de Catalunya,  
Barcelona, Spain*

## 5.1 Combined probability

In this section we study how to assign the combined probability  $p_*$  of observing the combined endpoint  $\mathcal{E}_* = \mathcal{E}_1 \cup \mathcal{E}_2$  when having two marginal components  $\mathcal{E}_1$  and  $\mathcal{E}_2$ . The value of  $p_*$  depends on the marginal probabilities  $p_1$  and  $p_2$  together with the correlation between  $\mathcal{E}_1$  and  $\mathcal{E}_2$  (Bahadur, 1961).<sup>4</sup> Hence, whenever the correlation is anticipated, the value of  $p_*$  is univocally determined by  $p_1$  and  $p_2$ . Otherwise, a range of plausible values can be proposed.



**Figure 5.1:** Combined probability  $p_R$  and hazard ratio  $HR_R(t)$  of a combined endpoint consisting of the union of Stroke and Myocardial infarction that forms the relevant endpoint.  $p_1, p_2$  and  $HR_1, HR_2$  corresponds to the probability and hazard ratio of each component.  $p_A$  and  $HR_A$  stands for the probability and hazard ratio of the additional endpoint.

### 5.1.1 Boundaries for the combined probability

Consider that we have two endpoint components  $\mathcal{E}_1$  and  $\mathcal{E}_2$  belonging to the combined endpoint  $\mathcal{E}_* = \mathcal{E}_1 \cup \mathcal{E}_2$ . Let  $p_1, p_2$  and  $p_*$  be the probability of observing  $\mathcal{E}_1, \mathcal{E}_2$  and  $\mathcal{E}_*$ , respectively.

*Proposition:* The probability  $p_*$  is bounded by:

$$\max(p_1, p_2) \leq p_* \leq p_1 + p_2. \quad (5.1)$$

*Proof:*

$p_* = P(\mathcal{E}_1 \cup \mathcal{E}_2) = P(\mathcal{E}_1) + P(\mathcal{E}_2) - P(\mathcal{E}_1 \cap \mathcal{E}_2) = p_1 + p_2 - P(\mathcal{E}_1 \cap \mathcal{E}_2)$ . Since  $0 \leq P(\mathcal{E}_1 \cap \mathcal{E}_2) \leq \min(p_1, p_2)$ , we have

$$p_1 + p_2 - \min(p_1, p_2) \leq p_* \leq p_1 + p_2, \text{ which implies} \\ \max(p_1, p_2) \leq p_* \leq p_1 + p_2, \text{ because } p_1 + p_2 - \min(p_1, p_2) = \max(p_1, p_2).$$

In case of independence between  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , that is if  $P(\mathcal{E}_1 \cap \mathcal{E}_2) = P(\mathcal{E}_1) \cdot P(\mathcal{E}_2)$ , we have  $p_* = p_1 + p_2 - p_1 p_2$ .

*Corollary:* Let  $p_1, \dots, p_m$  be the probabilities of observing each component  $\mathcal{E}_1, \dots, \mathcal{E}_m$ , respectively, and define  $p_{*m} = P(\mathcal{E}_1 \cup \dots \cup \mathcal{E}_m)$ . Bounds from (5.1) are straightforwardly generalized as:

$$\max(p_1, p_2, \dots, p_m) \leq p_{*m} \leq p_1 + p_2 + \dots + p_m.$$

### 5.1.2 Expression of the combined probability $p_*$ as a function of the correlation coefficient

Given two Bernoulli random variables  $X_1$  and  $X_2$  with  $p_k = P(X_k = 1) > 0$  ( $k = 1, 2$ ), it is possible to calculate the specific combined probability from the marginal probabilities and a correlation coefficient. Following Bahadur,<sup>4</sup> the joint probability of two correlated Bernoulli variables is determined by the marginal probabilities  $p_1, p_2$  and the second order correlation coefficient  $\delta$  (Pearson product-moment correlation coefficient) given by

$$\delta = E[Z_1 \cdot Z_2],$$

where  $Z_k = (X_k - p_k) / \sqrt{p_k(1 - p_k)}$ , ( $k = 1, 2$ ).

The joint probability function can be expressed as

$$P(X_1 = x_1, X_2 = x_2) = \prod_{k=1}^2 p_k^{x_k} \cdot q_k^{(1-x_k)} (1 + \delta \cdot z_1 \cdot z_2),$$

where  $z_k = \frac{x_k - p_k}{\sqrt{p_k q_k}}$  and  $q_k = 1 - p_k$ , ( $k = 1, 2$ ).

Hence, the combined probability  $p_*$  of observing at least one response is given by:

$$p_* = 1 - P(X_1 = 0, X_2 = 0) = 1 - q_1 q_2 - \delta \sqrt{p_1 q_1 p_2 q_2}.$$

Unlike in the continuous data case, the correlation  $\delta$  is not free to range over  $(-1, 1)$  (see Bahadur(1961) and Sozu(2010) for more details).<sup>4, 68</sup> The correlation coefficient  $\delta$  is bounded by:

$$\max \left\{ -\sqrt{\frac{p_1 p_2}{q_1 q_2}}, -\sqrt{\frac{q_1 q_2}{p_1 p_2}} \right\} \leq \delta \leq \min \left\{ \sqrt{\frac{p_1 q_2}{q_1 p_2}}, \sqrt{\frac{q_1 p_2}{p_1 q_2}} \right\}.$$

Note that given  $p_1$  and  $p_2$ , the higher (positive) correlation, the lower the combined probability  $p_*$  is. In Table 5.1 we indicate the possible range of values of the combined probability and correlation depending on the marginals  $p_1$  and  $p_2$ . For example, if we assume the two relevant marginal probabilities  $p_1 = 0.1$  and  $p_2 = 0.2$ , the possible values of the combined probability  $p_*$  will range over the interval  $(0.2, 0.3)$  whose probability boundaries will correspond to the correlation coefficients  $\delta$  equal to 0.67 and -0.17, respectively. In case of independence between  $X_1$  and  $X_2$ , then  $p_* = 0.28$ . When  $p_1 = p_2 = 0.5$ , it leads to the widest correlation's interval  $(-1, 1)$  and the widest probability interval  $(0.5, 1)$ .

		$p_2$																							
		0,01	0,05	0,10	0,20	0,50	0,60	0,90	0,999																
$p_1$	0,01	0,01	0,02	0,02	0,05	0,06	0,06	0,10	0,11	0,11	0,20	0,21	0,21	0,50	0,51	0,51	0,60	0,60	0,61	0,90	0,90	0,91	1,00	1,00	1,00
	0,05	<i>1,00</i>	<i>0</i>	<i>-0,01</i>	<i>0,44</i>	<i>0</i>	<i>-0,02</i>	<i>0,30</i>	<i>0</i>	<i>-0,03</i>	<i>0,20</i>	<i>0</i>	<i>-0,05</i>	<i>0,10</i>	<i>0</i>	<i>-0,10</i>	<i>0,08</i>	<i>0</i>	<i>-0,12</i>	<i>0,03</i>	<i>0</i>	<i>-0,30</i>	<i>0,00</i>	<i>0</i>	<i>-0,31</i>
	0,10	0,05	0,06	0,06	0,05	0,10	0,10	0,10	0,15	0,15	0,20	0,24	0,25	0,50	0,53	0,55	0,60	0,62	0,65	0,90	0,91	0,95	1,00	1,00	1,00
	0,20	<i>0,44</i>	<i>0</i>	<i>-0,02</i>	<i>1,00</i>	<i>0</i>	<i>-0,05</i>	<i>0,69</i>	<i>0</i>	<i>-0,08</i>	<i>0,46</i>	<i>0</i>	<i>-0,11</i>	<i>0,23</i>	<i>0</i>	<i>-0,23</i>	<i>0,19</i>	<i>0</i>	<i>-0,28</i>	<i>0,08</i>	<i>0</i>	<i>-0,69</i>	<i>0,01</i>	<i>0</i>	<i>-0,14</i>
	0,50	0,10	0,11	0,11	0,10	0,15	0,15	0,10	0,19	0,20	0,20	0,28	0,30	0,50	0,55	0,60	0,60	0,64	0,70	0,90	0,91	1,00	1,00	1,00	1,00
	0,60	<i>0,30</i>	<i>0</i>	<i>-0,03</i>	<i>0,69</i>	<i>0</i>	<i>-0,08</i>	<i>1,00</i>	<i>0</i>	<i>-0,11</i>	<i>0,67</i>	<i>0</i>	<i>-0,17</i>	<i>0,33</i>	<i>0</i>	<i>-0,33</i>	<i>0,27</i>	<i>0</i>	<i>-0,41</i>	<i>0,11</i>	<i>0</i>	<i>-1,00</i>	<i>0,01</i>	<i>0</i>	<i>-0,09</i>
	0,90	0,20	0,21	0,21	0,20	0,24	0,25	0,20	0,36	0,40	0,20	0,36	0,40	0,50	0,60	0,70	0,60	0,68	0,80	0,90	0,92	1,00	1,00	1,00	1,00
	0,999	<i>0,20</i>	<i>0</i>	<i>-0,05</i>	<i>0,46</i>	<i>0</i>	<i>-0,11</i>	<i>0,67</i>	<i>0</i>	<i>-0,17</i>	<i>1,00</i>	<i>0</i>	<i>-0,25</i>	<i>0,50</i>	<i>0</i>	<i>-0,50</i>	<i>0,41</i>	<i>0</i>	<i>-0,61</i>	<i>0,17</i>	<i>0</i>	<i>-0,67</i>	<i>0,02</i>	<i>0</i>	<i>-0,06</i>
		0,50	0,51	0,51	0,50	0,53	0,55	0,50	0,60	0,70	0,50	0,60	0,70	0,50	0,75	1,00	0,60	0,80	1,00	0,90	0,95	1,00	1,00	1,00	1,00
		<i>0,10</i>	<i>0</i>	<i>-0,10</i>	<i>0,23</i>	<i>0</i>	<i>-0,23</i>	<i>0,33</i>	<i>0</i>	<i>-0,33</i>	<i>0,50</i>	<i>0</i>	<i>-0,50</i>	<i>1,00</i>	<i>0</i>	<i>-1,00</i>	<i>0,82</i>	<i>0</i>	<i>-0,82</i>	<i>0,33</i>	<i>0</i>	<i>-0,33</i>	<i>0,03</i>	<i>0</i>	<i>-0,03</i>
	0,60	0,60	0,61	0,60	0,62	0,65	0,60	0,64	0,70	0,60	0,68	0,80	0,60	0,80	1,00	0,60	0,84	1,00	0,90	0,96	1,00	1,00	1,00	1,00	
	<i>0,08</i>	<i>0</i>	<i>-0,12</i>	<i>0,19</i>	<i>0</i>	<i>-0,28</i>	<i>0,27</i>	<i>0</i>	<i>-0,41</i>	<i>0,41</i>	<i>0</i>	<i>-0,61</i>	<i>0,82</i>	<i>0</i>	<i>-0,82</i>	<i>1,00</i>	<i>0</i>	<i>-0,67</i>	<i>0,41</i>	<i>0</i>	<i>-0,27</i>	<i>0,04</i>	<i>0</i>	<i>-0,03</i>	
	0,90	0,90	0,91	0,90	0,91	0,95	0,90	0,91	1,00	0,90	0,92	1,00	0,90	0,95	1,00	0,90	0,96	1,00	0,90	0,99	1,00	1,00	1,00	1,00	
	<i>0,03</i>	<i>0</i>	<i>-0,30</i>	<i>0,08</i>	<i>0</i>	<i>-0,69</i>	<i>0,11</i>	<i>0</i>	<i>-1,00</i>	<i>0,17</i>	<i>0</i>	<i>-0,67</i>	<i>0,33</i>	<i>0</i>	<i>-0,33</i>	<i>0,41</i>	<i>0</i>	<i>-0,27</i>	<i>1,00</i>	<i>0</i>	<i>-0,11</i>	<i>0,09</i>	<i>0</i>	<i>-0,01</i>	
	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	
	<i>0,00</i>	<i>0</i>	<i>-0,31</i>	<i>0,01</i>	<i>0</i>	<i>-0,14</i>	<i>0,01</i>	<i>0</i>	<i>-0,09</i>	<i>0,02</i>	<i>0</i>	<i>-0,06</i>	<i>0,03</i>	<i>0</i>	<i>-0,03</i>	<i>0,04</i>	<i>0</i>	<i>-0,03</i>	<i>0,09</i>	<i>0</i>	<i>-0,01</i>	<i>1,00</i>	<i>0</i>	<i>0,00</i>	

Table 5.1: Range of combined probability values (grey shadow) and the corresponding correlation coefficient (below, in italics) depending on the marginal probabilities  $p_1$  and  $p_2$ .

### 5.1.3 Expression of the combined probability $p_*$ as a function of several parameters.

#### Time-to-event data

We follow analogous steps that Gómez and Lagakos did to develop the *ARE* method, to calculate the combined probability. Considering time-to-event data, we can express the  $p_*$  in terms of several parameters depending on joint and marginal distribution assumptions.

#### Some notation

Define  $T_1$  and  $T_2$  as the marginal times to  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , respectively. The time to the combined endpoint  $\mathcal{E}_*$  is given by  $T_* = \min(T_1, T_2)$ . Assuming that  $\tau$  is the only non-informative censoring time, the combined probability  $p_*$  of observing the combined endpoint during the follow-up period is given by:

$$p_* = P((T_1 < \tau) \cup (T_2 < \tau)) = P(T_* < \tau) = 1 - S_*(\tau),$$

where  $S_*(\tau)$  is the survival function of the combined endpoint at time  $\tau$ .

We can approach the bivariate distribution  $(T_1, T_2)$  decoupling the joint survival of  $(T_1, T_2)$  into univariate components using a copula model given by.

$$S_*(t) = C(S_1(t), S_2(t); \theta),$$

where  $S_1(t)$  and  $S_2(t)$  are the marginal survival functions for  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , respectively; and  $\theta$  is the association parameter between  $T_1$  and  $T_2$ , which is biunivocally related to Spearman's rank correlation  $\rho$  ( $-1 < \rho < 1$ )<sup>(1)</sup>. Therefore, the probability  $p_*$  for a follow-up period  $\tau$  is in terms of the marginal laws of  $T_1$  and  $T_2$ , the chosen copula for the joint distribution of  $(T_1, T_2)$ , and  $\theta$ :

$$p_* = 1 - S_*(\tau) = 1 - C(S_1(\tau), S_2(\tau); \theta). \quad (5.2)$$

#### Expression of $p_*$ depending on whether $\mathcal{E}_1$ or $\mathcal{E}_2$ includes terminating events

The values of  $S_1(\tau)$  and  $S_2(\tau)$  depend on whether the relevant components include a terminating event since the observation of one component might preclude the observation of the other component (Cases 1,2,3 and 4 in Gómez and Lagakos<sup>27</sup>). We next derive the expression of  $S_1(\tau)$  and  $S_2(\tau)$  for each Case separately.

- **Neither  $\mathcal{E}_1$  nor  $\mathcal{E}_2$  includes a terminating event (Case 1)**

The expression of the marginal survival functions at time  $\tau$  is given by:

$$S_1(\tau) = 1 - p_1,$$

$$S_2(\tau) = 1 - p_2.$$

Hence, from (5.2) the expression of  $p_*$  in Case 1 is given by:

$$p_* = 1 - C(1 - p_1, 1 - p_2; \theta). \quad (5.3)$$

<sup>1</sup>See more details in the Appendix. Other copulas could also be considered (Plana-Ripoll O. and Gómez G., 2015).<sup>51</sup>



• **Either  $\mathcal{E}_1$  or  $\mathcal{E}_2$  includes a terminating event (Cases 2 and 3)**

In these cases, the occurrence of one endpoint might preclude the observation of the other. Considering  $\mathcal{E}_2$  as the terminating event (Case 2), the occurrence of  $\mathcal{E}_2$  precludes the observation of  $\mathcal{E}_1$  when  $T_1 > T_2$  (see Figure 5.2). Hence,  $S_1(\tau)$  has to be derived in terms of the probability  $p_1$  plus another unobservable probability  $U_1$ :

$$S_1(\tau) = 1 - [p_1 + P(T_1 > T_2, T_1 < \tau)] = 1 - [p_1 + U_1]$$

$$S_2(\tau) = 1 - p_2,$$

where  $U_1 = \int_0^\tau \int_0^u f_{(1,2)}(u, v) dv du$ , and  $f_{(1,2)}(u, v)$  is the joint density of  $(T_1, T_2)$ .

Thus, from the expression 5.2,  $p_*$  in Case 2 is given by:

$$p_* = 1 - C(1 - p_1 - U_1, 1 - p_2; \theta). \quad (5.4)$$

In Case 3, the occurrence of the terminating event  $\mathcal{E}_1$  precludes the observation of  $\mathcal{E}_2$  when  $T_1 < T_2$  (see Figure 5.2). Thus,  $S_2(\tau)$  is function of the probability  $p_2$  plus another unobservable probability  $U_2$ :

$$S_1(\tau) = 1 - p_1$$

$$S_2(\tau) = 1 - [p_2 + P(T_2 > T_1, T_2 < \tau)] = 1 - [p_2 + U_2],$$

where  $U_2 = \int_0^\tau \int_0^v f_{(1,2)}(u, v) du dv$ . Hence,  $p_*$  in Case 3 is given by:

$$p_* = 1 - C(1 - p_1, 1 - p_2 - U_2; \theta). \quad (5.5)$$

If we assume a Frank's Archimedean survival copula, the joint density of  $(T_1, T_2)$  is given by

$$f_{(1,2)}(t_1, t_2; \theta) = \frac{\theta e^{-\theta(S_1(t_1) + S_2(t_2))}}{e^{-2\theta S_{(1,2)}(t_1, t_2; \theta)} (e^{-\theta} - 1)} [f_1(t_1)] [f_2(t_2)]. \quad (5.6)$$

- **Both  $\mathcal{E}_1$  and  $\mathcal{E}_2$  include a terminating event (Case 4)**

In Case 4, both marginal survival functions are in terms of the unobservable probabilities  $U_1$  and  $U_2$  (defined previously for Cases 2 and 3):

$$S_1(\tau) = 1 - [p_1 + U_1],$$

$$S_2(\tau) = 1 - [p_2 + U_2].$$

Therefore, from (5.2),  $p_*$  in Case 4 is given by:

$$p_* = 1 - C(1 - p_1 - U_1, 1 - p_2 - U_2; \theta). \quad (5.7)$$

Note that the expression of  $p_*$  can be simplified as  $p_* = p_1 + p_2$ .

*Proof:*

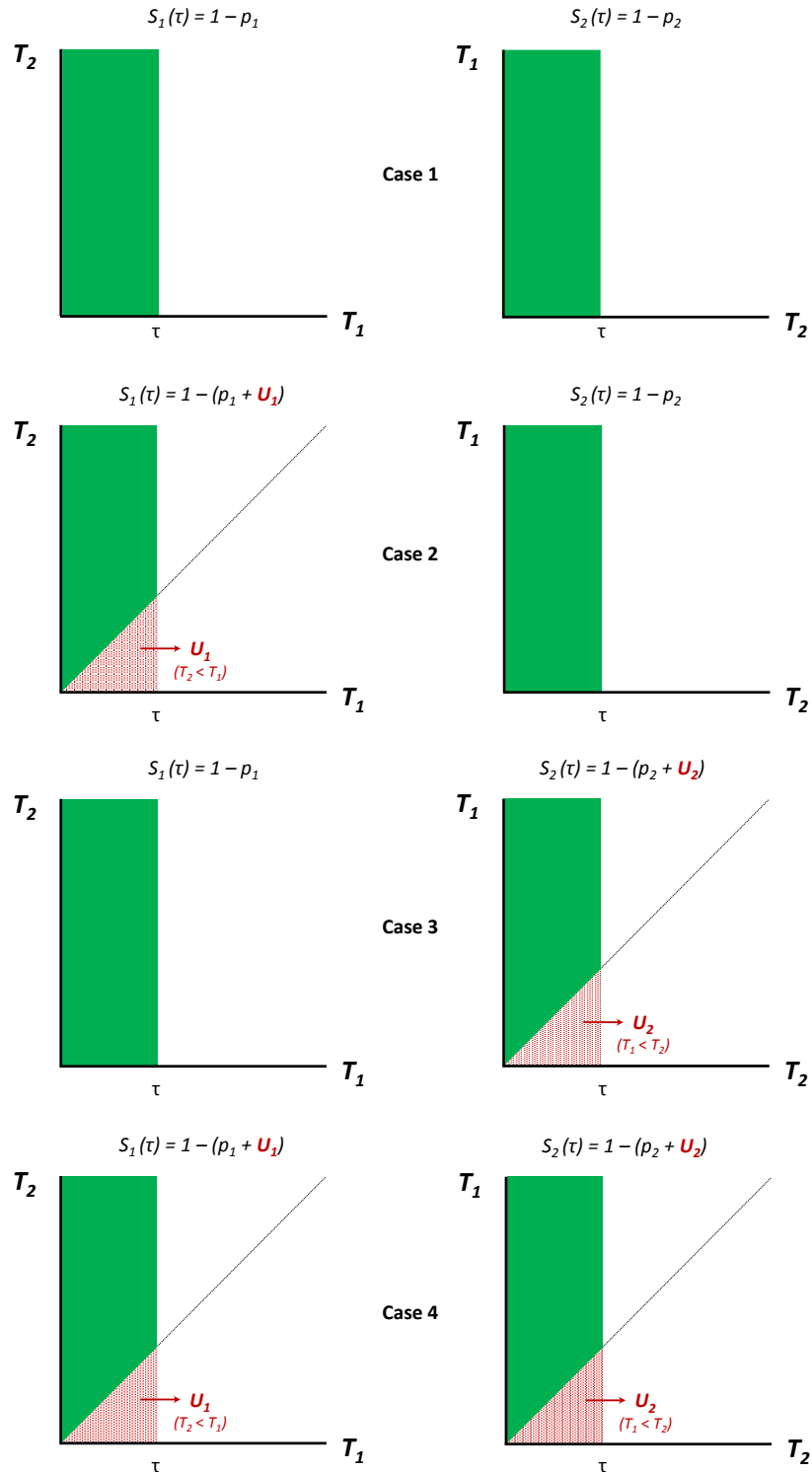
We have that  $p_* = P(\min(T_1, T_2) < \tau)$ . In Case 4, the observation of  $\mathcal{E}_1$  precludes the observation of  $\mathcal{E}_2$ , and vice versa. Thus,  $p_* = P(T_1 < \tau, T_1 < T_2) + P(T_2 < \tau, T_1 > T_2) = P(T_1 < \min(T_2, \tau)) + P(T_2 < \min(T_1, \tau)) = p_1 + p_2$ .

It is important to take into account whether each endpoint includes a terminating event. As commented in Rauch<sup>56</sup> and Gooley,<sup>31</sup> the calculation of the probabilities via  $1 - S(\tau)$  ignoring that one event is censoring the other might lead to inflated values above the true event probability. That is, ignoring the probabilities  $U_1$  and  $U_2$  might inflate the values of  $S_1(\tau)$  and  $S_2(\tau)$  for Cases 2, 3 and 4.

Summarizing, for Case 4 the combined probability  $p_*$  can be calculated as the sum of the marginal observed probabilities  $p_1$  and  $p_2$  (see Table 5.2). For Case 1,  $p_*$  can be calculated straightforwardly from expression (5.3) given  $p_1, p_2$  and the association parameter  $\theta$  for any copula, regardless of the marginal distribution laws of  $T_1$  and  $T_2$ . For Cases 2 and 3, we also need to set the laws for  $T_1$  and  $T_2$  in order to calculate the joint density  $f_{(1,2)}(t_1, t_2)$  for a chosen copula (expressions for Weibull distribution are detailed in the Appendix).

	$p_*$
Case 4	$p_1 + p_2$
Case 3	$1 - C(1 - p_1, 1 - p_2 - U_2; \theta)$
Case 2	$1 - C(1 - p_1 - U_1, 1 - p_2; \theta)$
Case 1	$1 - C(1 - p_1, 1 - p_2; \theta)$

**Table 5.2:** Expression of the combined probability  $p_*$  for each Case.  $p_1$  and  $p_2$  stands for the probability of observing  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , respectively;  $\theta$  is the association parameter between the time  $T_1$  and  $T_2$  to  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , respectively.  $U_1$  and  $U_2$  are the unobservable probabilities.  $C$  indicates the copula.



**Figure 5.2:** Calculation of the marginal survival functions  $S_1$  and  $S_2$  of the endpoints  $\mathcal{E}_1$  and  $\mathcal{E}_2$  at the end-of-study time  $\tau$  for each Case in terms of the observable probabilities  $p_1$  and  $p_2$  and the unobservable probabilities  $U_1$  and  $U_2$ . The green/red shadows represent the observable/unobservable combinations of times  $T_1$  and  $T_2$ . In Case 1,  $T_1$  and  $T_2$  will be always observable until  $\tau$ . In Case 2, the observation of  $T_2$  precludes the observation  $T_1$  when  $(T_1 > T_2)$ . In Case 3, the observation of  $T_1$  precludes the observation  $T_2$  when  $(T_1 < T_2)$ . In Case 4, the observation of  $T_1$  and  $T_2$  precludes the observation of  $T_2$  and  $T_1$  when  $T_1 < T_2$  and  $T_1 > T_2$ , respectively.

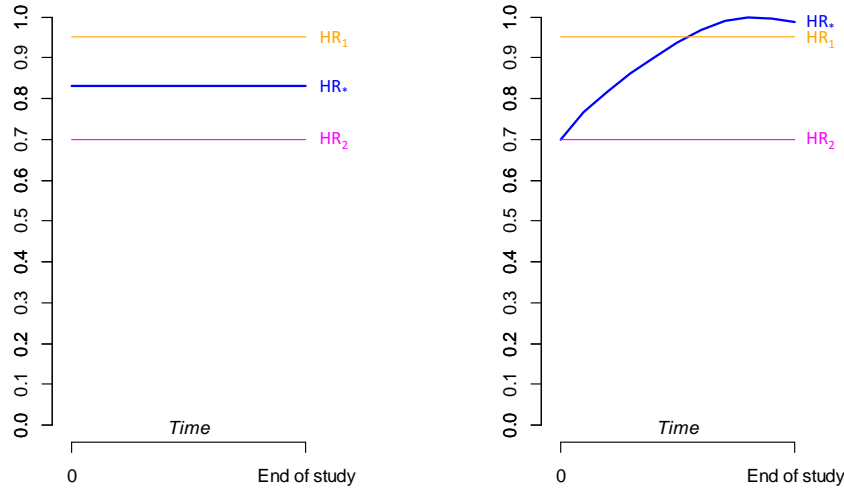
## 5.2 Combined hazard ratio

The combined hazard ratio  $HR_*(t)$  is defined as the ratio of the hazard functions  $\lambda_*^{(1)}(t)$  and  $\lambda_*^{(0)}(t)$ . In this section we propose a practical solution to assign  $HR_*(t)$  values in terms of the marginal hazard ratios. The specific expression for the combined  $HR_*(t)$  can be derived in terms of several parameter values. In cases where the  $HR_*(t)$  is constant over time, we can calculate the specific  $HR_*(t)$  in terms of these parameter values. Otherwise, since some of these parameters are difficult to anticipate by trialists, and since the  $HR_*(t)$  is not always constant over time, we propose to use a range of plausible values.

### 5.2.1 Notation and assumptions

We denote by  $T_1^{(j)}$  and  $T_2^{(j)}$  the marginal times to the endpoint components  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , respectively, for both groups  $j$  ( $j = 0, 1$ ). Define the time to the combined endpoint  $\mathcal{E}_* = \mathcal{E}_1 \cup \mathcal{E}_2$  as  $T_*^{(j)}$  for each group  $j$ . We assume for the remainder of this section:

- **Censoring:** The end-of-study censoring at time  $\tau$  is the only non-informative censoring cause for both groups.
- **Proportional hazards:** The hazard ratios between  $T_1^{(0)}$  and  $T_1^{(1)}$  and between  $T_2^{(0)}$  and  $T_2^{(1)}$  are constant. That is,  $HR_1(t) = \lambda_1^{(1)}(t)/\lambda_1^{(0)}(t) = HR_1$  and  $HR_2(t) = \lambda_2^{(1)}(t)/\lambda_2^{(0)}(t) = HR_2$  for all  $t$ . Note that although we are assuming that the hazard functions  $\lambda_1^{(j)}(t)$  and  $\lambda_2^{(j)}(t)$  ( $j = 0, 1$ ) are proportional, this does not imply the proportionality of the combined hazards  $\lambda_*^{(0)}(t)$  and  $\lambda_*^{(1)}(t)$  for the combined endpoint  $\mathcal{E}_*$  (see Figure 5.3, right).
- **Copula assumptions:** We consider a Frank's Archimedean copula relationship between  $T_1^{(0)}$  and  $T_2^{(0)}$ , which is in terms of an association parameter  $\theta^{(0)}$ , and between  $T_1^{(1)}$  and  $T_2^{(1)}$ , which is in terms of an association parameter  $\theta^{(1)}$ . We assume equal association parameter  $\theta = \theta^{(0)} = \theta^{(1)}$  for both groups 0 and 1, which is biunivocally related to Spearman's rank correlation  $\rho$  ( $-1 < \rho < 1$ ) (see Appendix for more details). Other copulas can also be used (Plana-Ripoll O. and Gómez G., 2015).<sup>51</sup>
- **Marginal laws:** We assume Weibull distributions for  $T_1^{(j)}$  and  $T_2^{(j)}$  ( $j = 0, 1$ ) with scale and shape parameters  $b_1^{(j)}, b_2^{(j)}$  and  $\beta_1^{(j)}, \beta_2^{(j)}$ , respectively. The shape parameters are chosen equal for both groups, that is  $\beta_k^{(0)} = \beta_k^{(1)} = \beta_k$  ( $k = 1, 2$ ), so that the assumption of proportionality of the marginal hazards holds.



**Figure 5.3:** Values of the combined hazard ratio  $HR_*(t)$  over time assuming constant hazard ratios  $HR_1 = 0.7$  and  $HR_2 = 0.95$ , with a probability  $p_1 = 0.15$  and  $p_2 = 0.15$  of observing the marginal component  $\mathcal{E}_1$  (terminating) and  $\mathcal{E}_2$  (non terminating), respectively. We assume Weibull distributions with constant hazard rates for  $\mathcal{E}_1$  and increasing hazard rates for  $\mathcal{E}_2$  with correlation  $\rho = 0.95$  (right); and constant hazards with null correlation (left).

### 5.2.2 Expression of $HR_*(t)$ as a function of several parameters

The expression of the combined hazard ratio is given by:

$$HR_*(t) = \frac{\lambda_*^{(1)}(t)}{\lambda_*^{(0)}(t)} = \frac{\frac{f_*^{(1)}(t)}{S_*^{(1)}(t)}}{\frac{f_*^{(0)}(t)}{S_*^{(0)}(t)}}, \quad (5.8)$$

where  $f_*^{(j)}(t)$  and  $S_*^{(j)}(t)$  ( $j = 0, 1$ ) are the density and survival functions of  $T_*^{(j)}$ , respectively.

In the same way that Gómez and Lagakos did to derive the hazard ratio for the composite endpoint of the relevant and the additional endpoint, we can derive the combined  $HR_*(t)$  from the following parameters:

- The marginal hazard ratios  $HR_1$  and  $HR_2$ ,
- the probabilities  $p_1$  and  $p_2$  of observing each component  $\mathcal{E}_1$  and  $\mathcal{E}_2$  in control group,
- the shape parameters  $\beta_1$  and  $\beta_2$  for  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , respectively,
- and the correlation  $\rho$  between the two components.

The expression of  $HR_*(t)$  also depends on whether the  $\mathcal{E}_1$  or  $\mathcal{E}_2$  has a terminating event or not (Cases 1,2,3 and 4).

### 5.2.3 Scenarios with constant $HR_*(t)$

We next describe how to calculate the combined hazard ratio  $HR_*(t)$  when it is constant over time (see Figure 5.3, left). In section (5.3) we will describe some practical solutions in order to supply a value for  $HR_*(t)$  when the proportionality of the hazards for the combined endpoint does not hold (see Figure 5.3, right). A brief summary of the results is shown in Table 5.8.

**a) Null treatment effect on  $\mathcal{E}_1$  and  $\mathcal{E}_2$  ( $HR_1 = HR_2 = 1$ ) implies  $HR_*(t) = 1$ .**

If there is no treatment effect on the marginal components ( $HR_1 = HR_2 = 1$ ), then there is no treatment effect on the combined endpoint ( $HR_*(t) = 1$ ). That is, if the null hypotheses of no treatment effect of  $\mathcal{E}_1$  and  $\mathcal{E}_2$  are true ( $H_0: HR_1 = 1$  and  $H_0: HR_2 = 1$ , respectively), it implies that the null hypothesis of no treatment effect of  $\mathcal{E}_*$  ( $H_{0*}: HR_*(t) = 1$ ) is also true.

*Proof:*

The expression of  $S_*^{(j)}$  for each group  $j$  is given by:

$$S_*^{(j)}(t) = C(S_1^{(j)}(t), S_2^{(j)}(t)), \quad (5.9)$$

where  $C$  indicates the copula model between the marginal times.

If  $HR_1(t) = \lambda_1^{(1)}(t)/\lambda_1^{(0)}(t) = 1$ , then  $\lambda_1^{(0)}(t) = \lambda_1^{(1)}(t)$ . Since the hazard function characterizes the law of the random variables, we have that  $S_1^{(0)}(t) = S_1^{(1)}(t)$ . Analogously, if  $HR_2(t) = 1$ , we have  $S_2^{(0)}(t) = S_2^{(1)}(t)$ . Therefore, from (5.9),  $S_*^{(0)}(t) = S_*^{(1)}(t)$  and hence  $f_*^{(0)}(t) = f_*^{(1)}(t)$ . Thus,

$$HR_*(t) = \frac{\frac{f_*^{(1)}(t)}{S_*^{(1)}(t)}}{\frac{f_*^{(0)}(t)}{S_*^{(0)}(t)}} = 1.$$

Note that it also applies to any marginal distribution, copula model, and correlation between marginal times.

**b) Null correlation between marginal times when  $\lambda_2^{(0)}(t)/\lambda_1^{(0)}(t)$  is constant implies constant  $HR_*(t)$ .**

When assuming null correlation, the combined  $HR_*(t)$  is constant when the ratio of the baseline hazard functions  $\lambda_1^{(0)}(t)$  and  $\lambda_2^{(0)}(t)$  is as well constant (Gómez G., 2011).<sup>21</sup> The expression of this ratio is given by:

$$\frac{\lambda_2^{(0)}(t)}{\lambda_1^{(0)}(t)} = \frac{HR_1 - HR_*(t)}{HR_*(t) - HR_2}. \quad (5.10)$$

*Proof:*

The expression of the survival function  $S_*^{(j)}(t)$  for each group  $j = 0, 1$  is given by:

$$S_*^{(j)}(t) = P(T_1 > t \cap T_2 > t).$$

Assuming independence between times (and hence  $\rho = 0$  for Spearman's correlation),  $S_*^{(j)}(t)$  is given by

$$S_*^{(j)}(t) = P(T_1^{(j)} > t) \cdot P(T_2^{(j)} > t) = S_1^{(j)}(t) \cdot S_2^{(j)}(t).$$

The expression of the density function  $f_*^{(j)}(t)$  for each group  $j$  is given by:

$$f_*^{(j)}(t) = \frac{-\partial S_*^{(j)}(t)}{\partial t} = \frac{-\partial(S_1^{(j)}(t) \cdot S_2^{(j)}(t))}{\partial t} = f_1^{(j)}(t) \cdot S_2^{(j)}(t) + S_1^{(j)}(t) \cdot f_2^{(j)}(t).$$

Hence, from (5.8) we have:

$$HR_*(t) = \frac{f_*^{(1)}(t)/S_*^{(1)}(t)}{f_*^{(0)}(t)/S_*^{(0)}(t)} = \frac{\left[ f_1^{(1)}(t) \cdot S_2^{(1)}(t) + S_1^{(1)}(t) \cdot f_2^{(1)}(t) \right] / S_1^{(1)}(t) \cdot S_2^{(1)}(t)}{\left[ f_1^{(0)}(t) \cdot S_2^{(0)}(t) + S_1^{(0)}(t) \cdot f_2^{(0)}(t) \right] / S_1^{(0)}(t) \cdot S_2^{(0)}(t)}.$$

We express  $f_1^{(j)}(t) = \lambda_1^{(j)}(t) \cdot S_1^{(j)}(t)$ , and  $f_2^{(j)}(t) = \lambda_2^{(j)}(t) \cdot S_2^{(j)}(t)$ :

$$\begin{aligned} HR_*(t) &= \frac{\left[ \lambda_1^{(1)}(t) \cdot S_1^{(1)}(t) \cdot S_2^{(1)}(t) + S_1^{(1)}(t) \cdot \lambda_2^{(1)}(t) \cdot S_2^{(1)}(t) \right] \cdot S_1^{(0)}(t) \cdot S_2^{(0)}(t)}{\left[ \lambda_1^{(0)}(t) \cdot S_1^{(0)}(t) \cdot S_2^{(0)}(t) + S_1^{(0)}(t) \cdot \lambda_2^{(0)}(t) \cdot S_2^{(0)}(t) \right] \cdot S_1^{(1)}(t) \cdot S_2^{(1)}(t)} \\ &= \frac{S_1^{(1)}(t) \cdot S_2^{(1)}(t) \cdot \left[ \lambda_1^{(1)}(t) + \lambda_2^{(1)}(t) \right] \cdot S_1^{(0)}(t) \cdot S_2^{(0)}(t)}{S_1^{(0)}(t) \cdot S_2^{(0)}(t) \cdot \left[ \lambda_1^{(0)}(t) + \lambda_2^{(0)}(t) \right] \cdot S_1^{(1)}(t) \cdot S_2^{(1)}(t)}. \end{aligned}$$

After canceling the survival functions, we have:

$$HR_*(t) = \frac{\lambda_1^{(1)}(t) + \lambda_2^{(1)}(t)}{\lambda_1^{(0)}(t) + \lambda_2^{(0)}(t)}.$$

Therefore, since  $\lambda_1^{(1)}(t) = HR_1 \cdot \lambda_1^{(0)}(t)$  and  $\lambda_2^{(1)}(t) = HR_2 \cdot \lambda_2^{(0)}(t)$ :

$$HR_*(t) = \frac{HR_1 \cdot \lambda_1^{(0)}(t) + HR_2 \cdot \lambda_2^{(0)}(t)}{\lambda_1^{(0)}(t) + \lambda_2^{(0)}(t)}. \quad (5.11)$$

The previous equality can be expressed as:

$$\frac{\lambda_2^{(0)}(t)}{\lambda_1^{(0)}(t)} = \frac{HR_1 - HR_*(t)}{HR_*(t) - HR_2}.$$

If  $\lambda_2^{(0)}(t)/\lambda_1^{(0)}(t) = c$  ( $c$  constant), then

$$c \cdot (HR_*(t) - HR_2) = HR_1 - HR_*(t).$$

Hence, we have:

$$HR_*(t) = \frac{HR_1 + c \cdot HR_2}{c + 1}.$$

Since  $HR_1$  and  $HR_2$  are constant, it implies that  $HR_*(t)$  is constant.

This result applies to any marginal distribution and copula model.

**c) Null correlation when  $HR_1 = HR_2 = k$  implies  $HR_*(t) = k$ .**

*Proof:*

Assuming  $\rho = 0$ , from (5.11), if  $HR_1 = HR_2 = k$ , we have:

$$HR_*(t) = \frac{k \cdot \lambda_1^{(0)}(t) + k \cdot \lambda_2^{(0)}(t)}{\lambda_1^{(0)}(t) + \lambda_2^{(0)}(t)} = k.$$

This result applies to any marginal distribution and copula model.

**d) Null correlation and Weibull distributions with  $(\beta_1 = \beta_2 = \beta)$  implies constant  $HR_*(t)$ .**

When the marginal laws of  $T_1$  and  $T_2$  are Weibull with equal shape parameters ( $\beta_1 = \beta_2 = \beta$ ) for both components  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , the combined  $HR_*(t)$  is constant and given by:

$$HR_*(t) = \frac{HR_1 + HR_2 \left( \frac{b_1^{(0)}}{b_2^{(0)}} \right)^\beta}{1 + \left( \frac{b_1^{(0)}}{b_2^{(0)}} \right)^\beta}. \quad (5.12)$$

*Proof:*

The ratio of the baseline hazard functions  $\lambda_1^{(0)}(t)$  and  $\lambda_2^{(0)}(t)$  is given by:

$$\frac{\lambda_2^{(0)}(t)}{\lambda_1^{(0)}(t)} = \frac{\frac{f_2^{(0)}(t)}{S_2^{(0)}(t)}}{\frac{f_1^{(0)}(t)}{S_1^{(0)}(t)}}.$$

Since we assume Weibull distributions with equal shape parameters ( $\beta_1 = \beta_2 = \beta$ ), from expressions (B.6) and (B.7) (Appendix), we have:

$$\frac{\lambda_2^{(0)}(t)}{\lambda_1^{(0)}(t)} = \frac{\frac{\beta}{(b_2^{(0)})^\beta} \cdot t^{(\beta-1)}}{\frac{\beta}{(b_1^{(0)})^\beta} \cdot t^{(\beta-1)}} = \left( \frac{b_1^{(0)}}{b_2^{(0)}} \right)^\beta. \quad (5.13)$$

From expressions (5.10) and (5.13), we can express the following equality:

$$\left( \frac{b_1^{(0)}}{b_2^{(0)}} \right)^\beta = \frac{HR_1 - HR_*(t)}{HR_*(t) - HR_2}.$$

Solving the equation for  $HR_*(t)$ , we have:

$$HR_*(t) = \frac{HR_1 + HR_2 \left( \frac{b_1^{(0)}}{b_2^{(0)}} \right)^\beta}{1 + \left( \frac{b_1^{(0)}}{b_2^{(0)}} \right)^\beta}.$$

This result applies to any copula model.



**e) Null correlation, Weibull distributions with  $(\beta_1 = \beta_2 = \beta)$  and  $p_1 = p_2$  for Case 1 implies  $HR_*(t) = (HR_1 + HR_2)/2$ .**

If the probability of observing both endpoints is the same ( $p_1 = p_2$ ) with no competing events (Case 1) and equal shape parameters ( $\beta_1 = \beta_2 = \beta$ ), then the combined  $HR_*(t)$  is equal to the mean of  $HR_1$  and  $HR_2$ .

*Proof:*

The equality of the marginal probabilities  $p_1$  and  $p_2$  of observing  $\mathcal{E}_1$  and  $\mathcal{E}_2$  in group 0 at censoring time  $\tau$ , respectively, for Case 1, implies that:

$$S_1^{(0)}(\tau) = S_2^{(0)}(\tau).$$

Since we assume Weibull distributions, we have:

$$e^{-(\tau/b_1^{(0)})^\beta} = e^{-(\tau/b_2^{(0)})^\beta},$$

which implies that

$$b_1^{(0)} = b_2^{(0)} = b^{(0)}.$$

Hence from (5.12),

$$HR_*(t) = \frac{HR_1 + HR_2 \left(\frac{b^{(0)}}{b^{(0)}}\right)^\beta}{1 + \left(\frac{b^{(0)}}{b^{(0)}}\right)^\beta} = \frac{HR_1 + HR_2}{2}.$$

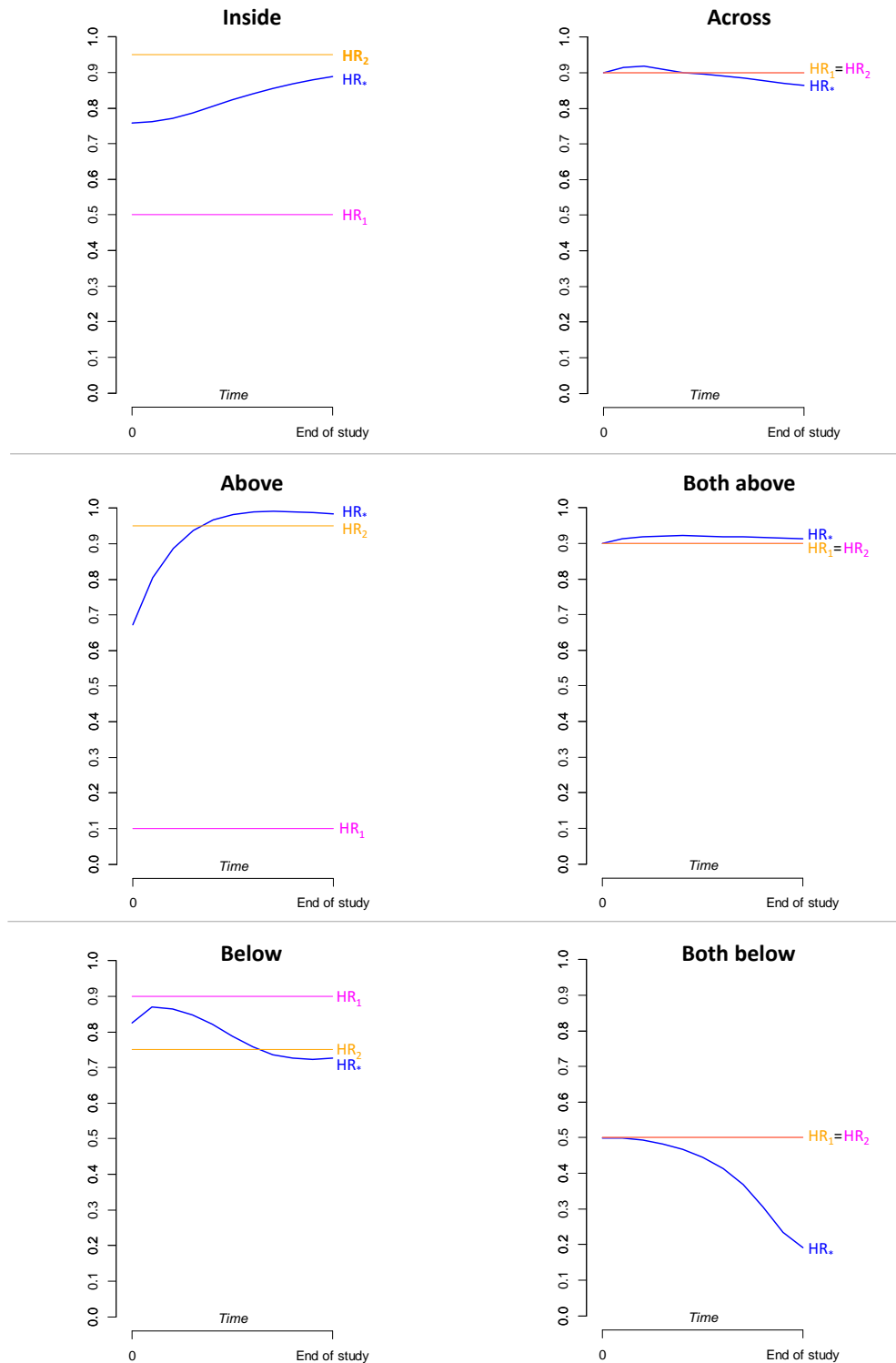
This result applies to any copula model.

## 5.3 Practical solutions to assign combined hazard ratios

### 5.3.1 General behavior of the $HR_*(t)$ over time

In the previous section 5.2 we have seen that it is possible to calculate the specific value of the combined  $HR_*(t)$  when investigators can anticipate all the required parameters. The main objective of this section is to assign a specific value  $HR_*(t)$  among all the values that the  $HR_*(t)$  can take over time when it is not constant or when investigators cannot anticipate all the needed parameters. To do so, a range of values for the combined hazard ratio will be calculated over the follow-up period from expression 5.8 and the assumptions detailed in subsection 5.2.1. For example, in Figure 5.4 (top left), we represent a specific scenario where the combined hazard ratio  $HR_*(t)$  ranges from 0.75 to 0.89.

We analyze how the combined  $HR_*(t)$  behaves taking the marginal  $HR_1$  and  $HR_2$  as fixed references. That is, the  $HR_*(t)$  interval, consisting of the minimum and maximum  $HR_*(t)$  over time, might have both boundaries inside the marginal hazard ratios, both outside or either one outside (see Figure 5.4). We consider multiple scenarios consisting of different parameter value combinations. Aggregated results will give us a general idea of how the  $HR_*(t)$  behaves depending on different parameter values. Thus, when investigators cannot anticipate all the needed parameter values (such as the correlation), they will be able to calculate a range of plausible values of  $HR_*(t)$  and act accordingly.



**Figure 5.4:** Behaviour of  $HR_*(t)$  over time with respect to the marginal hazard ratios  $HR_1$  and  $HR_2$ . That is,  $HR_*(t)$  can remain inside, outside or cross the marginal hazard ratios. Parameter values used to depict these plots belong to the parameter value setting used for the executions except for the plot "Both below", with which negative correlation is used only for illustration purpose.

### Parameter setting

We choose a wide range of values for each parameter in order to cover a wide range of scenarios (see Table 5.3). These chosen parametric values lead to 117,855 different scenarios and can be somehow understood as if we took a huge representative sample from the universe we want to study. We set a study duration from time  $t = 0$  to  $t = 1$  divided into intervals of width 0.1.

Only positive correlations and equal shape parameters  $\beta_1 = \beta_2 = \beta$  were included because they represent more realistic situations in clinical trials. Hazard ratios higher than 1 were excluded since investigators are not likely to include components with harmful effects at the design stage. Moreover, the ARE methodology is applied considering an alternative hypothesis of no treatment effect with hazard ratios lower than 1. We excluded probabilities larger than 0.5 and extremely low or high correlations in Cases 2,3 and 4. These excluded scenarios led to very low survival values (in the order of  $10^{-16}$ ) for some parameter value combinations at some time  $t < \tau = 1$ , indicating that the support of  $T_*$  was not  $[0,1]$ .

Note that since we are not including scenarios with neither  $HR_1 = HR_2 = 1$  nor  $\rho = 0$  (although they are very near 1 and 0, respectively), all the computed  $HR_*(t)$  lead to non-proportional hazards, as shown in section (5.2.3).

Parameters											
<b>Case 1</b>											
$p_1, p_2$	0.01	0.05	0.1	0.5	0.9	0.95	0.99				
$\rho$	0.01	0.1	0.5	0.9	0.99						
<b>Cases 2,3,4</b>											
$p_1, p_2$	0.01	0.05	0.1	0.5							
$\rho$	0.1	0.25	0.5	0.75	0.9						
$HR_1, HR_2$	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99		
$\beta_1 = \beta_2 = \beta$	0.5	1	2								
Time points	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Number of scenarios	Total	Case 1	Case 2	Case 3	Case 4						
	117,855	59,535	19,440	19,440	19,440						

**Table 5.3:** Parameter settings to compute the combined hazard ratios  $HR_*(t)$  for each parameter value combination.  $HR_1, p_1$  and  $HR_2, p_2$  stands for the marginal constant hazard ratio and probability in control group of each endpoint  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , respectively.  $\rho$  stands for the Spearman rank correlation between endpoint times.  $\beta_1$  and  $\beta_2$  are the shape parameters. For computational reasons, we use the values 0.01 and 0.99 instead of 0 and 1, respectively, for the hazard ratios and probabilities; 0.49 instead of 0.5 in Case 4 for the probabilities; and 0.0001 instead of 0 as the first time point.

## Results

### Behaviour of $HR_*(t)$ with respect to $HR_1$ and $HR_2$

For every pair of marginal hazard ratios we study when the different values of the  $HR_*(t)$  fall inside the area delimited by  $HR_1$  and  $HR_2$ . Whenever the  $HR_*(t)$  is outside this area for some  $t$ , we study the nature of the distinct departures (see Figure 5.4). Since Case 1 consists of a different parameter setting than Cases 2,3,4, we split the results by Cases. Due to their symmetry, we merge Cases 2 and 3 because they lead to the same aggregated results of  $HR_*(t)$ .

In 73% of scenarios, the combined  $HR_*(t)$  always remains between the marginal hazard ratios (see "Inside" in Table 5.4). For the rest of cases, we observe 5% of scenarios where the  $HR_*(t)$  crosses both the upper and lower marginal hazard ratios (which we call "Across"); 18% of scenarios where the  $HR_*(t)$  crosses or remains above the upper marginal HR ("Above"+"Both above"), and 5% of scenarios where the  $HR_*(t)$  crosses the lower marginal HR ("Below"). Note that the situation where the  $HR_*(t)$  remains below the lower marginal HR for all  $t$  never occurs in our setting. We executed some scenarios with negative correlations, and some cases led to this latter behaviour in the  $HR_*(t)$  (see bottom right in Figure 5.4).

	Global	Case 1	Cases 2&3	Case 4	$HR_1 \neq HR_2$	$HR_1 = HR_2$
<b>Inside</b>	73%	67%	79%	80%	82%	0%
<b>Across</b>	5%	8%	1%	2%	0%	41%
<b>Both above</b>	7%	3%	10%	10%	0%	59%
<b>Above</b>	11%	14%	8%	7%	12%	0%
<b>Below</b>	5%	8%	2%	1%	5%	0%
<b>Both below</b>	0%	0%	0%	0%	0%	0%

**Table 5.4:** Table representing the range of possible values of  $HR_*(t)$  over the follow-up time in relation to the marginal hazard ratios  $HR_1$  and  $HR_2$ . Whenever  $HR_1 \leq HR_2$ , we distinguish situations according to whether  $\max_t\{HR_*(t)\} \leq HR_1$  (both below),  $\min_t\{HR_*(t)\} \geq HR_2$  (both above),  $HR_1 \leq HR_*(t) \leq HR_2 \forall t$  (inside), as well as three other combinations. Analogously, the same is applied whenever ( $HR_1 \geq HR_2$ ).

Whenever  $HR_1 = HR_2$ , since we are excluding scenarios with  $\rho = 0$  (and hence non-constant  $HR_*(t)$ ), none of the scenarios will fall between the marginal hazard ratios for all  $t$ . In those cases,  $HR_*(t)$  either always remains above the marginals (59%) or crosses the marginal hazard ratio (41%) (see the last column in Table 5.4). Whenever  $HR_1 \neq HR_2$ , the  $HR_*(t)$  either is inside the marginals for all  $t$  (82%), or crosses one marginal hazard ratio (17%).

The correlation  $\rho$  between endpoint times plays an important role. The lower correlation, the higher percentage of scenarios where the combined  $HR_*(t)$  remains between the marginal hazard ratios for all  $t$  (see Table 5.5). Indeed, as stated in Gómez G.,<sup>21</sup> when having null correlation ( $\rho = 0$ ), the constant combined  $HR_*(t)$  lies always between  $HR_1$  and  $HR_2$ . That is, the treatment effect on the combined  $\mathcal{E}_*$  lies between the treatment effect on  $\mathcal{E}_1$  and  $\mathcal{E}_2$ . Whenever  $HR_1 \neq HR_2$ , the  $HR_*(t)$  remains always inside  $HR_1$  and  $HR_2$  for low correlations ( $\rho \leq 0.25$ ), and remains inside in the vast majority of cases whenever the correlation is moderate or high ( $0.5 \leq \rho \leq 0.75$ ).

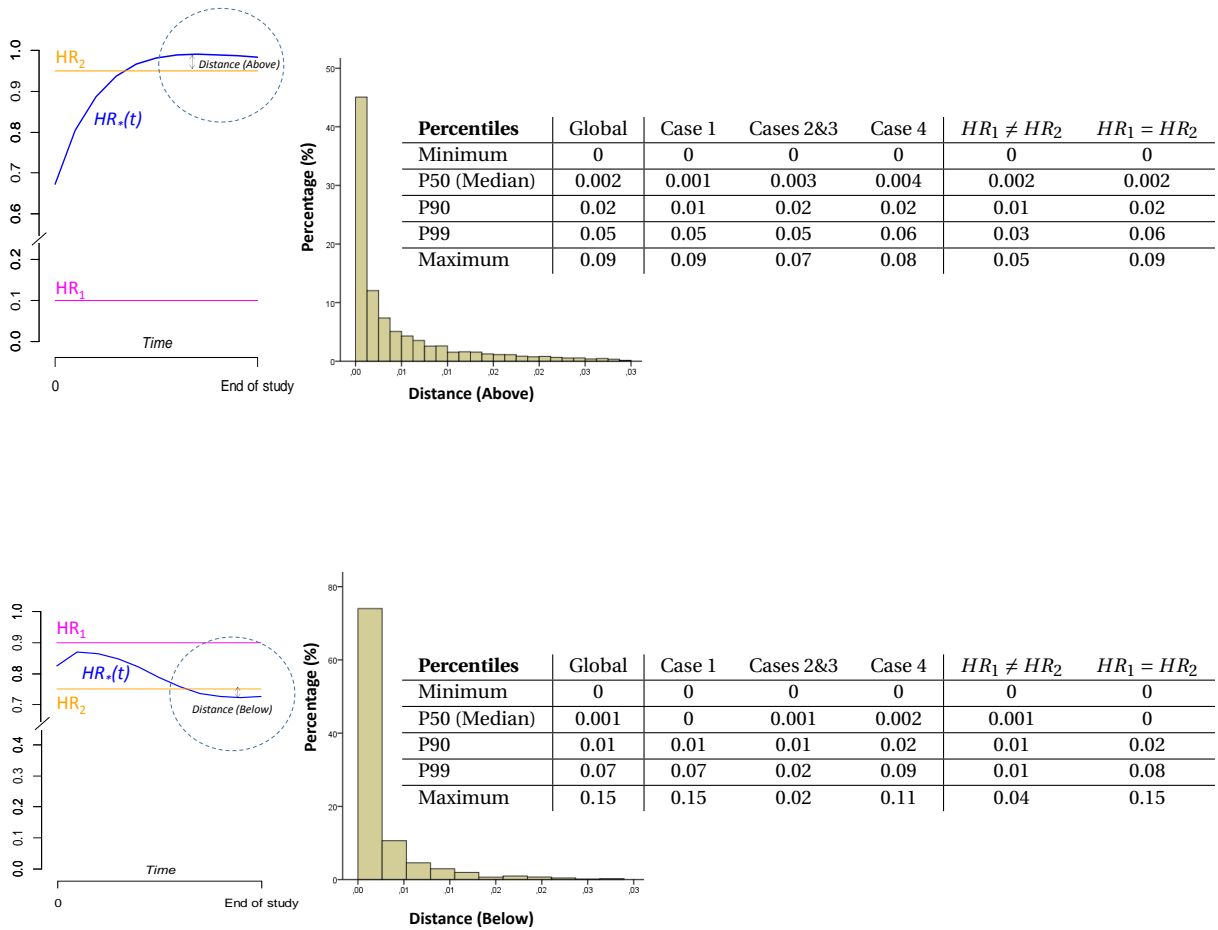
$\rho$	Global	Case 1	Cases 2&3	Case 4	$HR_1 \neq HR_2$
<b>0.01</b>	89%	89%	-	-	100%
<b>0.1</b>	89%	89%	89%	89%	100%
<b>0.25</b>	89%	-	89%	89%	100%
<b>0.5</b>	86%	84%	87%	87%	96%
<b>0.75</b>	72%	-	71%	74%	81%
<b>0.9</b>	51%	42%	60%	62%	58%
<b>0.99</b>	30%	30%	-	-	34%
<b>Global</b>	73%	75%	89%	90%	82%

**Table 5.5:** Scenarios with  $HR_*(t)$  inside the marginal hazard ratios  $HR_1$  and  $HR_2$ .

Summarizing, we observe that although  $HR_*(t)$  is not constant, their range of values lies between the marginal hazard ratios in 73% of cases, with this proportion reaching to almost 90% if the two components are close to be independent, and almost 100% whenever  $HR_1 \neq HR_2$  with moderate or low correlations. This behaviour informs the investigators about the plausible values for  $HR_*(t)$ , and hence to know that the assigned  $HR_*(t)$  might be between the marginal hazard ratios in the majority of scenarios, specially when having medium or low correlations.

#### Departure of $HR_*(t)$ with respect to $HR_1$ and $HR_2$

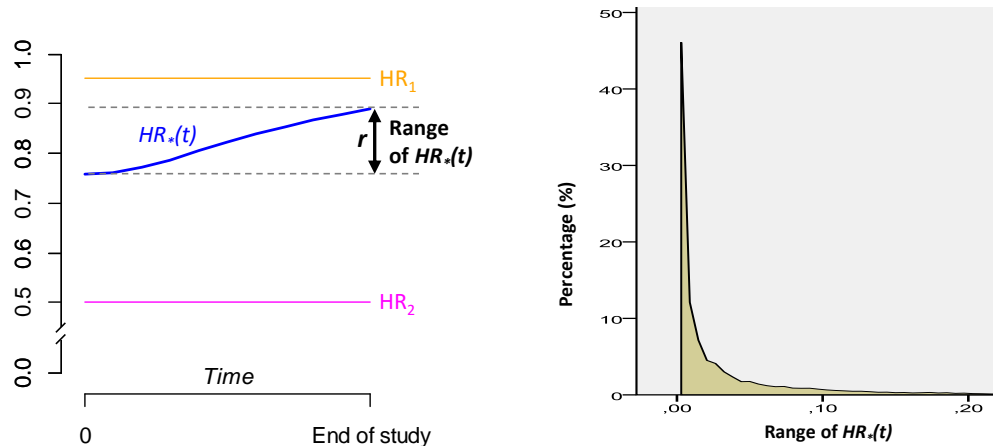
We now study the behaviour of  $HR_*(t)$  whenever it is outside the interval  $[HR_1, HR_2]$  (27% of scenarios). In particular, we analyze the departure of  $HR_*(t)$  from  $HR_1$  and  $HR_2$  for  $t$  such that  $HR_*(t) \notin [HR_1, HR_2]$ . In Figure 5.5, the distribution of the maximum distances between the marginal boundaries and the combined  $HR_*(t)$  is shown by means of a table with the percentiles in cases where the combined hazard ratio falls above or below the maximum and minimum marginal hazard ratios, respectively. We observe that these distances are very small in the majority of scenarios. Only 1% of scenarios reach distances above 0.05 and 0.07 in the upper and lower boundary. Therefore, given  $HR_1 \leq HR_2$ , the general combined  $HR_*(t)$  interval given by  $[HR_1 - 0.07, HR_2 + 0.05]$  would cover all the plausible  $HR_*(t)$  values over time in the 99% of the scenarios. Whenever  $HR_1 \neq HR_2$ , this interval shrinks to  $[HR_1 - 0.01, HR_2 + 0.03]$ . The interval is also shrunk when having low correlations (not shown). Scenarios reaching the widest interval consist of scenarios with  $HR_1 = HR_2$  and high correlation.



**Figure 5.5:** Illustrative examples, histograms and percentiles of the maximum distances between the  $HR_*(t)$  and the marginal hazard ratios  $HR_1$  and  $HR_2$  when  $HR_*(t)$  falls outside the marginal hazard ratios at some  $t$ . All the numbers are  $> 0$  although we do not specify all the decimals in order to make the table easier to read. Values in upper plot:  $(HR_1, HR_2) = (0.1, 0.95)$ ,  $(p_1, p_2) = (0.05, 0.1)$ . Values in bottom plot:  $(HR_1, HR_2) = (0.9, 0.75)$ ,  $p_1 = p_2 = 0.99$ . In both plots: Case 1 with  $\beta_1 = \beta_2 = 1$  and  $\rho = 0.99$ .

Departure from constancy of  $HR_*(t)$

Finally, we analyze  $r = \max_t\{HR_*(t)\} - \min_t\{HR_*(t)\}$  (see Figure 5.6) to assess how the  $HR_*(t)$  departs from being constant. In Figure 5.6 we depict the distribution of this measure by means of a histogram and we split in tables the median and maximum values by correlation and by absolute difference between the marginal hazard ratios ( $|HR_1 - HR_2|$ ). We observe that the larger the correlation and larger differences between marginal HR, the larger the distance  $r$  we get, and hence the larger departure from constancy of  $HR_*(t)$ . We remark that although these departures are close to zero in median, we have to take into consideration that small changes in hazard ratios might have a great impact in sample size calculations, as it is commented in Chapter 7 as future research.



$\rho$	Case 1		Cases 2&3		Case 4	
	Med.	Max.	Med.	Max.	Med.	Max.
<b>0.01</b>	0	0.02	-	-	-	-
<b>0.1</b>	0.003	0.04	0.002	0.04	0.002	0.04
<b>0.25</b>	-	-	0.005	0.11	0.006	0.09
<b>0.5</b>	0.011	0.21	0.012	0.22	0.015	0.19
<b>0.75</b>	-	-	0.023	0.33	0.03	0.31
<b>0.9</b>	0.017	0.42	0.035	0.41	0.05	0.4
<b>0.99</b>	0.017	0.47	-	-	-	-
<b>Global</b>	0.004	0.47	0.01	0.41	0.012	0.4

	Global	
	Med.	Max.
$HR_1 = HR_2$	0.002	0.22
$ HR_1 - HR_2  \leq 0.1$	0.002	0.07
$0.1 <  HR_1 - HR_2  \leq 0.25$	0.007	0.26
$0.25 <  HR_1 - HR_2  \leq 0.5$	0.011	0.37
$0.5 <  HR_1 - HR_2  \leq 0.75$	0.015	0.39
$ HR_1 - HR_2  > 0.75$	0.017	0.47
<b>Global</b>	0.007	0.47

**Figure 5.6:** Illustration of the range of  $HR_*(t)$  (top left) and histogram (top right). Descriptive results are split in tables according to the values of the correlation  $\rho$  and differences between the marginal hazard ratios ( $|HR_1 - HR_2|$ ). Med. and Max. stand for median and maximum values, respectively. All the numbers are  $> 0$  although we do not specify all the decimals in order to make the tables easier to read. Values used in the top-left plot:  $(HR_1, HR_2) = (0.5, 0.95)$ ,  $p_1 = p_2 = 0.1$ ,  $\beta_1 = \beta_2 = 2$ ,  $\rho = 0.9$  (Case 3).



## 5.4 Example from the cardiovascular area

The Losartan Intervention For Endpoint reduction in hypertension study (LIFE) trial (Dahlöf et al., 2002)<sup>10</sup> was performed to test the efficacy of Losartan-based antihypertensive treatment in patients with hypertension. The primary composite endpoint was composed by cardiovascular death, myocardial infarction and stroke. Cardiovascular death + myocardial infarction are considered the most clinically important components (Sankoh et al., 2014),<sup>62</sup> and hence we refer to them as the relevant endpoint  $\mathcal{E}_R$ , and stroke as the additional endpoint  $\mathcal{E}_A$  (see Table 5.6).

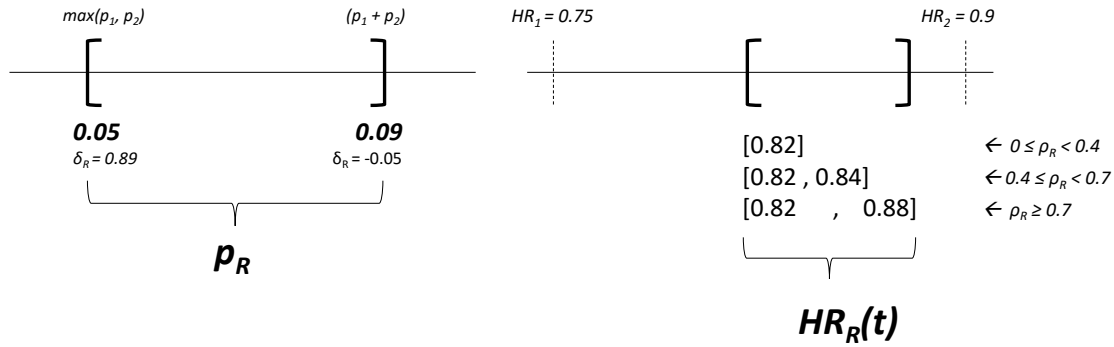
Endpoint	Type	Control treatment Probability (n)	Hazard ratio (CI, 95%)
Cardiovasc. mortality	$\mathcal{E}_1$	0.05 (234)	0.89 (0.73 – 1.07)
Myocardial infarction	$\mathcal{E}_2$	0.04 (188)	1.07 (0.88 – 1.31)
Stroke	$\mathcal{E}_A$	0.07 (309)	0.75 (0.63 – 0.89)
$\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_A$	Composite	0.13 (588)	0.87 (0.77 – 0.98)

**Table 5.6:** Probabilities in control group and hazard ratios of each endpoint from the LIFE trial.  $\mathcal{E}_1$  and  $\mathcal{E}_2$  stand for relevant endpoint component 1 and 2 respectively;  $\mathcal{E}_A$  stands for additional endpoint; and CI stands for the confidence interval.

The probability of observing the two components of the relevant endpoint  $\mathcal{E}_1$  and  $\mathcal{E}_2$  in control group is  $p_1 = 0.05$  and  $p_2 = 0.04$ , respectively. We assume exponential distributions for each component and a hazard ratio of Cardiovascular death and Myocardial infarction of  $HR_1 = 0.75$  and  $HR_2 = 0.9$ , respectively.

Following the calculations described in this chapter for the LIFE study, we observe that the plausible values for the combined probability, in terms of the correlation, would range between 0.05 and 0.09 (see Figure 5.7, left), and the possible values of the combined  $HR_R(t)$  would range from 0.82 to 0.88 (see Figure 5.7, right).

Whenever the investigator cannot anticipate the correlation, we can consider to use either the lowest, the average or the highest plausible values of  $p_R$  and  $HR_R(t)$ . We illustrate the results combining these possibilities in Table 5.7 which includes the Asymptotic Relative Efficiency (*ARE*) values for each scenario. Based on the *ARE* method, we observe that the decision of using the composite endpoint is always recommended since the *ARE* is always greater than 1. However, note that the required sample size can triple if we compare the smallest value of *ARE* (2.12) versus the highest (6.28).



**Figure 5.7:** Boundaries for the combined probability  $p_R$  and hazard ratio  $HR_R(t)$  in terms of Pearson's correlation  $\delta_R$  and Spearman's correlation  $\rho_R$ , respectively.  $p_1$  and  $p_2$  stand for the marginal probability of observing the relevant component  $\mathcal{E}_1$  and  $\mathcal{E}_2$  in control group, respectively.  $HR_1$  and  $HR_2$  stand for the marginal hazard ratio of each component. We assume constant hazard rates (exponential distributions) for each component. Component parameter values are taken from the LIFE study.

$p_R$	$HR_R(t)$	ARE
0.05	0.82	3.34
0.05	0.85	4.42
0.05	0.88	6.28
0.07	0.82	2.55
0.07	0.85	3.26
0.07	0.88	4.48
0.09	0.82	2.12
0.09	0.85	2.64
0.09	0.88	3.52

**Table 5.7:** Asymptotic Relative Efficiency (ARE) by different values of the probability  $p_R$  in control group and hazard ratio  $HR_R(t)$  of the relevant endpoint. Exponential distribution and moderate correlation ( $\rho_{RA} = 0.5$ ) is assumed between the relevant endpoint and additional endpoint (do not confuse with  $\rho_R$ , which corresponds to the correlation between the components of the relevant endpoint). With respect to the additional endpoint, we fix the probability in control group  $p_A = 0.07$  and the hazard ratio  $HR_A = 0.75$ .

A letter to the editor in response of Sankoh's paper is published. We discuss other scenarios which would have led to different results of ARE depending on different parameter combinations (Gómez and Gómez-Mateu, 2016).<sup>23</sup>

## 5.5 Conclusion

The *ARE* method is based on the assumption that, even when the relevant endpoint consist of several components, the combined probability in control group and the combined hazard ratio can be anticipated by researchers. Analogously, the same is assumed with respect to the additional endpoint. Since investigators may know the anticipated probabilities and hazard ratios of each one of the components rather than the combined probabilities and hazard ratios, we have discussed in this chapter how to derive these quantities based on the marginal components.

Based on binary data, the combined probability  $p_*$  can be calculated in terms of the probability  $p_1$  and  $p_2$  of observing the components  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , respectively, given the Pearson's correlation  $\delta$  between endpoints. It is also possible to calculate the combined probability values based on time-to-event variables. However, the calculation of the combined probability is based on several assumptions and on some parameter values that trialists might not know. It is for that reason that we would recommend to assign the combined  $p_*$  in terms of the Pearson correlation, except for Case 4, since  $p_*$  is straightforwardly calculated as  $p_* = p_1 + p_2$ . Whenever investigators cannot anticipate the correlation, we propose to assign a value among the plausible range of values that the combined probability can take as it is developed in 5.1.

Under some assumptions, the combined  $HR_*(t)$  can be derived in terms of the marginal hazard ratios  $HR_1$  and  $HR_2$ , the probabilities of observing  $\mathcal{E}_1$  and  $\mathcal{E}_2$  in control group, and the correlation  $\rho$  between endpoint times. Whenever the combined  $HR_*(t)$  is constant over time and investigators can anticipate all the required parameter values, a  $HR_*(t)$  can be specifically calculated. In particular, when having null correlation ( $\rho = 0$ ), the combined hazard ratio is constant and lies between  $HR_1$  and  $HR_2$ . In some specific scenarios, the combined  $HR_*(t)$  corresponds to the mean of  $HR_1$  and  $HR_2$ .

In cases where the  $HR_*(t)$  is not constant over time or trialist cannot anticipate all the parameter values, we propose to calculate the range of plausible values that the  $HR_*(t)$  can take as it has been described in this chapter. Among these range of values, trialists might have to assign a specific  $HR_*(t)$  that they consider appropriate based on their experience. We have seen that the interval range of values of  $HR_*(t)$  given by the general interval  $[HR_1 - 0.07, HR_2 + 0.05]$  ( $HR_1 \leq HR_2$ ) would cover almost all of the scenarios considered in our work. Whenever  $HR_1 \neq HR_2$  or for low correlations, this interval is narrower.

The more information that investigators can anticipate (such as the correlation), the more precise the assignment of the combined hazard ratio will be. We briefly summarize in table 5.8 the values and range of values of  $HR_*(t)$  depending on several parameters.

Special attention must be taken for scenarios with high correlation because they might lead to large departure from constancy of  $HR_*(t)$  and consequently they will have an important impact in sample size assessment. As a matter of fact, even small departures from constancy might have a considerable impact on sample size. We are working on that issue and it is detailed in the future research of this thesis.

We have implemented in *CompARE*, as advanced options, the practical solutions developed in this chapter. When possible, *CompARE* will calculate the specific values  $p_*$  and  $HR_*(t)$  when having several components in the relevant endpoint or in the additional endpoint. Whenever a parameter value, such as the correlation, is unknown, or whenever the  $HR_*(t)$  is not constant over time, *CompARE* will recommend a range of plausible values of  $p_*$  and  $HR_*(t)$ . Hence, *CompARE* will help investigators to choose a plausible value of  $p_*$  and  $HR_*(t)$  to calculate the corresponding *ARE* values for the choice of the primary endpoint of a clinical trial.

Whenever $\rho = 0$	$HR_*(t)$
$\frac{\lambda_2^{(0)}(t)}{\lambda_1^{(0)}(t)} = c$ $HR_1 = HR_2 = k$	$HR_*(t) = \frac{HR_1 + c \cdot HR_2}{c + 1}$
Weibull with $\beta_1 = \beta_2 = \beta$	$HR_*(t) = \frac{k + HR_1 + HR_2 \left( \frac{b_1^{(0)}}{b_2^{(0)}} \right)^\beta}{1 + \left( \frac{b_1^{(0)}}{b_2^{(0)}} \right)^\beta}$
Weibull with $\beta_1 = \beta_2 = \beta$ & $p_1 = p_2$	$HR_*(t) = \frac{HR_1 + HR_2}{2}$
Whenever $\rho > 0$	Range of values of $HR_*(t)$ *
Global $HR_1 = HR_2 = k$	$[HR_1 - 0.07, HR_2 + 0.05]$ $[k - 0.08, k + 0.06]$
$\rho \leq 0.1$	$[HR_1 - 0.02, HR_2 + 0.01]$
$0.25 \leq \rho \leq 0.5$	$[HR_1 - 0.07, HR_2 + 0.04]$
$0.75 \leq \rho \leq 0.99$	$[HR_1 - 0.08, HR_2 + 0.06]$

**Table 5.8:** Combined hazard ratio  $HR_*(t)$  depending on the parameter values.  $HR_1, HR_2, p_1, p_2, \beta_1, \beta_2$  and  $b_1^{(0)}, b_2^{(0)}$  stand for the hazard ratio, probability in control group, shape and scale (control group) parameters of endpoints  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , respectively.  $\rho$  stands for the correlation between the two endpoints. \* Corresponding intervals in 99% of scenarios executed in section 5.3 (that is, until percentile 99 for each cut). We consider  $HR_1 \leq HR_2$  to determine the intervals; analogously, the same applies for  $HR_1 \geq HR_2$ .

## APPLIED EXTENSIONS IN COMPARE

Throughout the previous chapters, we have seen that the efficiency of a composite endpoint can be evaluated in terms of several parameter values. In some cases, changing one of these parameters might influence the choice of the primary endpoint and might also have influence in the required sample size of the trial. Furthermore, some of the needed values to anticipate might be unknown by the investigators, specially when consisting of a union of several marginal components.

In the current chapter, we have extended *CompARE* to accommodate the comparison of distinct scenarios that could represent other realistic situations in the design of a clinical trial. First, the user can choose different marginal laws for the time to each endpoint, different degrees of correlation as well as different copulas. Second, *CompARE* has been extended to compute *ARE* values in cases where both the relevant endpoint (RE) and the additional endpoint (AE) include a terminating event.

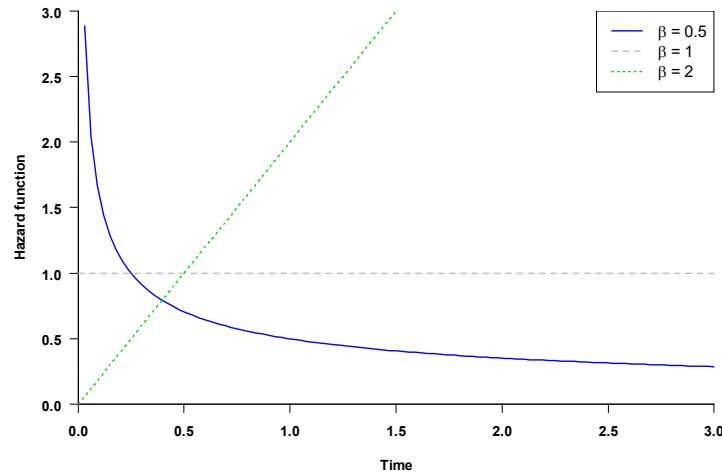
Third, *CompARE* has been also extended to quantify specific values for the combined probability and hazard ratios whenever it comes from a combination of several components. When the user cannot anticipate some of the needed parameters, *CompARE* provides a range of plausible values. Moreover, the departure from constancy of a combined hazard ratio, which might have a strong influence in treatment effect interpretation and in sample size assessment, can be explored by visualizing its shape over time for different scenarios. Finally, sample size computations based on the chosen scenarios are implemented as well in *CompARE*.



## 6.1 Extended options to change parameter values

By default, *CompARE* assumes exponential distributions for the time to the occurrence of the relevant endpoint ( $T_R$ ) and for the time to the additional endpoint ( $T_A$ ), moderate correlation ( $\rho = 0.5$ ) and Frank's copula relationship between times. We next detail how to modify these assumptions.

We have extended *CompARE* to allow different distributions, other than exponential, whenever investigators assume increasing or decreasing hazards of endpoints over the follow-up period of the trial (see Figure 6.1).



**Figure 6.1:** Hazard rate functions from Weibull distributions  $W(\beta, b)$  with increasing ( $\beta = 2$ ), constant ( $\beta = 1$ ) and decreasing shape parameter ( $\beta = 0.5$ ) and scale parameter  $b = 1$ .

From the drop-down menu in the advanced features box, the user can choose between the following options (see Figure 6.2, top):

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

where  $\beta$  is the shape parameter and the scale parameter  $b$  is derived automatically. The density and survival functions used in *CompARE* are parametrized as follows:

$$f(t) = \frac{\beta}{(b)^\beta} t^{\beta-1} e^{-(t/b)^\beta}$$

$$S(t) = e^{-(t/b)^\beta}.$$

### Advanced Features (Optional)

[ ]

	Terminating?*	Probability*	Hazard Ratio*	Shape parameter of the Weibull Distribution
Combined Relevant endpoint	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	Decreasing Hazard Rate ( $\beta: 0.5$ )
Combined Additional endpoint	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	
Correlation				Decreasing Hazard Rate ( $\beta: 0.5$ ) Constant Hazard Rate ( $\beta: 1$ ) (Exponential) Increasing Hazard Rate ( $\beta: 2$ )
Copula				

\* Only when the Relevant/Additional endpoint have several components.

	Terminating?*	Probability*	Hazard Ratio*	Shape parameter of the Weibull Distribution
Combined Relevant endpoint	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	
Combined Additional endpoint	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	
Correlation				
Copula				Very Strong ( $\rho: 0.9$ ) Strong ( $\rho: 0.7$ ) Moderate ( $\rho: 0.5$ ) Weak ( $\rho: 0.3$ ) Very Weak ( $\rho: 0.15$ ) No correlation ( $\rho: 0$ )

\* Only when the Relevant/Additional endpoint have several components.

Remove executions history?

	Terminating?*	Probability*	Hazard Ratio*	Shape parameter of the Weibull Distribution
Combined Relevant endpoint	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	
Combined Additional endpoint	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	
Correlation				
Copula				Frank Gumbel Clayton FGM Normal T Galambos HuslerReiss Tawn Tev Plackett

\* Only when the Relevant/Additional endpoint have several components.

Remove executions history?

Figure 6.2



As we have seen previously in this thesis, the correlation might play a crucial role. Trialist might need to compare the efficiency of one endpoint over the other in terms of the strength of the association between the relevant endpoint and the additional endpoint. Therefore, we have incorporated in *CompARE* the possibility to change Spearman's correlations as follows (see figure 6.2, middle):

- Very Strong ( $\rho = 0.9$ ),
- Strong ( $\rho = 0.7$ ),
- Moderate ( $\rho = 0.5$ ),
- Weak ( $\rho = 0.3$ ),
- Very Weak ( $\rho = 0.15$ ),
- No correlation ( $\rho = 0$ ),

where  $\rho$  indicates the Spearman correlation coefficient.

We have also implemented an extension of the methodology to allow to other copulas binding the relevant and the additional endpoint. The methodology was extended by Plana-Ripoll O. and Gómez G.<sup>51</sup> That is, the joint distribution of  $(T_R, T_A)$  can be derived in *CompARE* by means of the following copulas (see Figure 6.2, bottom):

- Frank,
- Gumbel,
- Clayton,
- GM,
- Normal,
- T,
- Galambos,
- HuslerReiss,
- Tawn,
- Tev,
- Plackett.

## 6.2 Computation of HR and probability for combined endpoints

CompARE is also extended to accommodate the computation of combined probabilities and combined hazard ratios based on the marginal components.

When the RE or the AE consist of several components, the user can specify each marginal endpoint in the input grid (see Figure 6.3, top). In this example, the union of cardiovascular death  $\mathcal{E}_{R1}$  and myocardial infarction  $\mathcal{E}_{R2}$  forms the relevant endpoint. The marginal probabilities and hazard ratios of each component  $\mathcal{E}_{R1}$  and  $\mathcal{E}_{R2}$  are (0.05, 0.75) and (0.04, 0.9), respectively. By default, *CompARE* uses the maximum probability and the average hazard ratio to calculate the *ARE* values (see chapter 5 for the rationale of this default choice). Alternatively, *CompARE* proposes a range of plausible values of the corresponding combined probability and combined hazard ratio (for the latter case, only implemented for two components so far). Based on this range of values, the user can introduce a specific value for the combined parameters by means of the advanced features box (see Figure 6.3, bottom).

### Information about all the candidate endpoints for your trial



(You can modify the parameter values and run it again)

Candidate endpoint $E$	Terminating? (click if yes)	Probability of observing $E$ in control group	Hazard Ratio	Type of endpoint	Definition of the composite	
CV death	<input checked="" type="checkbox"/>	0.05	0.75	Relevant component	<input checked="" type="checkbox"/>	
Myocardial infarction	<input type="checkbox"/>	0.04	0.9	Relevant component	<input checked="" type="checkbox"/>	
Stroke	<input type="checkbox"/>	0.07	0.75	Additional component	<input checked="" type="checkbox"/>	
	<input type="checkbox"/>				<input type="checkbox"/>	Add Rows? Image
	<input type="checkbox"/>				<input type="checkbox"/>	<input type="checkbox"/>

#### Advanced Features (Optional)

[+]

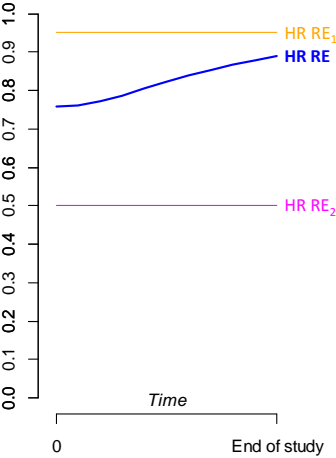
	Terminating?*	Probability*	Hazard Ratio*	Shape parameter of the Weibull Distribution
Combined Relevant endpoint	Yes	0.06	0.8	
Combined Additional endpoint				
Correlation				
Copula				

\* Only when the Relevant/Additional endpoint have several components.

Figure 6.3

In order to visually evaluate the departure from constancy that the combined hazard ratio might have, *CompARE* depicts the shape of the hazard ratio over time (see Figure 6.4)

**Marginal and combined hazard ratio of the relevant endpoint**



**Figure 6.4:** Marginal constant hazard ratios ( $HR RE_1$ ) and ( $HR RE_2$ ) of the relevant endpoint, and combined hazard ratio ( $HR RE$ ) over time.

### 6.3 Sample size

*CompARE* has also been designed to incorporate sample size computation. We start noticing that in survival analysis the power of the test depends on the number of events rather than on the needed sample size. The sample size, as we develop below, depends heavily on the allocation rate of patients into the RCT and the patterns of censoring including the loss to follow-up.

#### Required number of events

We define  $\alpha$  as the significance level, and  $\Pi = 1 - \beta$  as the statistical power, where  $\beta$  is the type II error. We assume proportional hazards for the marginal hazard ratio of the relevant endpoint RE, i.e.  $HR_R(t) = HR_R$ , and we consider the logrank test to compare treatment effects. We assume that the end-of-study censoring at time  $\tau$  is the only censoring cause in both groups.

Based on the asymptotic behaviour of the logrank statistic, Schoenfeld<sup>65</sup> estimate the required number of events  $e_{R,S}$  given by:

$$e_{R,S} = \frac{4(z_\alpha + z_\beta)^2}{(\ln(HR_R))^2},$$

where  $z_\alpha$  and  $z_\beta$  are the standard normal quantiles corresponding to the left tail probability  $\alpha$  and  $\beta$ , respectively.

Another approach was developed by Freedman.<sup>18</sup> The number of events  $e_{R,F}$  is estimated by means of the expected value and variance of the logrank statistic, and assuming that the ratio of the number of patients at risk in each group just before the event is equal to 1. The expression of  $e_{R,F}$  is given by:

$$e_{R,F} = \frac{(HR_R + 1)^2}{(HR_R - 1)^2} \left( z_\alpha + z_\beta \frac{2\sqrt{HR_R}}{HR_R + 1} \right)^2.$$

Taking the coefficient  $\frac{2\sqrt{HR_R}}{HR_R + 1}$  as approximately equal to 1 (it approaches 1 when  $HR \rightarrow 1$ ), the expression becomes

$$e_{R,F} = \frac{(HR_R + 1)^2 (z_\alpha + z_\beta)^2}{(HR_R - 1)^2}.$$

#### Required number of patients

If we assume that patients are randomized to receive one of the two treatments in the ratio (1 : A), for example treatment group triples the control group ( $A = 3$ ), the required number of patients is derived as follows (Machin, 1997):<sup>43</sup>

$$N_R = \frac{(1 + A)e_R}{p_R^{(0)} + A(p_R^{(1)})},$$

where  $e_R$  is the required number of events; and  $p_R^{(0)}$ ,  $p_R^{(1)}$  are the probability of observing the relevant endpoint in control and treatment group, respectively. Thus, the required number of patients to be recruited to the control and treatment group is  $N_R^{(0)} = N_R/(1+A)$  and  $N_R^{(1)} = N_RA/(1+A)$ , respectively. If we assume equal allocation between groups, that is  $A = 1$ , the total number of required patients  $N_R$  is given by:

$$N_R = \frac{2e_R}{p_R^{(0)} + p_R^{(1)}}.$$

Censoring due to loss to follow-up might affect the estimation of the required number of patients. Given an anticipated withdrawal proportion  $W$ , the required number of patients must be increased to:

$$N_{R,W} = \frac{N_R}{(1-W)}.$$

Different accrual rates and the duration of the trial will also affect sample size estimation and will be developed in future research.

In a recent article published by Abel et al. (2015),<sup>1</sup> the authors compare Schoenfeld and Freedman approaches. They state that, in terms of number of events, Schoenfeld's formula would be preferable because  $e_{R,S} < e_{R,F}$ . Indeed, the ratio  $r = e_{R,S}/e_{R,F}$  may be considerably lower than 1 for low and high (above 1) values of  $HR_R$ . Whenever  $HR_R$  approaches 1 (low treatment effect), the ratio  $r$  approaches 1, and hence both formulas would lead to similar number of events. The simulations included in their paper show that special attention must be taken whenever we have small sample sizes. In those cases, Schoenfeld's formula would lead to an overestimation of power, while the Freedman approach underestimates the power by almost the same absolute amount. Furthermore, loss of power may occur in case of deviations from the assumed distributions.

We remark that the formulas to calculate the number of events are based on the assumption that the  $HR_R(t)$  is constant over time. Whenever we do not have proportional hazards, the previous formulas are not appropriate.

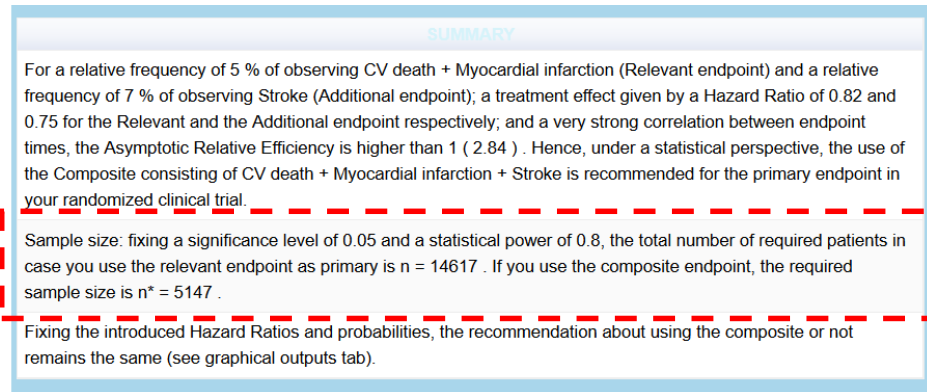
#### Sample size for the composite endpoint

Assume that the hazard ratio of the additional endpoint is constant, i.e.  $HR_A(t) = HR_A$ . As commented in Chapter 3, given the asymptotic relative efficiency  $ARE(Z_*, Z_R)$  of the logrank statistics  $Z_*$  and  $Z_R$  of the composite endpoint and the RE, respectively, the required number of patients  $N_*$  for the CE can be estimated as:

$$N_* = \frac{N_R}{ARE(Z_*, Z_R)}.$$

Based on the previous formulas, *CompARE* computes the sample size for each specific scenario and shows the results written in text (see Figure 6.5). The user can modify the parameter values to evaluate how the sample size changes depending on the different value sets. By default, *CompARE* uses

Schoenfeld's approach, equal allocation and no withdrawals to make the computations. We are extending the platform to derive sample sizes whenever the user considers other choices, and it is described in chapter 7 as future research.



*Figure 6.5*

## 6.4 Computations when both the relevant and additional endpoints include terminating events

As already mentioned in *Chapter 4*, before designing the platform *CompARE*, the code to execute the *ARE* method was written in MAPLE by Gómez G. to run scenarios where both the RE and AE do not include a terminating event (Case 1) or either RE or AE includes a terminating event (Cases 2 and 3). As part of my thesis work, I have adapted the MAPLE code to *R* and I have programmed, from scratch, the code for Case 4, where both the RE and the AE includes a terminating event.

The extension to Case 4 has not been straightforward because of the simultaneous solution of the following double integrals to derive the scale parameters  $b_R^{(0)}$  and  $b_A^{(0)}$  in control group for the RE and the AE, respectively:

$$\left. \begin{aligned} p_R &= \int_{UL}^1 \int_0^{VL^{(0)}(x)} g(x, y) dy dx \\ p_A &= \int_{VL}^1 \int_0^{UL^{(0)}(y)} g(x, y) dx dy, \end{aligned} \right\}$$

where

- $p_R$  and  $p_A$  are the probabilities of observing the RE and the AE in control group, respectively,
- $UL = \exp\{-(b_R^{(0)})^{-\beta_R}\}$ ,
- $VL^{(0)}(x) = \exp\left\{-\frac{(b_R^{(0)})^{\beta_A}(-\log x)^{\beta_A/\beta_R}}{(b_A^{(0)})^{\beta_A}}\right\}$ ,
- $VL = \exp\{-(b_A^{(0)})^{-\beta_A}\}$ , and
- $UL^{(0)}(y) = \exp\left\{-\frac{(b_A^{(0)})^{\beta_R}(-\log y)^{\beta_R/\beta_A}}{(b_R^{(0)})^{\beta_R}}\right\}$ .

The function  $g(x, y)$  is given by:

$$g(x, y) = \frac{\theta(1 - e^{-\theta})\exp\{-\theta(x + y)\}}{(e^{-\theta} + e^{-\theta(x+y)} - e^{-\theta x} - e^{-\theta y})^2},$$

where  $\theta$  the association parameter between  $T_R$  and  $T_A$ .

The code in *R* used to compute the *ARE* values is included in the Appendix.





## FUTURE RESEARCH

New lines of research are open after the development of this thesis. As a member of the research groups GRBIO<sup>33</sup> and GRASS,<sup>32</sup> we are collaborating with other partners and researchers of other groups in order to achieve new goals that might contribute in the improvement of trial designs.

### 7.1 Extension of the ARE method

The methodology of Gómez and Lagakos is based on several assumptions described in *Chapter 1*. We aim to relax some of these assumptions and to study the robustness of the method when some of the assumptions do not hold.

The lifetimes of the outcomes are assumed to be distributed as Weibull, because they are widely used in survival analysis. However, other distributions with different features could be applied. For example, the lognormal distribution has been applied successfully in cancer studies (like in chronic leukemia<sup>40</sup>) due to its pronounced right asymmetry. Other distributions like the log-logistic are sometimes preferred since the survival function is mathematically more tractable than the lognormal.<sup>37</sup>

So far, the method has been developed under the assumption that the end-of-study time is the only non-informative cause of censoring. Moreover, it is also assumed that the censoring is the same in both groups. Other types of censoring, even different in each group, could be implemented, such as that derived from loss to follow-up or withdrawals.

Equal Spearman's rank correlation  $\rho$  is assumed between the marginal times of the relevant and the additional endpoints for both groups. This assumption might sometimes lead to unlikely scenarios for some parameter value combinations that might not be realistic for the design of the study. We want to relax this assumption to analyze how the *ARE* values would change whenever the correlations are not equal.

The *ARE* methodology is also being extended to dichotomous variables (Gómez and Ayestaran, 2014;<sup>22</sup> Gómez, Gómez-Mateu and Bofill, 2016<sup>25</sup>). We want to extend the methodology to binary data with more than two components, perform an exhaustive literature search in several fields, and carry out simulations to confirm the interpretation of the *ARE* as the ratio of sample sizes for finite sample sizes.

Observational studies are commonly conducted whenever the exposure of interest cannot be randomized, such as smoking, or the goal of the trial is to detect a rare or late adverse treatment effect.<sup>78</sup> The *ARE* method can be extended to observational cohort trials, as it has been discussed in a recent paper (Gómez, Plana-Ripoll and Dafni, 2016).<sup>29</sup> They apply the methodology to a cohort of adults with coronary heart disease to study the effect of concurrent depression and stress. We plan to accommodate different censoring patterns and extend further the *ARE* method for observational studies.

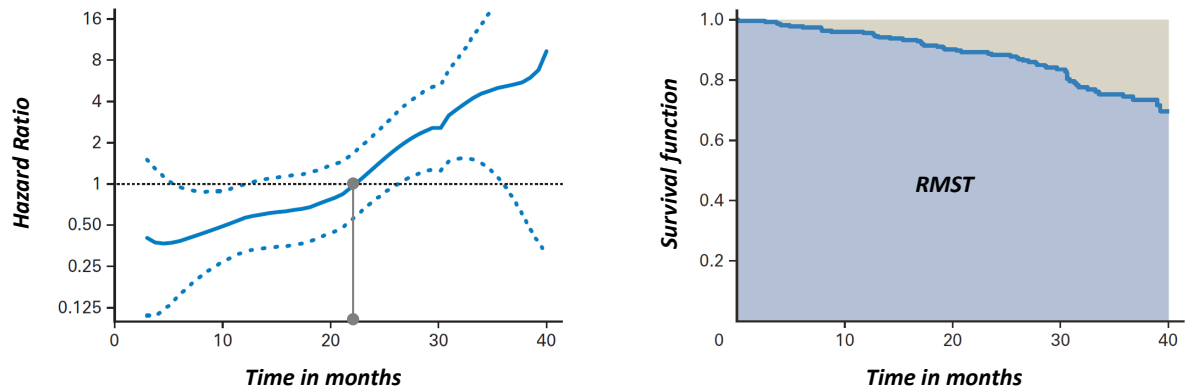
## 7.2 Non-proportional hazards and alternative measures

As we showed in *Chapter 5*, most of the times the hazard ratio of a composite endpoint is not constant over time. Based on the *ARE* method, from the sequence of contiguous alternatives to  $H_0$  when using composite endpoints given by  $H_{1,n} : \lambda_{R,n}^{(1)}(t) = \lambda_R^{(0)}(t)e^{g(t)/\sqrt{n}}$ , we have that  $g(t) \approx \log\left(\frac{\lambda_{R,n}^{(1)}(t)}{\lambda_R^{(0)}(t)}\right)\sqrt{n}$  is proportional to the log of the hazard ratio for sufficiently large  $n$ . Thus,  $g(t)$  allows us to evaluate the behaviour of the *ARE* when the hazard functions for the two groups are non-proportional.

The non-proportionality of the hazards may lead to difficulties in its interpretation as a measure of treatment effect. It may also have a great influence in *ARE* computations and hence, in sample size. Together with professors G. Gómez and K. Kim (University of Wisconsin-Madison), we are studying the different patterns and shapes that the composite hazard ratio  $HR_*(t)$  follows when it is not constant over time for realistic clinical trial scenarios. A study on the changes in sample size depending on how  $HR_*(t)$  departs from constancy will be of great help for the design of clinical trials.

We will also analyze the impact of non-proportionality on sample size whenever the investigator erroneously assumes constant hazard ratio. We will simulate scenarios where the  $HR_*(t)$  is clearly non constant and compare the required sample size using the commonly used formulas such as Schoenfeld's versus the corresponding sample size using the *ARE* method.

Alternative measures to assess treatment effects not relying on the proportionality of the hazards have been proposed (Uno et al, 2015).<sup>75</sup> The absolute and relative risk difference, derived from the difference in event rates at some fixed time point, provides a clinically interpretable comparison between groups. It is an appropriate measure for quantifying relatively long-term survival benefits. The difference and ratio of percentiles can also be considered as another plausible measure, since it has a simple mathematical interpretation. However, except for the median, it may not be as much intuitively interpretable to investigators or patients. Furthermore, whenever the event rate is relatively low, the censoring is high, or the follow-up is short, this measure might not be estimable from the observed data. Another alternative measure is the restricted mean survival time (RMST), which provides a clin-



**Figure 7.1:** Estimated hazard ratio (low dose over high dose) over time (left) with the corresponding 0.95 point-wise confidence band, and restricted mean survival time (right, blue area) with data from the Eastern Cooperative Oncology Group E4A03 study<sup>81</sup> up to 40 months for the low-dose group. Adapted from Figure 1 in Uno H, 2014.<sup>74</sup>

ically meaningful summary of treatment effect and gives more stable estimates than the differences of percentiles. It is defined as the expected event-free survival time experienced during a specified time, and it is estimated by the area under the survival curve up to that time point (see Figure 7.1, right). The RMST of a random variable  $T$  is the mean of the survival time  $Y = \min(T, \tau)$  limited to the time-to-follow up  $\tau$  of the study (Royston, 2013).<sup>61</sup> Its expression is given by:

$$RMST = E(Y) = \int_0^{\tau} S(t) dt,$$

where  $S(t)$  is the survival function.

Alternatively, we could consider the restricted mean time lost (RMTL), corresponding to the area above the survival curve up to  $\tau$ . It can be interpreted as the life lost during the follow-up of the study.

The Restricted Average Log Hazard ratio (RALH) can also be considered as an alternative measure of the effect. Its expression derives from the sample size  $n_*$  required to achieve a given power for a significance level  $\alpha$ . For Cases 1 and 3 (the additional endpoint is not terminating),  $n_*$  is given by:

$$n_* = \frac{1}{\pi(1-\pi)} \cdot \frac{(z_\alpha + z_\beta)^2 p_*^{(0)}}{RALH^2},$$

where  $\pi$  is the percentage of patients allocated to the control group,  $p_*^{(0)}$  is the probability of observing the CE in control group,  $z_\alpha$  and  $z_\beta$  are the standard normal quantiles corresponding to the left tail probability  $\alpha$  and type II error  $\beta$ , respectively, and RALH is given by

$$RALH = \frac{\int_0^1 \log(HR_*(t)) f_*^{(0)}(t) dt}{p_*^{(0)}},$$

where  $f_*^{(0)}(t)$  is the density function of the time  $T_*$  to the CE in control group.

We consider to use the logrank test specifying the null and alternative hypothesis in terms of the RALH and derive the needed sample size based on that measure.

We plan to carry out a research based on the convenience of using the aforementioned measures, specially when the proportionality of the hazards does not hold. We will study the impact of the non proportionality of the HR versus the rest of measures for plausible scenarios with considerable departure from constancy.

The time to follow up  $\tau$  is also crucial to assess treatment effect, and it does not affect equally on the results for each measure. Results based on HR might show different effects depending on  $\tau$ . As an illustrative example, in Figure 7.1 (left) we see that the estimated HR remains below 1 during the first 22 months approximately, but it shows no treatment effect afterwards. If we use RMST, the effect on each arm will always increase or keep constant over time, as it is shown in Figure 7.1 (right). That is, the longer  $\tau$ , the longer effect might exist on each group. By means of simulated scenarios, we are studying how different prespecified  $\tau$ 's might influence the treatment effect assessment for different measures, including scenarios with hazard ratios fluctuating over 1.

### 7.3 Composite endpoints in other medical fields

In the last years, the use of progression-free survival (PFS), defined as the union of death and progression, has increased in non-small cell lung cancer studies (Booth and Eisenhauer, 2012).<sup>5</sup> Progression, usually considered as the additional endpoint, may have a strong weight in the calculation of the *ARE* values, and hence in sample size calculations. However, information about progression in randomized clinical trials is almost never reported.

Together with professors G. Gómez and U. Dafni (University of Athens), we are carrying out a literature search based on the lung cancer field, in a similar way that we did in our cardiovascular research described in *Chapter 2*, to give general recommendations about the convenience of using PFS, taking into consideration the plausible values that progression can have. We will also analyze the impact in sample size and primary endpoint decision depending on variations of the parameter values from the study, such as the correlation between endpoints. Some case studies will be discussed in depth as well.

Furthermore, we will study alternative methods involving terminating and non-terminating events. As an example, Potthoff and Halabi<sup>53</sup> develop a test based on the time to death if the patient has died or the time to death based on a pre-specified model otherwise.

Other areas of interest involving the use of composite endpoints, such as HIV-AIDS and oncology studies, will be considered to our research as well.

## 7.4 Future extensions in CompARE

We are working to develop the following extensions in *CompARE* platform:

- We plan to extend computations of sample size. It will incorporate the choice of different accrual rates, unequal allocation, the duration of the trial and loss to follow-up for different values of the significance level  $\alpha$  to achieve a specific power.
- Computations based on different approaches, such as Schoenfeld's or Freedman's, will be implemented.
- We want as well to allow different distributional assumptions. On the one hand, we will allow the user to specify any value for the shape parameter  $\beta$  of the marginal Weibull distributions, other than the fixed by default (0.5, 1 and 2), to make the computations. On the other hand, other choices of distributions will be implemented as well on the platform.
- Based on the methodology for binary data, we will also adapt *CompARE* to make the computations for clinical trials where the primary endpoint is dichotomous. The platform will also be accommodated to deal with data from observational studies.
- We will create a new library in *R* from the code to compute the *ARE* values so as *CompARE* calls and keep all the functions updated throughout the on-line repository.
- Other written outputs in text will be incorporated to guide the decision depending on the aforementioned extensions. Moreover, graphical results will be improved by using other applications such as *Shiny*,<sup>67</sup> which allows to build interactive web applications straight from *R*, and includes the use of dynamical plots and instantaneous interaction with the user.



CHAPTER 

**REFERENCES**





## REFERENCES

- [1] Abel UR, Jensen K, Karapanagiotou-Schenkel I, Kieser M. Some Issues of Sample Size Calculation for Time-to-Event Endpoints Using the Freedman and Schoenfeld Formulas. *J Biopharm Stat.* 2015; 25(6):1285–311.
- [2] Amir E, Seruga B, Kwong R, Tannock IF, Ocaña A. Poor correlation between progression-free and overall survival in modern clinical trials: are composite endpoints the answer? *Eur J Cancer.* 2012; 48(3):385-8.
- [3] Ayestaran A. How to select between a primary dichotomous endpoint and a composite one in HIV Randomized Clinical Trials? Theoretical bases and application. 2014. Master thesis (Master's degree in Statistics and Operations Research -UPC).
- [4] Bahadur RR. A representation of the joint distribution of responses to n dichotomous items. *Stanford University Press.* 1961; 158–168.
- [5] Booth CM, Eisenhauer EA. Progression-free survival: meaningful or simply measurable? *J Clin Oncol.* 2012 Apr 1; 30(10):1030–3.
- [6] Chi GY. Some issues with composite endpoints in clinical trials. *Fundam Clin Pharmacol.* 2005; 19:609–619.
- [7] Ciszewski A, Bilinska ZT, Brydak LB, Kepka C, Kruk M, Romanowska M, Ksiezzycka E, Przulski J, Piotrowski W, Maczynska R, Ruzyllo W. Influenza vaccination in secondary prevention from coronary ischaemic events in coronary artery disease: FLUCAD study. *Eur Heart J.* 2008; 29:1350–1358.
- [8] Consolidated Standards of Reporting Trials (CONSORT). <http://www.consort-statement.org>.
- [9] Cox DR. Regression Models and Life-tables. *Journal of the Royal Statistical Society.* 1972; 34(2), 187–220.
- [10] Dahlöf B, Devereux RB, Kjeldsen SE, Julius S, Beevers G, de Faire U, Fyhrquist F, Ibsen H, Kristiansson K, Lederballe-Pedersen O, Lindholm LH, Nieminen MS, Omvik P, Oparil S, Wedel H; LIFE Study Group. Cardiovascular morbidity and mortality in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol. *Lancet.* 2002; 359:995–1003.
- [11] de Pedro X, Sánchez A. Usando de forma segura R vía web con Tiki. *II Jornadas de Usuarios de R en Castellano. Mieres, Oviedo, Spain.* 2010. <http://estbioinfo.stat.ub.es/wp-content/uploads/2010/12/RJ-II-Jornadas-R-ES-XavierdePedro.pdf>.
- [12] Editorial: Should protocols for observational research be registered? *The Lancet.* 2010; 375:348.
- [13] European network for Health Technology Assessment (EUnetHTA). Endpoints used for relative effectiveness assessment of pharmaceuticals. Composite endpoints. 2013.
- [14] Ferreira-González I, Busse JW, Heels-Ansdell D, Montori VM, Akl E, Bryant DM, Alonso-Coello P, Alonso J, Worster A, Upadhye S, Jaeschke R, Schünemann HJ, Permyer-Miralda G, Pacheco-Huergo V, Domingo-Salvany A, Wu P, Mills EJ, Guyatt GH. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ.* 2007; 334(7597):786.
- [15] Ferreira-González I, Permyer-Miralda G, Busse JW, Bryant DM, Montori VM, Alonso-Coello P, Walter SD, Guyatt GH. Composite endpoints in clinical trials: the trees and the forest. *Journal of Clinical Epidemiology.* 2007; 60(10.1016):660–661.

- [16] Ferreira-González I, Permanyer-Miralda G, Busse JW, Bryant DM, Montori VM, Alonso-Coello P, Walter SD, Guyatt GH. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *J Clin Epidemiol*. 2007; 60(7):651–7; discussion 658–62.
- [17] Free Software Foundation (FSF). <https://www.fsf.org>.
- [18] Freedman LS. Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine*. 1982; 1, 121–129.
- [19] Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA*. 2003; 289(19):2554–2559.
- [20] Freemantle N, Calvert M. Weighing the pros and cons for composite outcomes in clinical trials. *Journal of Clinical Epidemiology*. 2007; 60:658–659.
- [21] Gómez G. Some theoretical thoughts when using a composite endpoint to prove the efficacy of a treatment. *International Workshop on Statistical Modelling*. 2011; Proceedings of the 26th International Workshop on Statistical Modelling, 14–21.
- [22] Gómez G, Ayestaran A. Selection of the primary endpoint In a clinical trial: composite binary endpoints versus single outcomes. *Proceedings of the 2015 International Conference on Risk Analysis (ICRA6-RISK2015)*, Barcelona, 26-29 May 2015.
- [23] Gómez G, Gómez-Mateu M. Comments on "Use of composite endpoints in clinical trials" by Abdul J. Sankoh, Haihong Li and Ralph B. D'Agostino, Sr. *Statistics in Medicine*. 2016; 35, 317–318.
- [24] Gómez G, Gómez-Mateu M. The Asymptotic Relative Efficiency and the ratio of sample sizes when testing two different null hypotheses. *SORT*. 2014; 38, 73–88.
- [25] Gómez G, Gómez-Mateu M, Bofill M. Planning clinical trials involving composite endpoints. *Oral presentation in the XXXVI Congreso Nacional de Estadística e Investigación Operativa*, 5-7 September 2016.
- [26] Gómez G, Gómez-Mateu M, Dafni U. Informed Choice of Composite End Points in Cardiovascular Trials. *Circulation. Cardiovascular Quality and Outcomes*. 2014; 7, 170–178.
- [27] Gómez G, Lagakos SW. Statistical considerations when using a composite endpoint for comparing treatment groups. *Statistics in Medicine*. 2013; 32, 719–738.
- [28] Gómez-Mateu M, Gómez G. CompARE platform. <https://cinna.upc.edu/compare>. (Last date of access: May 11, 2016).
- [29] Gómez G, Plana-Ripoll O, Dafni U. Selection of the primary endpoint in an observational cohort study. *Journal of Epidemiology and Community Health*. 2016 (accepted for publication).
- [30] Gissi-HF Investigators, Tavazzi L, Maggioni AP, Marchioli R, Barlera S, Franzosi MG, Latini R, Lucci D, Nicolosi GL, Porcu M, Tognoni G. Effect of n-3 polyunsaturated fatty acids in patients with chronic heart failure (the GISSI-HF trial): a randomised, double-blind, placebo-controlled trial. *The Lancet*. 2008; 372:1223–30.
- [31] Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine*. 1999; 18(6):695–706.
- [32] Grup de Recerca en Anàlisi eStadística de la Supervivència (GRASS). <https://grass.upc.edu>.
- [33] Grup de Recerca en Bioestadística i Bioinformàtica (GRBIO). <https://genwebv4.upc.edu/grbio/en>.
- [34] Hirsch A, Windhausen F, Tijssen JG, Oude Ophuis AJ, van der Giessen WJ, van der Zee PM, Cornel JH, Verheugt FW, de Winter RJ; Invasive versus Conservative Treatment in Unstable coronary Syndromes Investigators. Diverging associations of an intended early invasive strategy compared with actual revascularization, and outcome in patients with non-ST-segment elevation acute coronary syndrome: the problem of treatment selection bias. *Eur Heart J*. 2009; 30:645–654.

- 
- [35] Huque ME, Alesh M, Bhole R. Addressing multiplicity issues of a composite endpoint and its components in clinical trials. *J Biopharm Stat.* 2011; 21:610–634.
- [36] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Efficacy Guidelines. [www.ich.org/products/guidelines/efficacy/article/efficacy-guidelines.html](http://www.ich.org/products/guidelines/efficacy/article/efficacy-guidelines.html).
- [37] Klein JP, Moeschberger ML. Survival Analysis: Techniques for Censored and Truncated Data. *Springer*. 1997. ISBN: 978-1-4757-2730-2.
- [38] Kleist P. Composite Endpoints for Clinical Trials. Current Perspectives. *International Journal of Pharmaceutical Medicine*. 2007; Volume 21, Issue 3, pp 187–198.
- [39] Kleist P. Composite Endpoints: Proceed with Caution. *Applied Clinical Trials online*. 2006. [www.appliedclinicaltrials.com/composite-endpoints-proceed-caution](http://www.appliedclinicaltrials.com/composite-endpoints-proceed-caution).
- [40] Lee ET. Statistical methods for survival data analysis. Second edition. *Wiley*. 1992. ISBN: 0-471-61592-7.
- [41] Lehmann EL, Romano JP. Testing Statistical Hypotheses. *Springer*. 2005. ISBN 978-0-387-98864-1.
- [42] Lindner AM, Szimayer A. A limit theorem for copulas. *Sonderforschungsbereich*. 2005; 86, Paper 433.
- [43] Machin D, Campbell MJ, Fayers PM, Pinol APY. Sample size tables for Clinical Studies. *Blackwell Science*. 1997. ISBN: 0-86542-870-0.
- [44] Meinert CL. Clinical Trials Dictionary: Terminology and Usage Recommendations. 2nd Edition. *Wiley*. 2012. ISBN: 978-1-118-29515-1.
- [45] Montori VM, Permyer-Miralda G, Ferreira-González I, Busse JW, Pacheco-Huergo V, Bryant D, Alonso J, Akl EA, Domingo-Salvany A, Mills E, Wu P, Schünemann HJ, Jaeschke R, Guyatt GH. Validity of composite end points in clinical trials. *BMJ*. 2005; 12;330(7491):594–6.
- [46] Neaton JD, Gray G, Zuckerman BD, Konstam MA. Key issues in endpoint selection for heart failure trials: composite end points. *J Card Fail*. 2005; 11:567–575.
- [47] Nelsen RB. An Introduction to Copulas. *Springer Series in Statistics*. 2006. ISBN 978-0-387-28678-5.
- [48] Noether GE. On a theorem of Pitman. *Ann. Math. Statist.* 1954; 25, 514–522.
- [49] Ogawa H, Nakayama M, Morimoto T, Uemura S, Kanauchi M, Doi N, Jinnouchi H, Sugiyama S, Saito Y; Japanese Primary Prevention of Atherosclerosis With Aspirin for Diabetes (JPAD) Trial Investigators. Low-dose aspirin for primary prevention of atherosclerotic events in patients with type 2 diabetes: a randomized controlled trial. *JAMA*. 2008; 300:2134–2141.
- [50] Oldgren J, Wallentin L, Afzal R, Bassand JP, Budaj A, Chrolavicius S, Fox KA, Granger CB, Mehta SR, Pais P, Peters RJ, Xavier D, Zhu J, Yusuf S; OASIS-6 Investigators. Effects of fondaparinux in patients with ST-segment elevation acute myocardial infarction not receiving reperfusion treatment. *Eur Heart J*. 2008; 29:315–323.
- [51] Plana-Ripoll O, Gómez G. Selecting the primary endpoint in a randomized clinical trial. The ARE method. *Journal of Biopharmaceutical Statistics*. 2015; 23:1–19.
- [52] Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J*. 2012; 33(2):176–82.
- [53] Potthoff RE, Halabi S. A novel test to compare two treatments based on endpoints involving both nonfatal and fatal events. *Pharm Stat*. 2015 Jul; 14(4): 273–283.
- [54] Rauch G, Jahn-Eimermacher A, Brannath W, Kiesera M. Opportunities and challenges of combined effect measures based on prioritized outcomes. *Statist. Med.* 2014; 33, 1104–1120.
- [55] Rauch G, Kieser M. An expected power approach for the assessment of composite endpoints and its components. *Computational Statistics & Data Analysis*. 2013; 60: 111–122.

- [56] Rauch G, Kieser M, Ulrich S, Doherty P, Rauch B, Schneider S, Riemer T, Senges J. Competing time-to-event endpoints in cardiology trials: a simulation study to illustrate the importance of an adequate statistical analysis. *European Journal of Preventive Cardiology*. 2014; 21(1):74–80.
- [57] Rauch G, Wirths M, Kieser M. Consistency-adjusted alpha allocation methods for a time-to-event analysis of composite endpoints. *Computational Statistics & Data Analysis*. 2014; 75 151–161.
- [58] Reynolds SJ, Makumbi F, Newell K, Kiwanuka N, Ssebowa P, Mondo G, Boaz I, Wawer MJ, Gray RH, Serwadda D, Quinn TC. Effect of daily aciclovir on HIV disease progression in individuals in Rakai, Uganda, co-infected with HIV-1 and herpes simplex virus type 2: a randomised, double-blind placebo-controlled trial. *Lancet Infect Dis*. 2012; 12(6):441–8.
- [59] Ross S. Composite outcomes in randomized clinical trials: arguments for and against. *Am J Obstet Gynecol*. 2007; 196(2):119.e1–6.
- [60] Roy D, Talajic M, Nattel S, Wyse DG, Dorian P, Lee KL, Bourassa MG, Arnold JM, Buxton AE, Camm AJ, Connolly SJ, Dubuc M, Ducharme A, Guerra PG, Hohnloser SH, Lambert J, Le Heuzey JY, O'Hara G, Pedersen OD, Rouleau JL, Singh BN, Stevenson LW, Stevenson WG, Thibault B, Waldo AL; Atrial Fibrillation and Congestive Heart Failure Investigators. Rhythm control versus rate control for atrial fibrillation and heart failure. *N Engl J Med*. 2008; 358:2667–2677.
- [61] Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol*. 2013 Dec 7; 13:152.
- [62] Sankoh AJ, Li H, D'Agostino RB. Use of composite endpoints in clinical trials. *Statistics in Medicine*. 2014; 33:4709–4714.
- [63] Sapir R. Tiki Essentials. What every Smarty needs to know about Tiki Wiki CMS Groupware. 2010. *Under a Creative Commons Attribution-Share Alike 3.0 License*.
- [64] Schoenfeld D. Sample-size formula for the proportional-hazards regression model. *Biometrics*. 1983; 39(2):499–503.
- [65] Schoenfeld D. The Asymptotic Properties of Nonparametric Tests for Comparing Survival Distributions. *Biometrika*. 1981; Vol. 68, No. 1, pp. 316–319.
- [66] Schweizer B, Wolff EF. On Nonparametric Measures of Dependence for Random Variables. *The Annals of Statistics*. 1981; Vol. 9, No. 4, pp. 879–885.
- [67] Shiny, by Rstudio. A web application framework for R. <http://shiny.rstudio.com>.
- [68] Sozu T. Sample size determination in clinical trials with multiple co-primary binary endpoints. *Statist. Med*. 2010; Vol 29, pp. 2169–2179.
- [69] Stone GW, Witzenbichler B, Guagliumi G, Peruga JZ, Brodie BR, Dudek D, Kornowski R, Hartmann F, Gersh BJ, Pocock SJ, Dangas G, Wong SC, Kirtane AJ, Parise H, Mehran R; HORIZONS-AMI Trial Investigators. Bivalirudin during primary PCI in acute myocardial infarction. *N Engl J Med*. 2008; 358:2218–2230.
- [70] Tardif JC, McMurray JJV, Klug E, Small R, Schumi J, Choi J, Cooper J, Scott R, Lewis EF, L'Allier PL, Pfeffer MA; Aggressive Reduction of Inflammation Stops Events (ARISE) Trial Investigators. Effects of succinobucol (AGI-1067) after an acute coronary syndrome: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2008; 371(9626), 1761–1768.
- [71] Tiki Documentation: PluginR. <https://doc.tiki.org/PluginR>.
- [72] Tomlinson G, Detsky AS. Composite end points in randomized trials: there is no free lunch. *JAMA*. 2010; 20;303:267–8.
- [73] Trivedi PK, Zimmer DM. Copula Modeling: An Introduction for Practitioners. *Foundations and Trends in Econometrics*. 2005; 1, pp 1–111.
- [74] Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, Schrag D, Takeuchi M, Uyama Y, Zhao L, Skali H, Solomon S, Jacobus S, Hughes M, Packer M, Wei LJ. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol*. 2014; 32(22):2380–5.

- 
- [75] Uno H, Wittes J, Fu H, Solomon SD, Claggett B, Tian L, Cai T, Pfeffer MA, Evans SR, Wei LJ. Alternatives to Hazard Ratios for Comparing the Efficacy or Safety of Therapies in Noninferiority Studies. *Ann Intern Med.* 2015; 163(2):127-34.
- [76] van Leth F, Lange JM. Use of composite end points to measure clinical events. *JAMA.* 2003; 17;290(11):1456-7; author reply 1457.
- [77] Vandembroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M; STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med.* 2007; 4:e297.
- [78] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandembroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ.* 2007; 335:806-8.
- [79] Wiki CMS groupware. <http://info.tiki.org>.
- [80] Wittkop L, Smith C, Fox Z, Sabin C, Richert L, Aboulker JP, Phillips A, Chêne G, Babiker A, Thiébaud R; NEAT-WP4. Methodological issues in the use of composite endpoints in clinical trials: examples from the HIV field. *Clin Trials.* 2010; 7(1):19-35.
- [81] Zukin M, Barrios CH, Pereira JR, Ribeiro Rde A, Beato CA, do Nascimento YN, Murad A, Franke FA, Precivale M, Araujo LH, Baldotto CS, Vieira FM, Small IA, Ferreira CG, Lilenbaum RC. Randomized phase III trial of single-agent pemetrexed versus carboplatin and pemetrexed in patients with advanced non-small-cell lung cancer and Eastern Cooperative Oncology Group performance status of 2. *J Clin Oncol.* 2013; 31(23):2849-53.





## APPENDIX: LIST OF REFERENCES FROM LITERATURE SEARCH IN CHAPTER 2

Duckworth, W.; Abaira, C.; Moritz, T.; Reda, D.; Emanuele, N.; Reaven, P. D.; Zieve, F. J.; Marks, J.; Davis, S. N.; Hayward, R.; Warren, S. R.; Goldman, S.; McCarren, M.; Vitek, M. E.; Henderson, W. G.; Huang, G. D. & Investigators, V. A. D. T. (2009), 'Glucose control and vascular complications in veterans with type 2 diabetes.', *N Engl J Med* 360(2), 129–139.

Hirsch, A.; Windhausen, F.; Tijssen, J. G. P.; Ophuis, A. J. M. O.; van der Giessen, W. J.; van der Zee, P. M.; Cornel, J. H.; Verheugt, F. W. A.; de Winter, R. J. & versus Conservative Treatment in Unstable coronary Syndromes Investigators, I. (2009), 'Diverging associations of an intended early invasive strategy compared with actual revascularization, and outcome in patients with non-ST-segment elevation acute coronary syndrome: the problem of treatment selection bias.', *Eur Heart J* 30(6), 645–654.

Menon, V.; Pearte, C. A.; Buller, C. E.; Steg, P. G.; Forman, S. A.; White, H. D.; Marino, P. N.; Katritsis, D. G.; Caramori, P.; Lasevitch, R.; Lobo-Grudzien, K.; Zurakowski, A.; Lamas, G. A. & Hochman, J. S. (2009), 'Lack of benefit from percutaneous intervention of persistently occluded infarct arteries after the acute phase of myocardial infarction is time independent: insights from Occluded Artery Trial.', *Eur Heart J* 30(2), 183–191.

ACCORD; Gerstein, H. C.; Miller, M. E.; Byington, R. P.; Goff, D. C.; Bigger, J. T.; Buse, J. B.; Cushman, W. C.; Genuth, S.; Ismail-Beigi, E.; Grimm, R. H.; Probstfield, J. L.; Simons-Morton, D. G. & Friedewald, W. T. (2008), 'Effects of intensive glucose lowering in type 2 diabetes.', *N Engl J Med* 358(24), 2545–2559.

ADVANCE; Patel, A.; MacMahon, S.; Chalmers, J.; Neal, B.; Billot, L.; Woodward, M.; Marre, M.; Cooper, M.; Glasziou, P.; Grobbee, D.; Hamet, P.; Harrap, S.; Heller, S.; Liu, L.; Mancia, G.; Mogensen, C. E.; Pan, C.; Poulter, N.; Rodgers, A.; Williams, B.; Bompoint, S.; de Galan, B. E.; Joshi, R. & Travert, F. (2008), 'Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes.', *N Engl J Med* 358(24), 2560–2572.

Albert, C. M.; Cook, N. R.; Gaziano, J. M.; Zaharris, E.; MacFadyen, J.; Danielson, E.; Buring, J. E. & Manson, J. E. (2008), 'Effect of folic acid and B vitamins on risk of cardiovascular events and total mortality among women at high risk for cardiovascular disease: a randomized trial.', *JAMA* 299(17), 2027–2036.

AMADEUS; Bousser, M. G.; Bouthier, J.; Büller, H. R.; Cohen, A. T.; Crijns, H.; Davidson, B. L.; Halperin, J.; Hankey, G.; Levy, S.; Pengo, V.; Prandoni, P.; Prins, M. H.; Tomkowiak, W.; Torp-Pedersen, C.; Thorp-Pedersen, C. & Wyse, D. G. (2008), 'Comparison of idraparinux with vitamin K antagonists for prevention of thromboembolism in patients with atrial fibrillation: a randomised, open-label, non-inferiority trial.', *Lancet* 371(9609), 315–321.

Baron, J. A.; Sandler, R. S.; Bresalier, R. S.; Lanus, A.; Morton, D. G.; Riddell, R.; Iverson, E. R. & Demets, D. L. (2008), 'Cardiovascular events associated with rofecoxib: final analysis of the APPROVe trial.', *Lancet* 372(9651), 1756–1764.

Busk, M.; Maeng, M.; Rasmussen, K.; Kelbaek, H.; Thayssen, P.; Abildgaard, U.; Vigholt, E.; Mortensen, L. S.; Thuesen, L.; Kristensen, S. D.; Nielsen, T. T.; Andersen, H. R. & Investigators, D. A. N. A. M. I.-2. (2008), 'The Danish multicentre randomized study of fibrinolytic therapy vs. primary angioplasty in acute myocardial infarction (the DANAMI-2 trial): outcome after 3 years follow-up.', *Eur Heart J* 29(10), 1259–1266.

Ciszewski, A.; Bilinska, Z. T.; Brydak, L. B.; Kepka, C.; Kruk, M.; Romanowska, M.; Ksiezyccka, E.; Przulski, J.; Piotrowski, W.; Maczynska, R. & Ruzyllo, W. (2008), 'Influenza vaccination in secondary prevention from coronary ischaemic events in coronary artery disease: FLUCAD study.', *Eur Heart J* 29(11), 1350–1358.

Cohen-Solal, A.; McMurray, J. J. V.; Swedberg, K.; Pfeffer, M. A.; Puu, M.; Solomon, S. D.; Michelson, E. L.; Yusuf, S.; Granger, C. B. & Investigators, C. H. A. R. M. (2008), 'Benefits and safety of candesartan treatment in heart failure are independent of age: insights from the Candesartan in Heart failure—Assessment of Reduction in Mortality and morbidity programme.', *Eur Heart J* 29(24), 3022–3028.

Ebbing, M.; Øyvind Bleie; Ueland, P. M.; Nordrehaug, J. E.; Nilsen, D. W.; Vollset, S. E.; Refsum, H.; Pedersen, E. K. R. & Nygård, O. (2008), 'Mortality and cardiovascular events in patients treated with homocysteine-lowering B vitamins after coronary angiography: a randomized controlled trial.', *JAMA* 300(7), 795–804.

Ellis, S. G.; O'Shaughnessy, C. D.; Martin, S. L.; Kent, K.; McGarry, T.; Turco, M. A.; Kereiakes, D. J.; Popma, J. J.; Friedman, M.; Koglin, J.; Stone, G. W. & Investigators, T. A. X. U. S. V. I. (2008), 'Two-year clinical outcomes after paclitaxel-eluting stent or brachytherapy treatment for bare metal stent restenosis: the TAXUS V ISR trial.', *Eur Heart J* 29(13), 1625–1634.

Ellis, S. G.; Tendera, M.; de Belder, M. A.; van Boven, A. J.; Widimsky, P.; Janssens, L.; Andersen, H. R.; Betriu, A.; Savonitto, S.; Adamus, J.; Peruga, J. Z.; Kosmider, M.; Katz, O.; Neunteufl, T.; Jorgova, J.; Dorobantu, M.; Grinfeld, L.; Armstrong, P.; Brodie, B. R.; Herrmann, H. C.; Montalescot, G.; Neumann, F.-J.; Effron, M. B.; Barnathan, E. S.; Topol, E. J. & Investigators, F. I. N. E. S. S. E. (2008), 'Facilitated PCI in patients with ST-elevation myocardial infarction.', *N Engl J Med* 358(21), 2205–2217.

Ferenc, M.; Gick, M.; Kienzle, R.-P.; Bestehorn, H.-P.; Werner, K.-D.; Comberg, T.; Kuebler, P.; Büttner, H. J. & Neumann, F.-J. (2008), 'Randomized trial on routine vs. provisional T-stenting in the treatment of de novo coronary bifurcation lesions.', *Eur Heart J* 29(23), 2859–2867.

Fox, K.; Ford, I.; Steg, P. G.; Tendera, M.; Ferrari, R. & Investigators, B. E. A. U. T. I. F. U. L. (2008), 'Ivabradine for patients with stable coronary artery disease and left-ventricular systolic dysfunction (BEAUTIFUL): a randomised, double-blind, placebo-controlled trial.', *Lancet* 372(9641), 807–816.

GALA; Lewis, S. C.; Warlow, C. P.; Bodenham, A. R.; Colam, B.; Rothwell, P. M.; Torgerson, D.; Dellagrammaticas, D.; Horrocks, M.; Liapis, C.; Banning, A. P.; Gough, M. & Gough, M. J. (2008), 'General anaesthesia versus local anaesthesia for carotid surgery (GALA): a multicentre, randomised controlled trial.', *Lancet* 372(9656), 2132–2142.

Galløe, A. M.; Thuesen, L.; Kelbaek, H.; Thayssen, P.; Rasmussen, K.; Hansen, P. R.; Bligaard, N.; Saunamäki, K.; Junker, A.; Aarøe, J.; Abildgaard, U.; Ravkilde, J.; Engstrøm, T.; Jensen, J. S.; Andersen, H. R.; Bøtker, H. E.; Galatius, S.; Kristensen, S. D.; Madsen, J. K.; Krusell, L. R.; Abildstrøm, S. Z.; Stephansen, G. B.; Lassen, J. F. & Investigators, S. O. R. T. O. I. (2008), 'Comparison of paclitaxel- and sirolimus-eluting stents in everyday clinical practice: the SORT OUT II randomized trial.', *JAMA* 299(4), 409–416.

Ghali, J. K.; Anand, I. S.; Abraham, W. T.; Fonarow, G. C.; Greenberg, B.; Krum, H.; Massie, B. M.; Wasserman, S. M.; Trotman, M.-L.; Sun, Y.; Knusel, B.; Armstrong, P. & of Anemia in Heart Failure Trial (STAMINA-HeFT) Group, S. (2008), 'Randomized double-blind trial of darbepoetin alfa in patients with symptomatic heart failure and anemia.', *Circulation* 117(4), 526–535.



Gissi-H.; Tavazzi, L.; Maggioni, A. P.; Marchioli, R.; Barlera, S.; Franzosi, M. G.; Latini, R.; Lucci, D.; Nicolosi, G. L.; Porcu, M. & Tognoni, G. (2008), 'Effect of n-3 polyunsaturated fatty acids in patients with chronic heart failure (the GISSI-HF trial): a randomised, double-blind, placebo-controlled trial.', *Lancet* 372(9645), 1223–1230.

Glineur, D.; Hanet, C.; Poncelet, A.; D'hoore, W.; Funken, J.-C.; Rubay, J.; Kefer, J.; Astarci, P.; Lacroix, V.; Verhelst, R.; Etienne, P. Y.; Noirhomme, P. & Khoury, G. E. (2008), 'Comparison of bilateral internal thoracic artery revascularization using in situ or Y graft configurations: a prospective randomized clinical, functional, and angiographic midterm evaluation.', *Circulation* 118(14 Suppl), S216–S221.

Gray, A.; Goodacre, S.; Newby, D. E.; Masson, M.; Sampson, F.; Nicholl, J. & Trialists, 3C. P. O. (2008), 'Noninvasive ventilation in acute cardiogenic pulmonary edema.', *N Engl J Med* 359(2), 142–151.

Greenberg, B.; Czerska, B.; Delgado, R. M.; Bourge, R.; Zile, M. R.; Silver, M.; Klapholz, M.; Haeusslein, E.; Mehra, M. R.; Mather, P.; Abraham, W. T.; Neaton, J. D.; Brown, B. S.; Parker, I. C.; Konstam, M. A.; Investigators, M. O. M. E. N. T. U. M. & Coordinators (2008), 'Effects of continuous aortic flow augmentation in patients with exacerbation of heart failure inadequately responsive to medical therapy: results of the Multicenter Trial of the Orqis Medical Cancion System for the Enhanced Treatment of Heart Failure Unresponsive to Medical Therapy (MOMENTUM).', *Circulation* 118(12), 1241–1249.

Gurm, H. S.; Yadav, J. S.; Fayad, P.; Katzen, B. T.; Mishkel, G. J.; Bajwa, T. K.; Ansel, G.; Strickman, N. E.; Wang, H.; Cohen, S. A.; Massaro, J. M.; Cutlip, D. E. & Investigators, S. A. P. P. H. I. R. E. (2008), 'Long-term results of carotid stenting versus endarterectomy in high-risk patients.', *N Engl J Med* 358(15), 1572–1579.

Jamerson, K.; Weber, M. A.; Bakris, G. L.; Dahlöf, B.; Pitt, B.; Shi, V.; Hester, A.; Gupte, J.; Gatlin, M.; Velazquez, E. J. & Investigators, A. C. C. O. M. P. L. I. S. H. T. (2008), 'Benazepril plus amlodipine or hydrochlorothiazide for hypertension in high-risk patients.', *N Engl J Med* 359(23), 2417–2428.

Kakkar, A. K.; Brenner, B.; Dahl, O. E.; Eriksson, B. I.; Mouret, P.; Muntz, J.; Soglian, A. G.; Pap, A. F.; Misselwitz, F.; Haas, S. & Investigators, R. E. C. O. R. D. (2008), 'Extended duration rivaroxaban versus short-term enoxaparin for the prevention of venous thromboembolism after total hip arthroplasty: a double-blind, randomised controlled trial.', *Lancet* 372(9632), 31–39.

Kastrati, A.; Neumann, F.-J.; Mehilli, J.; Byrne, R. A.; Iijima, R.; Büttner, H. J.; Khattab, A. A.; Schulz, S.; Blankenship, J. C.; Pache, J.; Minners, J.; Seyfarth, M.; Graf, I.; Skelding, K. A.; Dirschinger, J.; Richardt, G.; Berger, P. B.; Schömig, A. & Investigators, I. S. A. R.-R. E. A. C. T. 3. T. (2008), 'Bivalirudin versus unfractionated heparin during percutaneous coronary intervention.', *N Engl J Med* 359(7), 688–696.

Kelbaek, H.; Thuesen, L.; Helqvist, S.; Clemmensen, P.; Kløvgaard, L.; Kaltoft, A.; Andersen, B.; Thuesen, H.; Engstrøm, T.; Bøtker, H. E.; Saunamäki, K.; Krusell, L. R.; Jørgensen, E.; Hansen, H.-H. T.; Christiansen, E. H.; Ravkilde, J.; Køber, L.; Kofoed, K. F.; Terkelsen, C. J.; Lassen, J. F. & Investigators, D. E. D. I. C. A. T. I. O. N. (2008), 'Drug-eluting versus bare metal stents in patients with st-segment-elevation myocardial infarction: eight-month follow-up in the Drug Elution and Distal Protection in Acute Myocardial Infarction (DEDICATION) trial.', *Circulation* 118(11), 1155–1162.

Kolloch, R.; Legler, U. F.; Champion, A.; Cooper-Dehoff, R. M.; Handberg, E.; Zhou, Q. & Pepine, C. J. (2008), 'Impact of resting heart rate on outcomes in hypertensive patients with coronary artery disease: findings from the INTERNATIONAL Verapamil-SR/trandolapril Study (INVEST).', *Eur Heart J* 29(10), 1327–1334.

Køber, L.; Torp-Pedersen, C.; McMurray, J. J. V.; Gøtzsche, O.; Lévy, S.; Crijns, H.; Amlie, J.; Carlsen, J. & Group, D. S. (2008), 'Increased mortality after dronedarone therapy for severe heart failure.', *N Engl J Med* 358(25), 2678–2687.

LIMB; Leizorovicz, A. & Becker, F. (2008), 'Oral buflomedil in the prevention of cardiovascular events in patients with peripheral arterial obstructive disease: a randomized, placebo-controlled, 4-year study.', *Circulation* 117(6), 816–822.

Mario, C. D.; Dudek, D.; Piscione, F.; Mielecki, W.; Savonitto, S.; Murena, E.; Dimopoulos, K.; Manari, A.; Gaspardone, A.; Ochala, A.; Zmudka, K.; Bolognese, L.; Steg, P. G.; Flather, M. & in A. M. I. (Combined Abciximab RE-teplase Stent Study in Acute Myocardial Infarction) Investigators, C. A. R. E. S. S. (2008), 'Immediate angioplasty versus standard therapy with rescue angioplasty after thrombolysis in the Combined Abciximab RE-teplase Stent Study in Acute Myocardial Infarction (CARESS-in-AMI): an open, prospective, randomised, multicentre

trial., *Lancet* 371(9612), 559–568.

Massie, B. M.; Carson, P. E.; McMurray, J. J.; Komajda, M.; McKelvie, R.; Zile, M. R.; Anderson, S.; Donovan, M.; Iverson, E.; Staiger, C.; Ptaszynska, A. & Investigators, I.-P. R. E. S. E. R. V. E. (2008), 'Irbesartan in patients with heart failure and preserved ejection fraction.', *N Engl J Med* 359(23), 2456–2467.

Mehta, S. R.; Boden, W. E.; Eikelboom, J. W.; Flather, M.; Steg, P. G.; Avezum, A.; Afzal, R.; Piegas, L. S.; Faxon, D. P.; Widimsky, P.; Budaj, A.; Chrolavicius, S.; Rupprecht, H.-J.; Jolly, S.; Granger, C. B.; Fox, K. A. A.; Bassand, J.-P.; Yusuf, S.; 5, O. A. S. I. S. & Investigators, 6. (2008), 'Antithrombotic therapy with fondaparinux in relation to interventional management strategy in patients with ST- and non-ST-segment elevation acute coronary syndromes: an individual patient-level combined analysis of the Fifth and Sixth Organization to Assess Strategies in Ischemic Syndromes (OASIS 5 and 6) randomized trials.', *Circulation* 118(20), 2038–2046.

MEND-CABG.; Alexander, J. H.; Emery, R. W.; Carrier, M.; Ellis, S. J.; Mehta, R. H.; Hasselblad, V.; Menasche, P.; Khalil, A.; Cote, R.; Bennett-Guerrero, E.; Mack, M. J.; Schuler, G.; Harrington, R. A. & Tardif, J.-C. (2008), 'Efficacy and safety of pyridoxal 5'-phosphate (MC-1) in high-risk patients undergoing coronary artery bypass graft surgery: the MEND-CABG II randomized clinical trial.', *JAMA* 299(15), 1777–1787.

Murphy, S. A.; Antman, E. M.; Wiviott, S. D.; Weerakkody, G.; Morocutti, G.; Huber, K.; Lopez-Sendon, J.; McCabe, C. H.; Braunwald, E. & Investigators, T. R. I. T. O. N.-T. I. M. I. 38. (2008), 'Reduction in recurrent cardiovascular events with prasugrel compared with clopidogrel in patients with acute coronary syndromes from the TRITON-TIMI 38 trial.', *Eur Heart J* 29(20), 2473–2479.

Ndrepepa, G.; Kastrati, A.; Mehilli, J.; Neumann, F.-J.; ten Berg, J.; Bruskina, O.; Dotzer, F.; Seyfarth, M.; Pache, J.; Dirschinger, J.; Berger, P. B. & Schömig, A. (2008), 'One-year clinical outcomes with abciximab vs. placebo in patients with non-ST-segment elevation acute coronary syndromes undergoing percutaneous coronary intervention after pre-treatment with clopidogrel: results of the ISAR-REACT 2 randomized trial.', *Eur Heart J* 29(4), 455–461.

Nissen, S. E.; Nicholls, S. J.; Wolski, K.; Rodés-Cabau, J.; Cannon, C. P.; Deanfield, J. E.; Després, J.-P.; Kastelein, J. J. P.; Steinhubl, S. R.; Kapadia, S.; Yasin, M.; Ruzyllo, W.; Gaudin, C.; Job, B.; Hu, B.; Bhatt, D. L.; Lincoff, A. M.; Tuzcu, E. M. & Investigators, S. T. R. A. D. I. V. A. R. I. U. S. (2008), 'Effect of rimonabant on progression of atherosclerosis in patients with abdominal obesity and coronary artery disease: the STRADIVARIUS randomized controlled trial.', *JAMA* 299(13), 1547–1560.

Ogawa, H.; Nakayama, M.; Morimoto, T.; Uemura, S.; Kanauchi, M.; Doi, N.; Jinnouchi, H.; Sugiyama, S.; Saito, Y. & of Atherosclerosis With Aspirin for Diabetes (JPAD) Trial Investigators, J. P. P. (2008), 'Low-dose aspirin for primary prevention of atherosclerotic events in patients with type 2 diabetes: a randomized controlled trial.', *JAMA* 300(18), 2134–2141.

Oldgren, J.; Wallentin, L.; Afzal, R.; Bassand, J.-P.; Budaj, A.; Chrolavicius, S.; Fox, K. A. A.; Granger, C. B.; Mehta, S. R.; Pais, P.; Peters, R. J. G.; Xavier, D.; Zhu, J.; Yusuf, S. & Investigators, O. A. S. I. S.-6. (2008), 'Effects of fondaparinux in patients with ST-segment elevation acute myocardial infarction not receiving reperfusion treatment.', *Eur Heart J* 29(3), 315–323.

ONTARGET; Yusuf, S.; Teo, K. K.; Pogue, J.; Dyal, L.; Copland, I.; Schumacher, H.; Dagenais, G.; Sleight, P. & Anderson, C. (2008), 'Telmisartan, ramipril, or both in patients at high risk for vascular events.', *N Engl J Med* 358(15), 1547–1559.

Pitt, B.; Bakris, G.; Ruilope, L. M.; DiCarlo, L.; Mukherjee, R. & Investigators, E. P. H. E. S. U. S. (2008), 'Serum potassium and clinical outcomes in the Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study (EPHESUS).', *Circulation* 118(16), 1643–1650.

POISE; Devereaux, P. J.; Yang, H.; Yusuf, S.; Guyatt, G.; Leslie, K.; Villar, J. C.; Xavier, D.; Chrolavicius, S.; Greenspan, L.; Pogue, J.; Pais, P.; Liu, L.; Xu, S.; Málaga, G.; Avezum, A.; Chan, M.; Montori, V. M.; Jacka, M. & Choi, P. (2008), 'Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial.', *Lancet* 371(9627), 1839–1847.

Ridker, P. M.; Danielson, E.; Fonseca, F. A. H.; Genest, J.; Gotto, A. M.; Kastelein, J. J. P.; Koenig, W.; Libby, P.; Lorenzatti, A. J.; MacFadyen, J. G.; Nordestgaard, B. G.; Shepherd, J.; Willerson, J. T.; Glynn, R. J. & Group, J. U. P. I. T. E. R. S. (2008), 'Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein.', *N Engl J Med* 359(21), 2195–2207.

Rossebo, A. B.; Pedersen, T. R.; Boman, K.; Brudi, P.; Chambers, J. B.; Egstrup, K.; Gerds, E.; Gohlke-Bärwolf, C.; Holme, I.; Kesäniemi, Y. A.; Malbecq, W.; Nienaber, C. A.; Ray, S.; Skjaerpe, T.; Wachtell, K.; Willenheimer, R. & Investigators, S. E. A. S. (2008), 'Intensive lipid lowering with simvastatin and ezetimibe in aortic stenosis.', *N Engl J Med* 359(13), 1343–1356.

Rouleau, J. L.; Warnica, W. J.; Baillot, R.; Block, P. J.; Chocron, S.; Johnstone, D.; Myers, M. G.; Calciu, C.-D.; Dalle-Ave, S.; Martineau, P.; Mormont, C.; van Gilst, W. H. & with Accupril post-bypass Graft via Inhibition of the coNverting Enzyme Investigators, I. M. A. G. I. N. E. (I. M. (2008), 'Effects of angiotensin-converting enzyme inhibition in low-risk patients early after coronary artery bypass surgery.', *Circulation* 117(1), 24–31.

Roy, D.; Talajic, M.; Nattel, S.; Wyse, D. G.; Dorian, P.; Lee, K. L.; Bourassa, M. G.; Arnold, J. M. O.; Buxton, A. E.; Camm, A. J.; Connolly, S. J.; Dubuc, M.; Ducharme, A.; Guerra, P. G.; Hohnloser, S. H.; Lambert, J.; Heuzey, J.-Y. L.; O'Hara, G.; Pedersen, O. D.; Rouleau, J.-L.; Singh, B. N.; Stevenson, L. W.; Stevenson, W. G.; Thibault, B.; Waldo, A. L.; Fibrillation, A. & Investigators, C. H. F. (2008), 'Rhythm control versus rate control for atrial fibrillation and heart failure.', *N Engl J Med* 358(25), 2667–2677.

Sacco, R. L.; Diener, H.-C.; Yusuf, S.; Cotton, D.; Ounpuu, S.; Lawton, W. A.; Palesch, Y.; Martin, R. H.; Albers, G. W.; Bath, P.; Bornstein, N.; Chan, B. P. L.; Chen, S.-T.; Cunha, L.; Dahlöf, B.; Keyser, J. D.; Donnan, G. A.; Estol, C.; Gorelick, P.; Gu, V.; Hermansson, K.; Hilbrich, L.; Kaste, M.; Lu, C.; Machnig, T.; Pais, P.; Roberts, R.; Skvortsova, V.; Teal, P.; Toni, D.; Vandermaelen, C.; Voigt, T.; Weber, M.; Yoon, B.-W. & Group, P. R. E. S. S. S. (2008), 'Aspirin and extended-release dipyridamole versus clopidogrel for recurrent stroke.', *N Engl J Med* 359(12), 1238–1251.

Sesso, H. D.; Buring, J. E.; Christen, W. G.; Kurth, T.; Belanger, C.; MacFadyen, J.; Bubes, V.; Manson, J. E.; Glynn, R. J. & Gaziano, J. M. (2008), 'Vitamins E and C in the prevention of cardiovascular disease in men: the Physicians' Health Study II randomized controlled trial.', *JAMA* 300(18), 2123–2133.

Sever, P. S.; Poulter, N. R.; Dahlof, B.; Wedel, H.; Beevers, G.; Caulfield, M.; Collins, R.; Kjeldsen, S. E.; Kristinsson, A.; McInnes, G.; Mehlsen, J.; Nieminen, M. S.; O'Brien, E. T.; Ostergren, J. & Investigators, A. S. C. O. T. (2008), 'The Anglo-Scandinavian Cardiac Outcomes Trial lipid lowering arm: extended observations 2 years after trial closure.', *Eur Heart J* 29(4), 499–508.

Stone, G. W.; Midei, M.; Newman, W.; Sanz, M.; Hermiller, J. B.; Williams, J.; Farhat, N.; Mahaffey, K. W.; Cutlip, D. E.; Fitzgerald, P. J.; Sood, P.; Su, X.; Lansky, A. J. & Investigators, S. P. I. R. I. T. I. (2008), 'Comparison of an everolimus-eluting stent and a paclitaxel-eluting stent in patients with coronary artery disease: a randomized trial.', *JAMA* 299(16), 1903–1913.

Stone, G. W.; Witzenbichler, B.; Guagliumi, G.; Peruga, J. Z.; Brodie, B. R.; Dudek, D.; Kornowski, R.; Hartmann, F.; Gersh, B. J.; Pocock, S. J.; Dangas, G.; Wong, S. C.; Kirtane, A. J.; Parise, H.; Mehran, R. & Investigators, H. O. R. I. Z. O. N. S.-A. M. I. T. (2008), 'Bivalirudin during primary PCI in acute myocardial infarction.', *N Engl J Med* 358(21), 2218–2230.

Tardif, J.-C.; McMurray, J. J. V.; Klug, E.; Small, R.; Schumi, J.; Choi, J.; Cooper, J.; Scott, R.; Lewis, E. F.; L'Allier, P. L.; Pfeffer, M. A. & of Inflammation Stops Events (ARISE) Trial Investigators, A. R. (2008), 'Effects of succinobucol (AGI-1067) after an acute coronary syndrome: a randomised, double-blind, placebo-controlled trial.', *Lancet* 371(9626), 1761–1768.

Thiele, H.; Schindler, K.; Friedenberger, J.; Eitel, I.; Fürnau, G.; Grebe, E.; Erbs, S.; Linke, A.; Möbius-Winkler, S.; Kivelitz, D. & Schuler, G. (2008), 'Intracoronary compared with intravenous bolus abciximab application in patients with ST-elevation myocardial infarction undergoing primary percutaneous coronary intervention: the randomized Leipzig immediate percutaneous coronary intervention abciximab IV versus IC in ST-elevation myocardial infarction trial.', *Circulation* 118(1), 49–57.

Torre-Amione, G.; Anker, S. D.; Bourge, R. C.; Colucci, W. S.; Greenberg, B. H.; Hildebrandt, P.; Keren, A.; Motro, M.; Moyé, L. A.; Otterstad, J. E.; Pratt, C. M.; Ponikowski, P.; Rouleau, J. L.; Sestier, E.; Winkelmann, B. R.; Young, J. B. & of Immune Modulation Therapy Investigators, A. C. H. F. C. A. (2008), 'Results of a non-specific immunomodulation therapy in chronic heart failure (ACCLAIM trial): a placebo-controlled randomised trial.', *Lancet* 371(9608), 228–236.

TRANSCEND; Yusuf, S.; Teo, K.; Anderson, C.; Pogue, J.; Dyal, L.; Copland, I.; Schumacher, H.; Dagenais, G. & Sleight, P. (2008), 'Effects of the angiotensin-receptor blocker telmisartan on cardiovascular events in high-risk patients intolerant to angiotensin-converting enzyme inhibitors: a randomised controlled trial.', *Lancet* 372(9644), 1174–1183.

Valgimigli, M.; Campo, G.; Percoco, G.; Bolognese, L.; Vassanelli, C.; Colangelo, S.; de Cesare, N.; Rodriguez, A. E.; Ferrario, M.; Moreno, R.; Piva, T.; Sheiban, I.; Pasquetto, G.; Prati, F.; Nazzaro, M. S.; Parrinello, G.; Ferrari, R. & of Single High-Dose Bolus Tirofiban vs Abciximab With Sirolimus-Eluting Stent or Bare Metal Stent in Acute Myocardial Infarction Study (MULTISTRATEGY) Investigators, M. E. (2008), 'Comparison of angioplasty with infusion of tirofiban or abciximab and with implantation of sirolimus-eluting or uncoated stents for acute myocardial infarction: the MULTISTRATEGY randomized trial.', *JAMA* 299(15), 1788–1799.

Vlaar, P. J.; Svilaas, T.; van der Horst, I. C.; Diercks, G. F. H.; Fokkema, M. L.; de Smet, B. J. G. L.; van den Heuvel, A. F. M.; Anthonio, R. L.; Jessurun, G. A.; Tan, E.-S.; Suurmeijer, A. J. H. & Zijlstra, F. (2008), 'Cardiac death and reinfarction after 1 year in the Thrombus Aspiration during Percutaneous coronary intervention in Acute myocardial infarction Study (TAPAS): a 1-year follow-up study.', *Lancet* 371(9628), 1915–1920.

Werk, M.; Langner, S.; Reinkensmeier, B.; Boettcher, H.-E.; Tepe, G.; Dietz, U.; Hosten, N.; Hamm, B.; Speck, U. & Ricke, J. (2008), 'Inhibition of restenosis in femoropopliteal arteries: paclitaxel-coated versus uncoated balloon: femoral paclitaxel randomized pilot trial.', *Circulation* 118(13), 1358–1365.

Windecker, S.; Serruys, P. W.; Wandel, S.; Buszman, P.; Trznadel, S.; Linke, A.; Lenk, K.; Ischinger, T.; Klauss, V.; Eberli, F.; Corti, R.; Wijns, W.; Morice, M.-C.; di Mario, C.; Davies, S.; van Geuns, R.-J.; Eerdmans, P.; van Es, G.-A.; Meier, B. & Jüni, P. (2008), 'Biolimus-eluting stent with biodegradable polymer versus sirolimus-eluting stent with durable polymer for coronary revascularisation (LEADERS): a randomised non-inferiority trial.', *Lancet* 372(9644), 1163–1173.

Yusuf, S.; Diener, H.-C.; Sacco, R. L.; Cotton, D.; Ounpuu, S.; Lawton, W. A.; Palesch, Y.; Martin, R. H.; Albers, G. W.; Bath, P.; Bornstein, N.; Chan, B. P. L.; Chen, S.-T.; Cunha, L.; Dahlöf, B.; Keyser, J. D.; Donnan, G. A.; Estol, C.; Gorelick, P.; Gu, V.; Hermansson, K.; Hilbrich, L.; Kaste, M.; Lu, C.; Machnig, T.; Pais, P.; Roberts, R.; Skvortsova, V.; Teal, P.; Toni, D.; VanderMaelen, C.; Voigt, T.; Weber, M.; Yoon, B.-W. & Group, P. R. E. S. S. S. (2008), 'Telmisartan to prevent recurrent stroke and cardiovascular events.', *N Engl J Med* 359(12), 1225–1237.

## APPENDIX FOR CHAPTER 5

### Weibull distribution laws of $T_1$ and $T_2$ for Cases 2 and 3.

If we assume that the time  $T_k(k = 1, 2)$  to  $\mathcal{E}_k$  follows a Weibull distribution, with shape and scale parameters  $\beta_k$  and  $b_k$ , respectively, the density and survival functions are given by:

$$\begin{aligned} f_k(t) &= \frac{\beta_k}{(b_k)^{\beta_k}} t^{\beta_k-1} e^{-(t/b_k)^{\beta_k}} \\ S_k(t) &= e^{-(t/b_k)^{\beta_k}}. \end{aligned}$$

Fixing the shape parameters  $\beta_1$  and  $\beta_2$ , the scale parameters are derived specifically for each one of the two cases:

**Case 2:** The scale parameter  $b_1$  is a function of the joint density  $f_{(1,2)}(t_1, t_2; \theta)$  (see 5.6) and it is found as the solution of equation  $p_1 = \int_0^t \int_u^\infty f_{(1,2)}(u, v; \theta) dv du$ , where  $p_1$  is the probability of observing  $\mathcal{E}_1$  in control group. The scale parameter  $b_2$  is a function of  $p_2$  and  $\beta_2$  and given by  $b_2 = \frac{1}{(-\log(1-p_2))^{1/\beta_2}}$ , where  $p_2$  is the probability of observing  $\mathcal{E}_2$  in control group.

**Case 3:** The scale parameter  $b_1$  is a function of  $p_1$  and  $\beta_1$  and given by  $b_1 = \frac{1}{(-\log(1-p_1))^{1/\beta_1}}$ . The scale parameter  $b_2$  is a function of the joint density  $f_{(1,2)}(t_1, t_2; \theta)$  and it is found as the solution of equation  $p_2 = \int_0^t \int_v^\infty f_{(1,2)}(u, v; \theta) du dv$ .

### Copula assumptions

We consider a Frank's Archimedean survival copula relationship between the marginal times  $T_1^{(j)}$  and  $T_2^{(j)}$  ( $j = 0, 1$ ). The copula is given by:

$$C(t_1, t_2; \theta) = -\frac{1}{\theta} \log \left\{ 1 + \frac{(e^{-\theta t_1} - 1)(e^{-\theta t_2} - 1)}{e^{-\theta} - 1} \right\}, \quad (\text{B.1})$$

where the association parameter  $\theta$  is biunivocally related to Spearman's rank correlation  $\rho$  given by  $\rho = \rho(\theta) = 1 - \frac{12}{\theta} [\frac{1}{\theta} \int_0^\theta \frac{t}{e^t - 1} dt - \frac{2}{\theta^2} \int_0^\theta \frac{t^2}{e^t - 1} dt]$ .

Assuming equal association parameter  $\theta = \theta^{(0)} = \theta^{(1)}$  for both groups 0 and 1, the joint survival and joint density for  $(T_1^{(j)}, T_2^{(j)})$  are given by:

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

$$f_{(1,2)}^{(j)}(t_1, t_2; \theta) = \frac{\theta e^{-\theta(S_1^{(j)}(t_1) + S_2^{(j)}(t_2))}}{e^{-2\theta S_{(1,2)}^{(j)}(t_1, t_2; \theta)} (e^{-\theta} - 1)} [f_1^{(j)}(t_1)] [f_2^{(j)}(t_2)], \quad (\text{B.3})$$

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 marginal survival and marginal densities of  $T_1^{(j)}$  and  $T_2^{(j)}$ , respectively.

The survival functions of  $T_*^{(j)} = \min\{T_1^{(j)}, T_2^{(j)}\}$  are given by:

$$\begin{aligned} S_*^{(j)}(t; \theta) &= P(T_*^{(j)} > t) = P(T_1^{(j)} > t, T_2^{(j)} > t) = S_{(1,2)}^{(j)}(t, t; \theta) = \\ &= C(S_1^{(j)}(t), S_2^{(j)}(t)) = -\frac{1}{\theta} \log \left\{ 1 + \frac{(e^{-\theta S_1^{(j)}(t)} - 1)(e^{-\theta S_2^{(j)}(t)} - 1)}{e^{-\theta} - 1} \right\}. \end{aligned} \quad (\text{B.4})$$

The density function of  $T_*^{(j)}$  can be calculated as  $f_*^{(j)}(t; \theta) = -\partial S_*^{(j)}(t; \theta) / \partial t$ :

$$f_*^{(j)}(t; \theta) = \left\{ \frac{e^{-\theta S_1^{(j)}(t)}(e^{-\theta S_2^{(j)}(t)} - 1)f_1^{(j)}(t) + e^{-\theta S_2^{(j)}(t)}(e^{-\theta S_1^{(j)}(t)} - 1)f_2^{(j)}(t)}{(e^{-\theta} - 1) + (e^{-\theta S_1^{(j)}(t)} - 1)(e^{-\theta S_2^{(j)}(t)} - 1)} \right\}. \quad (\text{B.5})$$

### Marginal laws of $T_k^{(j)}$

We assume that the time  $T_k^{(j)}$  to observe each  $k$  component for each group  $j$  follows a Weibull distribution, with scale parameter  $b_k^{(j)}$  and shape parameter  $\beta_k^{(0)} = \beta_k^{(1)} = \beta_k$  (so that constant hazard ratios' assumption holds). The density and survival functions are given by:

$$f_k^{(j)}(t) = \frac{\beta_k}{(b_k^{(j)})^{\beta_k}} t^{\beta_k-1} e^{-(t/b_k^{(j)})^{\beta_k}} \quad (\text{B.6})$$

$$S_k^{(j)}(t) = e^{-(t/b_k^{(j)})^{\beta_k}}. \quad (\text{B.7})$$

### Scale parameters for group 0

The scale parameters  $b_1^{(0)}$ ,  $b_2^{(0)}$ , for group 0, have to be derived specifically for each one of the four Cases taking into account whether endpoints  $\mathcal{E}_1$  and  $\mathcal{E}_2$  are terminating as it is shown in the equations below:

**Case 1:** For  $k = 1, 2$ , the scale parameter  $b_k^{(0)}$  is a function of  $p_k$  and  $\beta_k$  and given by  $b_k^{(0)} = \frac{1}{(-\log(1-p_k))^{1/\beta_k}}$ .

**Case 2:** The scale parameter  $b_1^{(0)}$  is a function of the joint density  $f_{(1,2)}^{(0)}(t_1, t_2; \theta)$  and it is found as the solution of equation  $p_1 = \int_0^{\tau} \int_u^{\infty} f_{(1,2)}^{(0)}(u, v; \theta) dv du$ . The scale parameter  $b_2^{(0)}$  is a function of  $p_2$  and  $\beta_2$  and given by  $b_2^{(0)} = \frac{1}{(-\log(1-p_2))^{1/\beta_2}}$ .

**Case 3:** The scale parameter  $b_1^{(0)}$  is a function of  $p_1$  and  $\beta_1$  and given by  $b_1^{(0)} = \frac{1}{(-\log(1-p_1))^{1/\beta_1}}$ . The scale parameter  $b_2^{(0)}$  is a function of the joint density  $f_{(1,2)}^{(0)}(t_1, t_2; \theta)$  and it is found as the solution of equation  $p_2 = \int_0^{\tau} \int_v^{\infty} f_{(1,2)}^{(0)}(u, v; \theta) dudv$ .

**Case 4:** The scale parameters  $b_k^{(0)}$  ( $k = 1, 2$ ) are functions of the joint density  $f_{(1,2)}^{(0)}(t_1, t_2; \theta)$  and are found as the simultaneous solution of equations  $p_1 = \int_0^{\tau} \int_u^{\infty} f_{(1,2)}^{(0)}(u, v; \theta) dv du$  and  $p_2 = \int_0^{\tau} \int_v^{\infty} f_{(1,2)}^{(0)}(u, v; \theta) dudv$ .

### Scale parameters for group 1

The scale parameters for group 1,  $b_1^{(1)}$  and  $b_2^{(1)}$ , are derived in such a way that the marginal hazard ratio  $HR_k$  ( $k = 1, 2$ ) is constant and given by

$$HR_k = \frac{\lambda_k^{(1)}(t)}{\lambda_k^{(0)}(t)} = \left( \frac{b_k^{(0)}}{b_k^{(1)}} \right)^{\beta_k}, \quad (\text{B.8})$$

where  $\lambda_k^{(0)}(t)$  and  $\lambda_k^{(1)}(t)$  are the hazard functions of  $T_k^{(0)}$  and  $T_k^{(1)}$ , respectively.

Hence, the scale parameters for group 1,  $b_k^{(1)}$ , can be expressed as a function of  $b_k^{(0)}$  and  $HR_k$ , that is,  $b_k^{(1)} = \frac{b_k^{(0)}}{HR_k^{1/\beta_k}}$ .







## APPENDIX: R CODE TO COMPUTE THE ASYMPTOTIC RELATIVE EFFICIENCY (ARE) VALUES

```
#####
#
# ARE_case1234.R
#####
# Computation of the Asymptotic relative Efficiency (ARE) values for censoring
# cases 1,2,3 and 4, for several copulas and Weibull distributions.
#
#####
#
# CASE 1: The composite endpoint does not include a fatal event (i.e. Death)
# neither in the Relevant endpoint nor in the Additional endpoint.
#
# CASE 2: The composite endpoint does not include a fatal event
# in the Relevant endpoint but it does in the Additional endpoint.
#
# CASE 3: The composite endpoint does include a fatal event
# in the Relevant endpoint but it does not in the Additional endpoint.
#
# CASE 4: The composite endpoint does include a fatal event
# both in the Relevant endpoint and in the Additional endpoint.
#
# Last update: 8/03/2016
#
# R version: R 3.2.3
#
# Authors: Moisés Gómez Mateu (moises.gomez.mateu@upc.edu)
# Oleguer Plana Ripoll (oleguerplana@gmail.com)
#
#####
```

```

#
# References:
# - [1] Gómez G. and Lagakos S.W. (2013). Statistical considerations when using a composite
# endpoint for comparing treatment groups. Statistics in Medicine, 32, 719–38.
# - [2] Gómez G. and Lagakos S. (2013). Web-based Supporting Materials for "Statistical
# Considerations when Using a Composite Endpoint for Comparing Treatment Groups"
# by G. Gómez and S.W. Lagakos.
# - [3] Gómez G. and Gómez-Mateu M (2014). The Asymptotic Relative Efficiency and the ratio
# of sample sizes when testing two different null hypotheses. SORT. 38, 73–88.
#
#####

if ("copula" %in% rownames(installed.packages()) == FALSE) { install.packages("copula") }
library(copula)

if ("numDeriv" %in% rownames(installed.packages()) == FALSE) { install.packages("numDeriv") }
library(numDeriv)

if ("rootSolve" %in% rownames(installed.packages()) == FALSE) { install.packages("rootSolve") }
library('rootSolve')

#####
# Function: ARE
#
#####
# Description: It computes the ARE value for the given arguments
#
# rho0 Spearman's coefficient between T1 and T2 in control group
# rho1 Spearman's coefficient between T1 and T2 in treatment group
# 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 Relevant events in control group
# p2 Proportion of Additional events in control group
# case Censoring case -- > 1 (default), 2, 3 or 4
# copula Copula used:
# Archimedean: "Frank" (default), "Gumbel" or "Clayton"
# Elliptical: "Normal" or "T"
# Extreme Value: "Galambos", "HuslerReiss", "Gumbel", "Tawn" or "Tev"
# Others: "FGM" or "Plackett"
#####

ARE<-function(rho0,rho1=rho0, beta1, beta2, HR1, HR2, p1, p2, case = 1, copula="Frank")
{
#####
##### 0. WARNINGS AND ERRORS
if (rho0>1 | rho0<(-1)) {
stop("correlation_rho_must_be_a_number_between_-1_and_1", call.=FALSE)
}

if (p1<0 | p2<0 | p1>1 | p2>1) {
stop("probabilities_p1_and_p2_must_be_between_0_and_1.", call.=FALSE)
}

case_check<-0
if (case==1|case==2|case==3|case==4) {case_check=1}
if (case_check==0){stop("Please ,_introduce_a_valid_Case_value:_1,2,3_or_4.", call.=FALSE)}

# Note: Warning for copula validation already implemented. See CopulaSelection function.
#####

```

```

#####
##### 1. SELECTION OF THE COPULA
copula0<-CopulaSelection (copula , rho0)
theta<-copula0[[2]]
which.copula0<-copula0[[1]]
which.copula1<-CopulaSelection (copula , rho1)[[1]]
#####

#####
##### 2. SELECTION OF THE MARGINAL DISTRIBUTIONS
MarginSelec<-MarginalsSelection (beta1 , beta2 , HR1,HR2,p1 , p2 , case , theta)
T1dist<-MarginSelec[[1]]
T2dist<-MarginSelec[[2]]
T1pdist<-MarginSelec[[3]]
T2pdist<-MarginSelec[[4]]
T10param<-MarginSelec[[5]]
T20param<-MarginSelec[[6]]
T11param<-MarginSelec[[7]]
T21param<-MarginSelec[[8]]
#####

#####
##### 3. ARE EXPRESSION FOLLOWING GÓMEZ AND LAGAKOS (See reference [2], pages 2 and 3).

# Bivariate distribution in control and treatment groups
distribution0 <- mvdc(copula = which.copula0 , margins = c(T1dist , T2dist) , paramMargins = list(T10param , T20param))
distribution1 <- mvdc(copula = which.copula1 , margins = c(T1dist , T2dist) , paramMargins = list(T11param , T21param))

if (case==1|case==3) {

# Inside the integral in the numerator (See reference [2], page 2).
inside_integral <- function (t) {
Sstar0<-Sstar (x=t , dist1=T1pdist , dist2=T2pdist , param1=T10param , param2=T20param , dist_biv= distribution0)
Sstar1<-Sstar (x=t , dist1=T1pdist , dist2=T2pdist , param1=T11param , param2=T21param , dist_biv= distribution1)
fstar0<-(-grad (Sstar , x=t , dist1=T1pdist , dist2=T2pdist , param1=T10param , param2=T20param , dist_biv= distribution0))
fstar1<-(-grad (Sstar , x=t , dist1=T1pdist , dist2=T2pdist , param1=T11param , param2=T21param , dist_biv= distribution1))
Lstar0 <- (fstar0 / Sstar0)
Lstar1 <- (fstar1 / Sstar1)
HRstar <- (Lstar1 / Lstar0)
logHRstar <- log (HRstar)
return (logHRstar*fstar0)
}

# Integral in the numerator
integral<-integrate (inside_integral , lower=0 , upper=1 , subdivisions=1000 , stop.on.error = FALSE)
numerator<-(integral$value)^2

# Denominator
Sstar0_1<-Sstar (x=1 , dist1=T1pdist , dist2=T2pdist , param1=T10param , param2=T20param , dist_biv= distribution0)
ST10_1 <- 1-do.call (T1pdist , c(q=1 , T10param))
denominator <- ((log (HR1))^2)*(1-Sstar0_1)*(1-ST10_1)

# ARE value
AREstarT <- (numerator/denominator)

# If the integral is not computed, we assign a missing value
if (integral$message!="OK") {AREstarT <- NA}

} else
if (case==2|case==4) {

# Computation of the scale parameter values b10, b20

if (case==2) {
# Compute b20
b20 <- 1/(-log(1-p2))^(1/beta2)

```

```

# Compute b10
Fb10<-function(b10,p1){
  integral<-integrate(function(u) {
    sapply(u, function(u) {
      integrate(function(v) ( (theta*(1-exp(-theta))*exp(-theta*(u+v)))/(exp(-theta)+ exp(-theta*(u+v)))
      - exp(-theta*u)-exp(-theta*v))^2 ) , lower=0, upper= exp((b10*(-log(u))^(1/beta1))^beta2*log(1-p2)) )$value
    })
  }, lower= exp(-1/b10^beta1), upper=1)$value
  return(integral-p1)
}

limits <- c(0.00001,10000) # The first and the last values must be in opposite signs for the function
b10 <- uniroot(Fb10, interval=limits,p1=p1)$root # Find the root (value which equals the function to zero)

}

if(case==4) {
# We need to create x[1] and x[2] to run 'multiroot' function (library: rootSolve) (NA's initially assigned)
x<-NA
y<-NA
x[1]<-x
x[2]<-y

# We need to change the name of variables as (b10=x[1],b20=[2]) to execute 'multiroot'
# Compute b10
Fb10<-function(b10,b20,p1){
  b10->x[1]
  b20->x[2]
  integral<-integrate(function(u) {
    sapply(u, function(u) {
      integrate(function(v) ( (theta*(1-exp(-theta))*exp(-theta*(u+v))
      / (exp(-theta)+ exp(-theta*(u+v)) - exp(-theta*u)-exp(-theta*v))^2 ) , lower=0,
      upper= exp((x[1]*(-log(u))^(1/beta1))^beta2*( -1/(x[2]^beta2) )) )$value
    })
  }, lower= exp(-1/x[1]^beta1), upper=1)$value
  return(integral-p1)
}

# Compute b20
Fb20<-function(b10,b20,p2) {
  b10->x[1]
  b20->x[2]
  integral<-integrate(function(v) {
    sapply(v, function(v) {
      integrate(function(u) ((theta*(1-exp(-theta))*exp(-theta*(u+v)))
      / (exp(-theta)+exp(-theta*(u+v))-exp(-theta*u)-exp(-theta*v))^2 ), lower=0,
      upper=exp(-((( -log(v))^beta2)*x[2])/x[1]^beta1))$value
    })
  },
  lower= exp(-1/x[2]^beta2), upper=1)$value
  return(integral-p2)
}

model <- function(x){
  c(Fb10(x[1],x[2],p1), Fb20(x[1],x[2],p2))
}

(sol <- multiroot(f = model, start = c(1, 1)))

sol<-as.data.frame(sol[1])
b10<-sol[1,]
b20<-sol[2,]
}

```

```

## Computation of the numerator
# Note: Only marginal Weibull distributions for fT10, fT20, ST10, ST20.

fT10 <- function(t){
(beta1/b10) * ( (t/b10)^(beta1-1) ) * (exp(-(t/b10)^beta1))
}

ST10 <- function(t){
exp(-(t/b10)^beta1)
}

fT20 <- function(t){
(beta2/b20) * ( (t/b20)^(beta2-1) ) * (exp(-(t/b20)^beta2))
}

ST20 <- function(t){
exp(-(t/b20)^beta2)
}

# Sstar0 and fstar0 for any copula
Sstar0 <- function(t){
Sstar0<-Sstar(x=t, dist1=T1pdist, dist2=T2pdist, param1=T10param, param2=T20param, dist_biv= distribution0)
}

fstar0<- function(t){
fstar0<-(-grad(Sstar, x=t, dist1=T1pdist, dist2=T2pdist, param1=T10param, param2=T20param, dist_biv= distribution0))
}

aux21 <- function(t, y){
theta*exp(-theta*(ST10(t)+y))*(1-exp(-theta))/(exp(-theta)-exp(-theta*ST10(t))-exp(-theta*y)+exp(-theta*(ST10(t)+y)))^2
}

aux22<-function(u){
integrate(aux21, 0, ST20(u), t=u, subdivisions=10000)$value
}

lambdaC10<-function(t){
aux22(t)*fT10(t)/Sstar0(t)
}

lambdaC11<-function(t){
HR1*lambdaC10(t)
}

aux23<-function(x, t){
theta*exp(-theta*(x+ST20(t)))*(1-exp(-theta))/(exp(-theta)-exp(-theta*x)-exp(-theta*ST20(t))+exp(-theta*(x+ST20(t))))^2
}

aux24<-function(u){
integrate(aux23, 0, ST10(u), t=u, subdivisions=10000)$value
}

lambdaC20<-function(t){
aux24(t)*fT20(t)/Sstar0(t)
}

lambdaC21<-function(t){
HR2*lambdaC20(t)
}

# EVALUATION OF LambdaC20 BEFORE COMPUTATION (IT MAY FAIL IN CASES 214 FOR BETAS = 0.5 BECAUSE IT IS NOT ALWAYS
# EVALUABLE AT T=0)
LambdaC20_check<-tryCatch(
LambdaC20<-function(t){
integrate(lambdaC20, lower=0, upper=t, subdivisions=10000)$value
, error = function(e) e)

```

```

# WHENEVER LambdaC20 FAILS, WE INCREASE THE LOWER LIMIT OF INTEGRATION
lower_LambdaC20<-0
while (inherits (LambdaC20_check, "error")== "TRUE" ) {
lower_LambdaC20=lower_LambdaC20+0.001
LambdaC20_check<-tryCatch (
LambdaC20<-function (t){
integrate (lambdaC20, lower=0+lower_LambdaC20, upper=t, subdivisions=10000)$value
, error = function (e) e)
}

LambdaC20<-function (t){
integrate (lambdaC20, lower=0+lower_LambdaC20, upper=t, subdivisions=10000)$value
}

# Computation of the hazards for both groups

Lstar0<-function (t){
lambdaC10(t)+lambdaC20(t)
}

Lstar1<-function (t){
lambdaC11(t)+lambdaC21(t)
}

# Computation of HRstar

HRstar <- function (t){
Lstar1(t)/Lstar0(t)
}

logHRstar <- function (t){
log(Lstar1(t)/Lstar0(t))
}

temp3<- function (t){
logHRstar(t)*fstar0(t)
}

# EVALUATION OF temp4 BEFORE COMPUTATION (IT MAY FAIL IN CASES 2/4 FOR BETAS = 0.5 BECAUSE IT IS NOT
# ALWAYS EVALUABLE AT T=0)

temp4_check<-tryCatch (
temp4<-integrate (temp3,0,1, subdivisions=10000)$value
,error = function (e) e)

# WHENEVER temp4 FAILS, WE INCREASE THE LOWER LIMIT OF INTEGRATION

lower_temp4<-0
while (inherits (temp4_check, "error")== "TRUE" ) {
lower_temp4=lower_temp4+0.001
temp4_check<-tryCatch (
temp4<-integrate (temp3, lower_temp4, 1, subdivisions=10000)$value
,error = function (e) e)
}

temp4<-integrate (temp3,0+lower_temp4, 1, subdivisions=10000)$value
numerator<-(temp4)^2

## Computation of PROBT1UNC

PROBT1UNC_temp_num <- function (t){
exp(-HR2*LambdaC20(t))*Sstar0(t)*lambdaC10(t)
}

PROBT1UNC_temp_den <- function (t){

```

```

exp(-LambdaC20(t))*1/2 + exp(-HR2*LambdaC20(t))*1/2
}

PROBTIUNC_temp <- function(t){
PROBTIUNC_temp_num(t) / PROBTIUNC_temp_den(t)
}

PROBTIUNC_int_check<- tryCatch (integrate (PROBTIUNC_temp,lower=0, upper=1,subdivisions=10000)$value, error = function(e) e)

#####
#####

# WE EVALUATE THE FUNCTION PROBTIUNC_int BECAUSE IT MAY FAIL IN CASES 2/4 FOR BETAS = 0.5 PROBABLY DUE
# TO THE LOWER LIMITS OF THE INTERGATES lambdaC20 AND temp4.
# WHEN IT FAILS, WE SEARCH FOR THE MINIMUM EVALUABLE LIMIT OF INTEGARTION FOR lambdaC20 AND temp4;
# AND WE SET A LOWER LIMITS OF 0.001 FOR THE REST OF INTEGRATES TO ENSURE CONVERGENCE.

lower_PROBTIUNC_int<-0
inc_lower<-0

while (inherits (PROBTIUNC_int_check, "error")== "TRUE" ) {
lower_PROBTIUNC_int<-0.001
inc_lower<-inc_lower+0.001
aux22<-function(u){
integrate (aux21,0.001, ST20(u), t=u, subdivisions=10000)$value
}
lambdaC10<-function(t){
aux22(t)*fT10(t) / Sstar0(t)
}
lambdaC11<-function(t){
HR1*lambdaC10(t)
}
aux23<-function(x,t){
theta*exp(-theta*(x+ST20(t)))*(1-exp(-theta))/(exp(-theta)-exp(-theta*x)-exp(-theta*ST20(t))+exp(-theta*(x+ST20(t))))^2
}
aux24<-function(u){
integrate (aux23,0.001,ST10(u), t=u, subdivisions=10000)$value
}
lambdaC20<-function(t){
aux24(t)*fT20(t) / Sstar0(t)
}
lambdaC21<-function(t){
HR2*lambdaC20(t)
}
LambdaC20<-function(t){
integrate (lambdaC20,lower=lower_LambdaC20 + inc_lower, upper=t, subdivisions=10000)$value
}
Lstar0<-function(t){
lambdaC10(t)+lambdaC20(t)
}
Lstar1<-function(t){
lambdaC11(t)+lambdaC21(t)
}
HRstar <- function(t){
Lstar1(t) / Lstar0(t)
}
logHRstar <- function(t){
log(Lstar1(t)/Lstar0(t))
}
temp3<- function(t){
logHRstar(t)*fstar0(t)
}

temp4<-integrate (temp3,lower_temp4 + inc_lower,1, subdivisions=10000)$value
numerator<-(temp4)^2

```

```

PROBTIUNC_temp_num <- function(t){
exp(-HR2*LambdaC20(t))*Sstar0(t)*lambdaC10(t)
}
PROBTIUNC_temp_den <- function(t){
exp(-LambdaC20(t))*1/2 + exp(-HR2*LambdaC20(t))*1/2
}
PROBTIUNC_temp <- function(t){
PROBTIUNC_temp_num(t) / PROBTIUNC_temp_den(t)
}
PROBTIUNC_int_check<- tryCatch(integrate(PROBTIUNC_temp,lower=0.001, upper=1,subdivisions=10000)$value, error = function(e) e)
}

PROBTIUNC_int<-integrate(PROBTIUNC_temp,lower=lower_PROBTIUNC_int, upper=1,subdivisions=10000)$value

#####
#####

AREstarT <- numerator / ((log(HR1)^2) * PROBTIUNC_int * (1-Sstar0(1)))
AREstarT

}

return(AREstarT)
}

#####
# Function: CopulaSelection
#
#####
# Description: Constructs a copula class object from the family given and the
#               the corresponding dependence parameter from the given correlation
#
# copula Copula given:
#           Archimedean: "Frank" (default), "Gumbel" or "Clayton"
#           Elliptical: "Normal" or "T"
#           Extreme Value: "Galambos", "HuslerReiss", "Gumbel", "Tawn" or "Tev"
#           Other: "FGM" or "Plackett"
# rho Spearman's coefficient between the 2 marginal distributions
#####

CopulaSelection <- function(copula,rho) {
if (copula=="Frank") {
theta<-iRho(frankCopula(1),rho)
which.copula <- archmCopula(family = "frank", dim = 2, param = theta)
return(c(which.copula,theta))
} else
if (copula=="Gumbel") {
theta<-iRho(gumbelCopula(2),rho)
which.copula <- archmCopula(family = "gumbel", dim = 2, param = theta)
return(c(which.copula,theta))
} else
if (copula=="Clayton") {
theta<-iRho(claytonCopula(1),rho)
which.copula <- archmCopula(family = "clayton", dim = 2, param = theta)
return(c(which.copula,theta))
} else
if (copula=="FGM") {
theta<-iRho(fgmCopula(1),rho)
which.copula <- fgmCopula(dim = 2, param = theta)
return(c(which.copula,theta))
} else
if (copula=="Normal") {
theta<-iRho(normalCopula(0.5),rho)
which.copula <- normalCopula(dim = 2, param = theta)
return(c(which.copula,theta))
} else

```



```

if (copula=="T") {
theta<-iRho(tCopula(0.5),rho)
which.copula <- tCopula(dim = 2, param = theta)
return (c(which.copula, theta))
} else
if (copula=="Galambos") {
theta<-iRho(galambosCopula(0.5),rho)
which.copula <- galambosCopula(param = theta)
return (c(which.copula, theta))
} else
if (copula=="HuslerReiss") {
theta<-iRho(huslerReissCopula(0.5),rho)
which.copula <- huslerReissCopula(param = theta)
return (c(which.copula, theta))
} else
if (copula=="Tawn") {
theta<-iRho(tawnCopula(0.5),rho)
which.copula <- tawnCopula(param = theta)
return (c(which.copula, theta))
} else
if (copula=="Tev") {
theta<-iRho(tevCopula(0.5),rho)
which.copula <- tevCopula(param = theta)
return (c(which.copula, theta))
} else
if (copula=="Plackett") {
theta<-iRho(plackettCopula(0.5),rho)
which.copula <- plackettCopula(param = theta )
return (c(which.copula, theta))
} else { stop(paste("Not_implemented_for",copula,"copula.")) }
return (c(which.copula, theta))
}

#####
# Function: MarginalsSelection
#
#####
# Description: Returns the family distribution and parameters of the marginals
# (ONLY WEIBULL DISTRIBUTIONS SO FAR)
#
# 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), 2, 3 or 4
# theta Dependence parameter for the bivariate distribution in control group
#####

MarginalsSelection<-function(beta1,beta2,HR1,HR2,p1,p2,case,theta)
{
# Scale parameters for group 0 b10,b20

if(case==1) {
b10 <- 1/((-log(1-p1))^(1/beta1))
b20 <- 1/((-log(1-p2))^(1/beta2))
} else
if (case==2) {
Fb10<-function(b10,p1){
integral<-integrate(function(u) {
sapply(u, function(u) {
integrate(function(v) ( (theta*(1-exp(-theta))*exp(-theta*(u+v)))/ (exp(-theta)+ exp(-theta*(u+v)) -
exp(-theta*u)-exp(-theta*v))^2 ) , lower=0, upper= exp((b10*(-log(u))^(1/beta1))^beta2*log(1-p2)) )$value
})
})
}, lower= exp(-1/b10^beta1), upper=1)$value
}
}

```

```

return(integral-p1)
}
limits <- c(0.00001,10000) # The first and the last values must be in opposite signs for the function
b10 <- uniroot(Fb10, interval=limits,p1=p1)$root # Find the root (value which equals the function zero)
b20 <- 1/(-log(1-p2))^(1/beta2)
} else
if (case==3) {
b10 <- 1/((-log(1-p1))^(1/(beta1)))

Fb20<-function(b20,p2) {
integral<-integrate(function(v) {
sapply(v,function(u) {
integrate(function(u)((theta*(1-exp(-theta))*exp(-theta*(u+v)))
/(exp(-theta)+exp(-theta*(u+v))-exp(-theta*u)-exp(-theta*v))^2),lower=0,
upper=exp(-(((log(v))^(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
} else
if (case==4) {

# We need to create x[1] and x[2] (we assign NA's)
x<-NA
y<-NA
x[1]<-x
x[2]<-y

# We need to change the name of variables as (b10=x[1],b20=[2]) to execute 'multiroot'
# Compute b10
Fb10<-function(b10,b20,p1){
b10->x[1]
b20->x[2]
integral<-integrate(function(u) {
sapply(u,function(v) {
integrate(function(v) ((theta*(1-exp(-theta))*exp(-theta*(u+v)))
/(exp(-theta)+ exp(-theta*(u+v)) - exp(-theta*u)-exp(-theta*v))^2 ) , lower=0,
upper= exp((x[1]*(-log(u))^(1/beta1))^beta2*( -1/(x[2]^beta2) )) )$value
}))
}, lower= exp(-1/x[1]^beta1), upper=1)$value
return(integral-p1)
}

# Compute b20
Fb20<-function(b10,b20,p2) {
b10->x[1]
b20->x[2]
integral<-integrate(function(v) {
sapply(v,function(u) {
integrate(function(u)((theta*(1-exp(-theta))*exp(-theta*(u+v)))
/(exp(-theta)+exp(-theta*(u+v))-exp(-theta*u)-exp(-theta*v))^2),lower=0,
upper=exp(-(((log(v))^(1/beta2))*x[2])/x[1]^beta1))$value
}))
},
lower= exp(-1/x[2]^beta2), upper=1)$value
return(integral-p2)
}

model <- function(x){
c(Fb10(x[1],x[2],p1), Fb20(x[1],x[2],p2))
}

(sol <- multiroot(f = model, start = c(1, 1)))

```

```

sol<-as.data.frame(sol[1])
b10<-sol[1,]
b20<-sol[2,]

}

# Scale parameters for group 1 b11,b21 (Although we do not need to compute the scale parameters for
# group 1 (b11, b21) to calculate the ARE)
b11 <- b10/HR1^(1/beta1)
b21 <- b20/HR2^(1/beta2)

T1dist<-"weibull"
T2dist<-"weibull"

T1pdist<-pweibull
T2pdist<-pweibull

T10param<-list(shape = beta1, scale = b10)
T20param<-list(shape = beta2, scale = b20)

T11param<-list(shape = beta1, scale = b11)
T21param<-list(shape = beta2, scale = b21)

return(list(T1dist,T2dist,T1pdist,T2pdist,T10param,T20param,T11param,T21param))
}

#####
# Function: Sstar
#
#####
# Description: Returns the value of the survival function of S* at point x given the
#              marginal distributions and the bivariate distributions via copula
#
# x           Point in which to be evaluated
# dist1      Distribution function of the marginal T1 (pweibull)
# dist2      Distribution function of the marginal T2 (pweibull)
# param1     Parameters of the marginal distribution function T1 (pweibull)
# param2     Parameters of the marginal distribution function T2 (pweibull)
# dist_biv   Distribution function of the bivariate distribution via copula
#####

Sstar<-function(x,dist1,dist2,param1,param2,dist_biv) {
y <- if(length(x) == 1) c(x,x) else cbind(x,x)
return(
1
- do.call(dist1,c(list(q=x),param1))
- do.call(dist2,c(list(q=x),param2))
+ (pMvdc(y, dist_biv))
)
}

```



## APPENDIX: PUBLICATIONS RELATED TO THIS THESIS

- Gómez G, Gómez-Mateu M, Dafni U. Informed Choice of Composite End Points in Cardiovascular Trials (2014). *Circulation. Cardiovascular Quality and Outcomes*, 7, 170–178.
- Gómez G, Gómez-Mateu M. The Asymptotic Relative Efficiency and the ratio of sample sizes when testing two different null hypotheses (2014). *SORT*, 38, 73–88.
- Gómez G, Gómez-Mateu M. Comments on "Use of composite endpoints in clinical trials" by Abdul J. Sankoh, Haihong Li and Ralph B. D'Agostino, Sr (2016). *Statistics in Medicine*, 35, 317–318.



**Informed Choice of Composite End Points in Cardiovascular Trials**

Guadalupe Gómez, Moisés Gómez-Mateu and Urania Dafni

*Circ Cardiovasc Qual Outcomes*. published online January 14, 2014;  
*Circulation: Cardiovascular Quality and Outcomes* is published by the American Heart Association, 7272  
Greenville Avenue, Dallas, TX 75231  
Copyright © 2014 American Heart Association, Inc. All rights reserved.  
Print ISSN: 1941-7705. Online ISSN: 1941-7713

**ATTENTION !**

Pages 134 to 144 of the thesis are available at the editor's web

The online version of this article, along with updated information and services, is located on the  
World Wide Web at:

<http://circoutcomes.ahajournals.org/content/early/2014/01/13/CIRCOUTCOMES.113.000149>

**Permissions:** Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Circulation: Cardiovascular Quality and Outcomes* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the [Permissions and Rights Question and Answer](#) document.

**Reprints:** Information about reprints can be found online at:  
<http://www.lww.com/reprints>

**Subscriptions:** Information about subscribing to *Circulation: Cardiovascular Quality and Outcomes* is online at:  
<http://circoutcomes.ahajournals.org/subscriptions/>

# The asymptotic relative efficiency and the ratio of sample sizes when testing two different null hypotheses

Guadalupe Gómez<sup>\*,1</sup> and Moisés Gómez-Mateu<sup>1</sup>

---

## Abstract

Composite endpoints, consisting of the union of two or more outcomes, are often used as the primary endpoint in time-to-event randomized clinical trials. Previously, Gómez and Lagakos provided a method to guide the decision between using a composite endpoint instead of one of its components when testing the effect of a treatment in a randomized clinical trial. Consider the problem of testing the null hypotheses of no treatment effect by means of either the single component or the composite endpoint. In this paper we prove that the usual interpretation of the asymptotic relative efficiency as the reciprocal ratio of the sample sizes required for two test procedures, for the same null and alternative hypothesis, and attaining the same power at the same significance level, can be extended to the test procedures considered here for two different null and alternative hypotheses. A simulation to study the relationship between asymptotic relative efficiency and finite sample sizes is carried out.

---

MSC: 62N03, 62P10

*Keywords:* Asymptotic relative efficiency, composite endpoint, logrank test, sample size, simulation, survival analysis.

## 1. Introduction

In clinical trials research, one of the most important issues that investigators have to solve at the design stage of the study is the appropriate choice of the primary endpoint. Composite endpoints (CE) consisting of the union of two or more outcomes are commonly used as primary endpoints. For example, in the cardiovascular area the

---

\* Corresponding author e-mail: lupe.gomez@upc.edu

<sup>1</sup> Departament d'Estadística i I.O., Universitat Politècnica de Catalunya, Jordi Girona 1-3, 08034. Barcelona, Spain.

Received: May 2013

Accepted: January 2014



relevant endpoint of death is often combined with other additional endpoints such as myocardial infarction, stroke or hospitalization. Pros and cons on the use of CE have been extensively discussed (Freemantle et al., 2003; Ferreira-González et al., 2007, among many others). One of the main advantages of using a CE relies in the fact that by means of a CE the problem of multiplicity is adequately addressed and the bias associated with competing risks (Wittkop et al., 2010) is avoided. Also, with a CE the number of observed events will be higher and, hopefully, the power of the test will increase. However, as it has been discussed (Montori et al., 2005) and shown in Gómez and Lagakos (2013), adding inappropriate components to the relevant endpoint might actually lead to a decrease in the power of the test statistic, consequently having a larger chance to fail in detecting a real effect of the treatment under study.

Gómez and Lagakos (2013) developed a methodology to help to decide when it is worthwhile to base the analysis on the composite endpoint  $\mathcal{E}_* = \mathcal{E}_1 \cup \mathcal{E}_2$  where  $\mathcal{E}_1$  and  $\mathcal{E}_2$  are two candidate relevant endpoints to evaluate the effect of a treatment instead of sticking to one of them,  $\mathcal{E}_1$ , say. In order to do so, they compared how more efficient than  $\mathcal{E}_1$  would  $\mathcal{E}_*$  be to justify its use. Let  $H_0$  be the null hypothesis of no treatment effect evaluated on  $\mathcal{E}_1$  and denote by  $H_a$  an alternative hypothesis, for instance, claiming to delay the event  $\mathcal{E}_1$ . Analogously, define  $H_0^*$  and  $H_a^*$  the null and alternative hypotheses if the treatment effect is to be evaluated on  $\mathcal{E}_*$ . Since when comparing two treatment groups based on time-to-event endpoints, the primary analysis would be based, very commonly, on a logrank test, their method considers the logrank test  $Z$  to test  $H_0$  versus  $H_a$  and the logrank test  $Z_*$  to test  $H_0^*$  versus  $H_a^*$ . The asymptotic relative efficiency (ARE) of  $Z_*$  versus  $Z$  is the measure proposed to choose between  $\mathcal{E}_1$  and  $\mathcal{E}_*$ , with values larger than 1 in favour of  $\mathcal{E}_*$ . This relative measure can be computed as  $(\mu_*/\mu)^2$  where  $\mu$  and  $\mu_*$  are, respectively, the asymptotic means of  $Z$  and  $Z_*$ , under alternative contiguous hypotheses to  $H_0$  and  $H_0^*$ . The purpose of this paper is to prove that the usual interpretation of the ARE, as the ratio of sample sizes,  $n$  and  $n_*$ , needed to attain the same power for a given significance level, still holds even though two different sets of hypothesis ( $H_0$  versus  $H_a$  and  $H_0^*$  versus  $H_a^*$ ) are compared.

To clarify the purpose of our investigation consider the following. If we were to test  $H_0$  versus  $H_a$  with two different test statistics  $S_n$  and  $T_m$ , Pitman's relative efficiency would be defined as the ratio  $m/n$ , where  $n$  and  $m$  are the required sample sizes for  $S_n$  and  $T_m$ , respectively, to attain the same power for a given significance level. Furthermore, if both  $S_n$  and  $T_m$  are asymptotically normal with unit variance and means  $\mu_S$  and  $\mu_T$ , it can be proved that Pitman's ARE corresponds to the square of the ratio of the noncentrality parameters, that is  $(\mu_S/\mu_T)^2$ . Gómez and Lagakos' method compares the logrank statistics:  $Z$  and  $Z_*$  derived for two different set of hypotheses  $H_0$  versus  $H_a$  and  $H_0^*$  versus  $H_a^*$  and do so using, as definition of the ARE, the ratio  $(\mu_*/\mu)^2$  where  $\mu$  and  $\mu_*$  are, respectively, the asymptotic means of  $Z$  and  $Z_*$ , under alternative contiguous hypotheses to  $H_0$  and  $H_0^*$ .

This paper is organized as follows. In Section 2 the notation, assumptions and main results from Gómez and Lagakos' paper are introduced. Section 3 establishes

the limiting relationship between ARE and sample sizes and proves that the usual interpretation of the ARE as the ratio of sample sizes holds. Section 4 presents a simulation to study under which conditions and for finite sample sizes, the relationship  $\text{ARE}(Z_*, Z) = (\mu_*/\mu)^2 = n/n_*$  holds where  $n$  and  $n_*$  are the needed sample sizes for  $Z$  and  $Z_*$ , respectively, to attain the same power for a given significance level. Section 5 concludes the paper with a discussion.

## 2. Notation, the logrank test and the asymptotic relative efficiency

### 2.1. The logrank tests for the relevant and for the composite endpoints

Assume that we have a two-arm study involving random assignment to an active ( $X = 1$ ) or control treatment ( $X = 0$ ) aiming to prove the efficacy of the new active treatment. The effect of treatment is to be evaluated on the time  $T_1^{(j)}$  to a relevant event  $\mathcal{E}_1$ , where the superscript  $j$  indicates the treatment group ( $j = 0$  for the control group and  $j = 1$  for the treatment group). Let  $\lambda_1^{(j)}(t)$  denote the hazard function of  $T_1^{(j)}$  ( $j = 0, 1$ ). The null hypothesis of no effect is given by  $H_0: \text{HR}_1(t) = \lambda_1^{(1)}(t)/\lambda_1^{(0)}(t) = 1$  and the alternative that the new treatment improves survival by  $H_a: \text{HR}_1(t) < 1$ . The logrank test  $Z$  is used to test that the new treatment improves survival.

Assume now that an additional endpoint  $\mathcal{E}_2$  is considered as component of the primary endpoint and the composite endpoint  $\mathcal{E}_* = \mathcal{E}_1 \cup \mathcal{E}_2$  is to be used, instead, to prove the efficacy of the new treatment. The effect of treatment would then be evaluated on the time  $T_*^{(j)}$  to  $\mathcal{E}_*$  where  $T_*^{(j)} = \min\{T_1^{(j)}, T_2^{(j)}\}$  and  $T_2^{(j)}$  stands for the time to  $\mathcal{E}_2$  ( $j = 0, 1$ ). Let  $\lambda_2^{(j)}(t)$  and  $\lambda_*^{(j)}(t)$  denote, respectively, the hazard functions of  $T_2^{(j)}$  and  $T_*^{(j)}$  ( $j = 0, 1$ ). The treatment effect on  $\mathcal{E}_*$  would then be tested with the logrank test  $Z_*$  to compare  $H_0^*: \text{HR}_*(t) = \lambda_*^{(1)}(t)/\lambda_*^{(0)}(t) = 1$  versus  $H_a^*: \text{HR}_*(t) < 1$ .

Observation of endpoints  $\mathcal{E}_1$  and  $\mathcal{E}_2$  depends on whether or not they include a terminating event and yield four different situations referred, in Gómez and Lagakos (2013), as Cases 1, 2, 3 and 4. In this paper we assume that the additional endpoint does not include a terminating event, which corresponds to Case 1 when neither the relevant nor the additional endpoint includes a terminating event, and Case 3, when the relevant endpoint includes a terminating event.

Schoenfeld (1981) studies the asymptotic behaviour of the logrank statistic and proves that under the null hypothesis of no treatment difference, the logrank is asymptotically  $N(0, 1)$  and, under a sequence of alternatives contiguous to the null, the logrank is asymptotically normal with unit variance and finite mean. Gómez and Lagakos apply Schoenfeld's results and proceed as follows. They consider  $\lambda_1^{(0)}(t)$  as fixed and define a sequence of alternatives  $H_{a,n}$  consisting of instantaneous hazard functions close enough to  $\lambda_1^{(0)}(t)$ , for instance taking  $\lambda_{1,n}^{(1)}(t) = \lambda_1^{(0)}(t)e^{g(t)/\sqrt{n}}$  for some  $g(t)$  function. These sequence of alternatives, formulated equivalently as  $\text{HR}_{1,n}(t) = e^{g(t)/\sqrt{n}}$ , include pro-

portional hazard alternatives, i.e, taking  $g(t) = \beta$  for a fixed real value  $\beta$ . Logrank  $Z$  is asymptotically  $N(0, 1)$  under the null hypothesis of no treatment difference ( $H_0 : \text{HR}_1(t) = 1$ ) and asymptotically normal with unit variance and mean  $\mu$  given in equation (1) under the sequence of alternatives  $H_{a,n} : \text{HR}_{1,n}(t) = e^{g(t)/\sqrt{n}} < 1$ . Analogously, fix  $\lambda_*^{(0)}(t)$  and define  $H_0^* : \text{HR}_*(t) = 1$  and the sequence of alternatives  $H_{a,n}^* : \text{HR}_{*,n}(t) = e^{g_*(t)/\sqrt{n}} < 1$  for a given function  $g_*(t)$ . It follows that  $Z^*$  is asymptotically  $N(0, 1)$  under  $H_0^*$  and asymptotically normal with unit variance and mean  $\mu_*$  given in equation (2) under the sequence  $H_{a,n}^*$ . The asymptotic means of  $Z$  and  $Z^*$  are given by

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

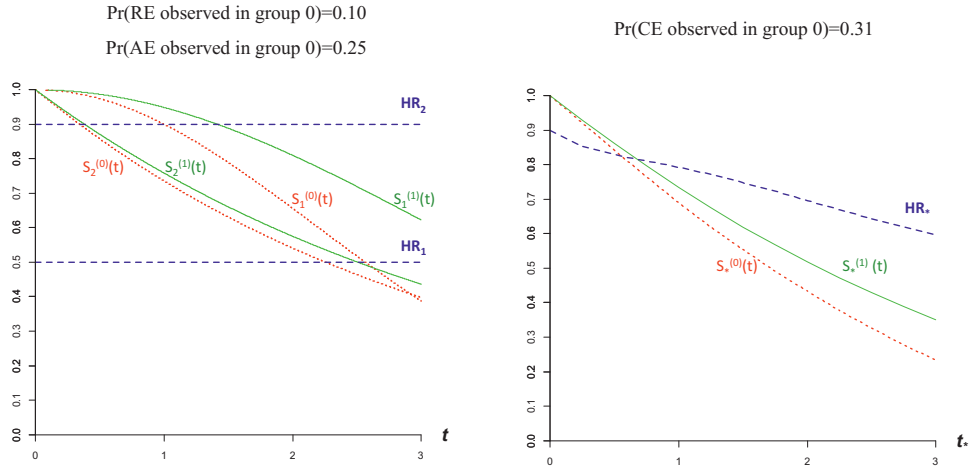
$$\mu_* = \frac{\int_0^\infty g_*(t)p_*(t)[1-p_*(t)]\Pr_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)dt}{\sqrt{\int_0^\infty p_*(t)[1-p_*(t)]\Pr_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)dt}}, \quad (2)$$

where  $U = \min\{T_1, C\}$  (in Cases 1 and 3) and  $U_* = \min\{T_*, C\}$  denote the observed outcome;  $C$  denotes the censoring time;  $p(t) = \Pr_{H_0}\{X = 1|U \geq t\}$  and  $p_*(t) = \Pr_{H_0^*}\{X = 1|U_* \geq t\}$  are the null probabilities that someone at risk at time  $t$  is in treatment group 1;  $\Pr_{H_0}\{U \geq t\}$  and  $\Pr_{H_0^*}\{U_* \geq t\}$  are the null probabilities that someone is still at risk at time  $t$  and  $\Pr_{H_0}\{U \geq t\}\lambda_1^{(0)}(t)$  and  $\Pr_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)$  correspond to the probabilities, under the null hypothesis, of observing events  $\mathcal{E}_1$  and  $\mathcal{E}_*$ , respectively, by time  $t$ .

## 2.2. Asymptotic relative efficiency

Efficiency calculations throughout the paper will assume that end-of-study censoring at time  $\tau$  ( $\tau = 1$  without loss of generality) is the only non-informative censoring cause for both groups; this assumption indirectly implies that the censoring mechanism is the same for both groups. It is as well assumed that the hazard functions  $\lambda_1^{(j)}(t)$  and  $\lambda_2^{(j)}(t)$  ( $j = 0, 1$ ) are proportional, that is,  $\text{HR}_1(t) = \text{HR}_1$  and  $\text{HR}_2(t) = \text{HR}_2$ , for all  $t$ , where  $\text{HR}_1(t) = \lambda_1^{(1)}(t)/\lambda_1^{(0)}(t)$  and  $\text{HR}_2(t) = \lambda_2^{(1)}(t)/\lambda_2^{(0)}(t)$  are the hazard ratios between  $T_1^{(0)}$  and  $T_1^{(1)}$  and between  $T_2^{(0)}$  and  $T_2^{(1)}$ , respectively. Note that although we are assuming that the hazard functions  $\lambda_1^{(j)}(t)$  and  $\lambda_2^{(j)}(t)$  ( $j = 0, 1$ ) are proportional, this does not imply the proportionality of hazards  $\lambda_*^{(0)}(t)$  and  $\lambda_*^{(1)}(t)$  for the composite endpoint  $T_*$  (see Figure 1).

To assess the difference in efficiency between using logrank test  $Z$ , based on the relevant endpoint  $\mathcal{E}_1$ , and logrank test  $Z_*$ , based on the composite endpoint  $\mathcal{E}_*$ , Gómez and Lagakos base their strategy on the behaviour of the asymptotic relative efficiency (ARE) of  $Z_*$  versus  $Z$ . The ARE is a measure of the relative power of two tests that can



**Figure 1:** Survival and hazard ratio for the relevant endpoint (RE),  $T_1$ , for the additional endpoint (AE),  $T_2$  and for the composite endpoint (CE),  $T_* = \min\{T_1, T_2\}$ .  $T_1 \sim$  Weibull with shape parameter  $\beta_1 = 2$  (increasing hazard) for treatment groups 0 and 1 and  $T_2 \sim$  Weibull with shape parameter  $\beta_2 = 1$  (constant hazard) for treatment groups 0 and 1. Scale parameters for  $T_1$  and  $T_2$  have been calculated such that  $\Pr\{T_1$  observed in group 0 $\}=0.1$ ,  $\Pr\{T_2$  observed in group 0 $\}=0.25$ ,  $HR_1 = 0.5$ ,  $HR_2 = 0.9$  and Spearman's  $\rho(T_1, T_2) = 0.45$  assuming Frank's copula between  $T_1$  and  $T_2$ . Considering the RE as a terminating event (case 3), in this setting  $ARE(Z_*, Z) = 0.21$ .

be interpreted, when the two tests are for the same null and alternative hypothesis, as the ratio of the required sample sizes to detect a specific treatment effect to attain the same power for a given significance level (Lehmann and Romano, 2005). In this case, a value of  $ARE = 0.6$  would mean that we only need 60% as many cases to reach a given power if we use  $\mathcal{E}_1$  as we would need if we used  $\mathcal{E}_*$ . Whenever the tests under consideration,  $Z$  and  $Z_*$ , are asymptotically  $N(0,1)$  under  $H_0$  and  $H_0^*$ , respectively, and asymptotically normal with variance 1 under a sequence of contiguous alternatives to the null hypothesis, a different definition for Pitman's relative efficiency as the square of the ratio of the non-centrality parameters  $\mu$  and  $\mu_*$  is appropriate

$$ARE(Z_*, Z) = \left( \frac{\mu_*}{\mu} \right)^2, \tag{3}$$

where  $\mu$  and  $\mu_*$  are to be replaced by expressions (1) and (2).

Before providing the expression that is being used to evaluate the ARE, and for the sake of clarity, we enumerate the assumptions that have been taken into account:

- End-of-study censoring at time  $\tau$  is the only non-informative censoring cause for both groups.
- The additional endpoint does not include a terminating event.

- The hazard ratios between  $T_1^{(0)}$  and  $T_1^{(1)}$  and between  $T_2^{(0)}$  and  $T_2^{(1)}$  are proportional, that is,  $\text{HR}_1(t) = \lambda_1^{(1)}(t)/\lambda_1^{(0)}(t) = \text{HR}_1$  and  $\text{HR}_2(t) = \lambda_2^{(1)}(t)/\lambda_2^{(0)}(t) = \text{HR}_2$  for all  $t$ .
- Effect of treatment on  $\mathcal{E}_1$  is tested establishing  $H_0: \text{HR}_1 = 1$  versus a sequence of alternatives  $H_{a,n}: \lambda_{1,n}^{(1)}(t) = \lambda_1^{(0)}(t)e^{g(t)/\sqrt{n}}$  for some  $g(t)$  function. Note that  $g(t)/\sqrt{n} = \log\{\lambda_{1,n}^{(1)}(t)/\lambda_1^{(0)}(t)\}$ .
- Effect of treatment on  $\mathcal{E}_*$  is tested establishing  $H_0^*: \text{HR}_*(t) = 1$  versus a sequence of alternatives  $H_{a,n}^*: \text{HR}_{*,n}(t) = e^{g_*(t)/\sqrt{n}} < 1$  for a given function  $g_*(t)$ . Note that  $g_*(t)/\sqrt{n} = \log\{\text{HR}_{*,n}(t)\}$ .

Under the above assumptions expression (3) becomes

$$\text{ARE}(Z_*, Z) = \frac{\left(\int_0^1 \log\{\lambda_*^{(1)}(t)/\lambda_*^{(0)}(t)\} f_*^{(0)}(t) dt\right)^2}{(\log\{\text{HR}_*\})^2 \left(\int_0^1 f_*^{(0)}(t) dt\right) \left(\int_0^1 f_1^{(0)}(t) dt\right)}, \quad (4)$$

where  $f_1^{(0)}(t)$  and  $f_*^{(0)}(t)$  are the density functions of  $T_1^{(0)}$  and  $T_*^{(0)}$ , respectively.

**Remark** The density function  $f_*^{(0)}(t)$  is the density of the  $T_*^{(0)} = \min\{T_1^{(0)}, T_2^{(0)}\}$ , computed from the joint density between  $T_1^{(0)}$  and  $T_2^{(0)}$ , which itself is built from the marginals of  $T_1^{(0)}$  and  $T_2^{(0)}$  by means of a bivariate copula.

### 3. Relationship between ARE and sample sizes

We start establishing that if the hazard ratios for  $T_1^{(j)}$  ( $j = 0, 1$ ) and for  $T_2^{(j)}$  ( $j = 0, 1$ ) approach the unity as  $n$  gets large, so does the hazard ratio of the minimum  $T_*^{(j)}$  between  $T_1^{(j)}$  and  $T_2^{(j)}$  ( $j = 0, 1$ ).

**Lemma 1** Given two sequences of hazard ratios  $\{\text{HR}_{1,n}(t) = \lambda_{1,n}^{(1)}(t)/\lambda_1^{(0)}(t)\}$  and  $\{\text{HR}_{2,n}(t) = \lambda_{2,n}^{(1)}(t)/\lambda_2^{(0)}(t)\}$ , both converging uniformly to 1 as  $n \rightarrow \infty$ , the sequence corresponding to the hazard ratio of  $T_*^{(j)} = \min\{T_1^{(j)}, T_2^{(j)}\}$ , namely  $\{\text{HR}_{*,n}(t) = \lambda_{*,n}^{(1)}(t)/\lambda_*^{(0)}(t)\}$ , tends to 1 as  $n \rightarrow \infty$ . In particular, this lemma holds whenever  $\log(\lambda_{k,n}^{(1)}(t)/\lambda_k^{(0)}(t)) = O(n^{-1/2})$ , which in turn, is true if  $\log(\lambda_{k,n}^{(1)}(t)/\lambda_k^{(0)}(t)) = g_k(t)/\sqrt{n}$ , for any bounded real function  $g_k(t)$  ( $k = 1, 2$ ).

*Proof* It follows immediately that for fixed  $t$ ,  $\lim_{n \rightarrow \infty} \lambda_{1,n}^{(1)}(t) = \lambda_1^{(0)}(t)$  and  $\lim_{n \rightarrow \infty} \lambda_{2,n}^{(1)}(t) = \lambda_2^{(0)}(t)$ . Furthermore, it follows that the corresponding densities and

survival functions  $f_{1,n}^{(1)}(t)$ ,  $f_{2,n}^{(1)}(t)$ ,  $S_{1,n}^{(1)}(t)$  and  $S_{2,n}^{(1)}(t)$ , converge to  $f_1^{(0)}(t)$ ,  $f_2^{(0)}(t)$ ,  $S_1^{(0)}(t)$  and  $S_2^{(0)}(t)$ , respectively. Taking into account that the survival function of the minimum,  $S_{*,n}^{(1)}(t)$  is expressed in terms of the marginal survival functions  $S_{1,n}^{(1)}(t)$  and  $S_{2,n}^{(1)}(t)$  of  $T_1^{(1)}$  and  $T_2^{(1)}$  via a copula  $C$ , that is,

$S_{*,n}^{(1)}(t) = C(S_{1,n}^{(1)}(t), S_{2,n}^{(1)}(t))$ , it remains to prove that  $\lim_{n \rightarrow \infty} S_{*,n}^{(1)}(t) = S_*^{(0)}(t)$ . This result will imply that

$\lim_{n \rightarrow \infty} f_{*,n}^{(1)}(t) = f_*^{(0)}(t)$ ,  $\lim_{n \rightarrow \infty} \lambda_{*,n}^{(1)}(t) = \lambda_*^{(0)}(t)$  and hence the sequence  $\text{HR}_{*,n}(t) \rightarrow 1$  as  $n \rightarrow \infty$ , as we wanted to prove.

The convergence of  $S_{*,n}^{(1)}(t)$  to  $S_*^{(0)}(t)$  is guaranteed by the convergence of  $S_{1,n}^{(1)}(t)$  and  $S_{2,n}^{(1)}(t)$  to  $S_1^{(0)}(t)$  and  $S_2^{(0)}(t)$ , respectively, together with the fact that bivariate copulas  $C$  are bivariate distribution functions with uniform marginals. The reader is referred to Lindner and Szimayer (2005) for the corresponding technical proofs.  $\square$

**Proposition 1** Consider two test procedures  $\phi_n$  and  $\phi_n^*$  to test  $H_0 : \text{HR}_1(t) = 1$  against  $H_{a,n} : \text{HR}_{1,n}(t) < 1$  and  $H_0^* : \text{HR}_*(t) = 1$  against  $H_{a,n}^* : \text{HR}_{*,n}(t) < 1$ , respectively. Let  $n$  and  $n_*$  be the sample sizes required for  $\phi_n$  and  $\phi_n^*$ , respectively, to have power at least  $\Pi$  at level  $\alpha$ . Assume the sequences  $\phi = \{\phi_n\}$  and  $\phi^* = \{\phi_n^*\}$  are based on the logrank statistics  $Z$  and  $Z^*$ , respectively, converging, to Normal  $(\mu, 1)$  and Normal  $(\mu_*, 1)$  with  $\mu$  and  $\mu_*$  given in (1) and (2), under sequences of local alternatives  $\text{HR}_{k,n}(t)$  ( $k = 1, 2$ ) converging uniformly to 1 as  $n \rightarrow \infty$ . Given  $0 < \alpha < \Pi < 1$ ,

$$\lim_{\substack{\text{HR}_{1,n}(t) \rightarrow 1 \\ \text{HR}_{2,n}(t) \rightarrow 1}} \frac{n}{n_*} = \text{ARE}(Z_*, Z).$$

The usual interpretation of the ARE as the reciprocal ratio of the sample sizes holds even when two different sets of hypotheses ( $H_0$  versus  $H_{a,n}$  and  $H_0^*$  versus  $H_{a,n}^*$ ) are tested. As a consequence of this proposition, the interpretation of the ARE is the following. If  $\text{ARE}(Z_*, Z) = 0.7$ , then, asymptotically, we only need 70% as many cases to attain a given power if we use  $Z$  as we would need if we used  $Z_*$ .

*Proof 2* By Lemma 1, uniform convergence to 1 of  $\{\text{HR}_{1,n}(t)\}$  and  $\{\text{HR}_{2,n}(t)\}$  imply that  $\lim \text{HR}_{*,n}(t) \rightarrow 1$ . Under the sequence of contiguous alternatives to the null  $H_{a,n} : \{\text{HR}_{1,n}(t) = \lambda_{1,n}^{(1)}(t)/\lambda_1^{(0)}(t)\} \rightarrow 1$  and  $H_{a,n}^* : \{\text{HR}_{*,n}(t) = \lambda_{*,n}^{(1)}(t)/\lambda_*^{(0)}(t)\} \rightarrow 1$ , both  $Z$  and  $Z^*$  are asymptotically  $N(\mu, 1)$  and  $N(\mu_*, 1)$ , respectively. The power function for a one-sided test with size  $\alpha$  is therefore given, respectively, by

$$\begin{aligned} \Pi_1 &= \lim_{n \rightarrow \infty} \text{Prob}\{Z < z_{1-\alpha} | H_{a,n}\} = 1 - \Phi(-z_{1-\alpha} + \mu) \\ \Pi_* &= \lim_{n \rightarrow \infty} \text{Prob}\{Z_* < z_{1-\alpha} | H_{a,n}^*\} = 1 - \Phi(-z_{1-\alpha} + \mu_*) \end{aligned} \tag{5}$$

where  $\Phi$  is the distribution function of the standard normal and  $z_{1-\alpha}$  is the standard normal quantile corresponding to the left tail probability  $\alpha$ . It immediately follows that  $\Pi_1 = \Pi_*$  is equivalent to  $\mu = \mu_*$ .

The equivalence of powers ( $\Pi_1 = \Pi_*$ ) implies that  $\mu = \mu_*$ , given by (1) and (2). Equivalently

$$\left(\frac{\mu_*}{\mu}\right)^2 = 1 \iff \left(\frac{\frac{\int_0^\infty g(t)p(t)[1-p(t)]\Pr_{H_0}\{U \geq t\}\lambda_1^{(0)}(t)dt}{\sqrt{\int_0^\infty p(t)[1-p(t)]\Pr_{H_0}\{U \geq t\}\lambda_1^{(0)}(t)dt}}}{\frac{\int_0^\infty g_*(t)p_*(t)[1-p_*(t)]\Pr_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)dt}{\sqrt{\int_0^\infty p_*(t)[1-p_*(t)]\Pr_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)dt}}}\right)^2 = 1. \quad (6)$$

Since

$$p(t) = \frac{\Pr_{H_0}\{U \geq t|X = 1\}\pi}{\Pr_{H_0}\{U \geq t\}} = \frac{\Pr_{H_0}\{U^{(j)} \geq t\}\pi}{\Pr_{H_0}\{U \geq t\}}$$

where  $\pi = \Pr_{H_0}\{X = 1\}$ , we have

$$p(t)(1-p(t))\Pr_{H_0}\{U \geq t\} = \frac{\Pr_{H_0}\{U^{(1)} \geq t\}\pi\Pr_{H_0}\{U^{(0)} \geq t\}(1-\pi)}{\Pr_{H_0}\{U^{(0)} \geq t\}(1-\pi) + \Pr_{H_0}\{U^{(1)} \geq t\}\pi}.$$

Based on the stated assumptions, because  $T_1^{(j)}$  is right-censored by the end-of-study at time  $\tau$ , and under the null hypothesis of no effect ( $S_1^{(0)}(t) = S_1^{(1)}(t)$ ), we have  $\Pr_{H_0}\{U^{(j)} \geq t\} = S_1^{(0)}(t)1\{[0, 1]\}(t)$ , for  $j = 0, 1$ . Replacing in (1), the noncentrality parameter  $\mu$  becomes

$$\mu = \frac{\sqrt{\pi(1-\pi)}\int_0^1 g(t)S_1^{(0)}(t)\lambda_1^{(0)}(t)dt}{\sqrt{\int_0^1 S_1^{(0)}(t)\lambda_1^{(0)}(t)dt}} = \frac{\sqrt{\pi(1-\pi)}\int_0^1 g(t)f_1^{(0)}(t)dt}{\sqrt{\int_0^1 f_1^{(0)}(t)dt}}$$

where  $f_1^{(0)}(t)$  is the marginal density function for  $T_1^{(0)}$ . Analogously, it can be seen that

$$\mu_* = \frac{\sqrt{\pi(1-\pi)}\int_0^1 g_*(t)f_*^{(0)}(t)dt}{\sqrt{\int_0^1 f_*^{(0)}(t)dt}}$$

where  $f_*^{(0)}(t)$  is the density function for  $T_*^{(0)}$ . The reader is addressed to the online supporting material of Gómez and Lagakos paper for other technical details.

If we would replace  $g(t)$  and  $g_*(t)$  by  $\sqrt{n} \log \left( \frac{\lambda_{1,n}^{(1)}(t)}{\lambda_1^{(0)}(t)} \right) = \sqrt{n} \log(\text{HR}_1)$  and  $\sqrt{n_*} \log \left( \frac{\lambda_{*,n}^{(1)}(t)}{\lambda_*^{(0)}(t)} \right)$ , respectively, equality (6), after cancelling  $\pi(1 - \pi)$ , becomes equal to

$$\lim_{\substack{\text{HR}_{1,n}(t) \rightarrow 1 \\ \text{HR}_{2,n}(t) \rightarrow 1}} \frac{\sqrt{n_*} \frac{\int_0^1 \log \{ \lambda_*^{(1)}(t) / \lambda_*^{(0)}(t) \} f_*^{(0)}(t) dt}{\sqrt{\int_0^1 f_*^{(0)}(t) dt}}}{\sqrt{n} \log(\text{HR}_1) \sqrt{\int_0^1 f_1^{(0)}(t) dt}} = 1$$

which in turn is equivalent to

$$\lim_{\substack{\text{HR}_{1,n}(t) \rightarrow 1 \\ \text{HR}_{2,n}(t) \rightarrow 1}} \frac{n}{n_*} = \frac{\left( \int_0^1 \log \{ \lambda_*^{(1)}(t) / \lambda_*^{(0)}(t) \} f_*^{(0)}(t) dt \right)^2}{(\log(\text{HR}_1))^2 \left( \int_0^1 f_*^{(0)}(t) dt \right) \left( \int_0^1 f_1^{(0)}(t) dt \right)} \quad (7)$$

and it follows that  $\text{ARE}(Z_*, Z) = \lim_{\substack{\text{HR}_{1,n}(t) \rightarrow 1 \\ \text{HR}_{2,n}(t) \rightarrow 1}} \frac{n}{n_*}$ , as we wanted to prove.  $\square$

Note that (7) implies

$$\frac{\left( \int_0^1 \log \{ \lambda_*^{(1)}(t) / \lambda_*^{(0)}(t) \} f_*^{(0)}(t) dt \right)^2}{(\log(\text{HR}_1))^2 \left( \int_0^1 f_*^{(0)}(t) dt \right)^2} = \lim_{\substack{\text{HR}_{1,n}(t) \rightarrow 1 \\ \text{HR}_{2,n}(t) \rightarrow 1}} \frac{n \left( \int_0^1 f_1^{(0)}(t) dt \right)}{n_* \left( \int_0^1 f_*^{(0)}(t) dt \right)} \approx \frac{\text{expected number } \mathcal{E}_1}{\text{expected number } \mathcal{E}_*}$$

and whenever  $\lambda_*^{(1)}(t) / \lambda_*^{(0)}(t)$  is approximately constant and equal to  $\text{HR}_*$ , we would have

$$\frac{\left( \frac{1}{\log(\text{HR}_1)} \right)^2}{\left( \frac{1}{\log(\text{HR}_*)} \right)^2} = \lim_{\substack{\text{HR}_{1,n}(t) \rightarrow 1 \\ \text{HR}_{2,n}(t) \rightarrow 1}} \frac{n \left( \int_0^1 f_1^{(0)}(t) dt \right)}{n_* \left( \int_0^1 f_*^{(0)}(t) dt \right)} \approx \frac{\text{expected number } \mathcal{E}_1}{\text{expected number } \mathcal{E}_*}$$

## 4. Simulation

### 4.1. Simulation

Our next aim is to simulate data to empirically check how close we are to the limiting relationship  $n/n_* = \text{ARE}(Z_*, Z)$  when  $\Pi_1 = \Pi_*$  for different finite sample sizes. To



conduct the simulations we will assume, as Gómez and Lagakos did, that  $T_1^{(j)}$  and  $T_2^{(j)}$  follow Weibull distributions. Weibull distributions are chosen for their wide use in the field of survival analysis due to its flexibility, allowing decreasing, constant and increasing hazard rates. The corresponding shape and scale parameters are denoted by  $\beta_k$  and  $b_k^{(j)}$  ( $j = 0, 1, k = 1, 2$ ) (shape parameters for both groups are taken equal so that the assumption of the proportionality of the hazard ratios holds). To establish the bivariate distribution of  $(T_1^{(0)}, T_2^{(0)})$  we consider Frank's Archimedean survival copula, again as Gómez and Lagakos did. Other choices of copulas would be possible, although main conclusions and recommendations will not differ (Plana-Ripoll and Gómez, 2014). Frank's copula depends on an association parameter  $\theta$  between  $T_1^{(0)}$  and  $T_2^{(0)}$  which is biunivocally related to Spearman's rank correlation  $\rho$ . Different scenarios will be simulated according to several choices of  $(\beta_1, \beta_2, p_1^{(0)}, p_2^{(0)}, \text{HR}_1, \text{HR}_2, \rho)$  where  $p_1^{(0)}$  and  $p_2^{(0)}$  are the probability of observing events  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , respectively, for treatment group 0,  $\text{HR}_1$  and  $\text{HR}_2$  are relative treatment hazard ratios for  $T_j^{(1)}$  versus  $T_j^{(0)}$  ( $j = 1, 2$ , respectively) and  $\rho$  is Spearman's rank correlation between  $T_1^{(0)}$  and  $T_2^{(0)}$ .

Given a set of values for  $(\beta_1, \beta_2, p_1^{(0)}, p_2^{(0)}, \text{HR}_1, \text{HR}_2, \rho)$ , for a given power  $\Pi$  and a significance level  $\alpha$ , the simulation steps are the following:

1. **Computations for the relevant endpoint  $\mathcal{E}_1$ .** The scale parameters  $b_1^{(0)}$  and  $b_1^{(1)}$  and the probability  $p_1^{(1)}$  of observing the relevant endpoint in group 1 are derived as:

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

$$b_1^{(1)} = \frac{b_1^{(0)}}{\text{HR}_1^{(1/\beta_1)}}$$

$$p_1^{(1)} = 1 - e^{-(1/b_1^{(1)})^{\beta_1}}$$

2. **Computations for the additional endpoint  $\mathcal{E}_2$ .** The scale parameters  $b_2^{(0)}$  and  $b_2^{(1)}$  and the probability  $p_2^{(1)}$  of observing the additional endpoint in group 1 are derived as:

$$b_2^{(0)} = \begin{cases} \frac{1}{(-\log(1 - p_2^{(0)}))^{1/\beta_2}} & \text{for Case 1} \\ * & \text{for Case 3} \end{cases}$$

$$b_2^{(1)} = \frac{b_2^{(0)}}{\text{HR}_2^{(1/\beta_2)}}$$

$$p_2^{(1)} = 1 - e^{-(1/b_2^{(1)})^{\beta_2}}$$

\* For Case 3,  $b_2^{(0)}$  is found as the solution of equation  $p_2^{(1)} = \int_0^1 \int_u^1 f_{(1,2)}^{(0)}(u, v; \rho) dv du$ , where  $f_{(1,2)}^{(0)}(\cdot, \cdot; \rho)$  is the joint density between  $T_1^{(0)}$  and  $T_2^{(0)}$  and  $\rho$  is Spearman's  $\rho$  coefficient between  $T_1^{(0)}$  and  $T_2^{(0)}$ .

**3. Computation of sample sizes  $n$  and  $n_*$**

(a) Compute  $n$  (per group) following Freedman (1982) formulas as follows

$$n = \frac{E}{p_1^{(0)} + p_1^{(1)}} \tag{8}$$

where

$$E = \frac{(HR_1 + 1)^2(z_{1-\alpha} + z_{\Pi})^2}{(HR_1 - 1)^2} \tag{9}$$

(b) Compute  $ARE(Z_*, Z)$  based on  $(\beta_1, \beta_2, p_1^{(0)}, p_2^{(0)}, HR_1, HR_2, \rho)$ .

(c) Compute  $n_* = n / ARE(Z_*, Z)$ .

(d) Compute  $N = \max\{n, n_*\}$ .

**4. Simulation of  $T_1^{(0)}, T_1^{(1)}, T_2^{(0)}, T_2^{(1)}, T_*^{(0)}, T_*^{(1)}$**

Simulate 1000 samples of size  $N$  for the 4 endpoints  $T_k^{(j)}$  from Weibull  $(b_k^{(j)}, \beta_k)$  ( $j = 0, 1, k = 1, 2$ ). Compute  $T_*^{(j)} = \min\{T_1^{(j)}, T_2^{(j)}\}$ .

**5. Computation of empirical powers  $\hat{\Pi}_1$  and  $\hat{\Pi}_*$**

For each sample of size  $n$  ( $n_*$ ), compute the logrank statistic  $Z$  ( $Z_*$ ) to compare the treatment effect between  $T_1^{(0)}$  and  $T_1^{(1)}$  ( $T_*^{(0)}$  and  $T_*^{(1)}$ ). For a given significance level  $\alpha$ , the rejection region comprises all observed  $Z$  ( $Z_*$ ) such that  $Z < z_{1-\alpha}$  ( $Z_* < z_{1-\alpha}$ ) where  $z_{1-\alpha}$  is the standard normal quantile corresponding to the left tail probability  $\alpha$ . The empirical powers, denoted by  $\hat{\Pi}_1$  ( $\hat{\Pi}_*$ ), are calculated as the proportion of samples for which  $Z < z_{1-\alpha}$  ( $Z_* < z_{1-\alpha}$ ).

We note here that whenever  $n_* < n$ , we only use, for each sample, the first  $n_*$  simulated values to compute  $\hat{\Pi}_*$ , while when  $n < n_*$ , we only use the first  $n$  simulated values to compute  $\hat{\Pi}_1$ .

**6. Comparison between  $\hat{\Pi}_1$  and  $\hat{\Pi}_*$**

For each scenario  $(\beta_1, \beta_2, p_1^{(0)}, p_2^{(0)}, HR_1, HR_2, \rho)$ , we compare the differences between the two empirical powers  $\hat{\Pi}_1$  and  $\hat{\Pi}_*$  obtained from the 1000 simulations.

**Table 1:** Values of parameters  $\beta_1$ ,  $\beta_2$ ,  $p_1$ ,  $p_2$ ,  $HR_1$ ,  $HR_2$  and  $\rho$  used for the simulations. There are 624 different configurations, excluding those yielding sample sizes larger than 1100 and  $ARE(Z_*, Z) > 10$ .

Parameters						
$\beta_1 = \beta_2$	0.5	1	2			
$(p_1, p_2)$	(0.05, 0.01)	(0.05, 0.15)	(0.05, 0.35)	(0.1, 0.01)	(0.1, 0.15)	(0.1, 0.35)
$(p_1, p_2)$	(0.15, 0.01)	(0.15, 0.15)	(0.15, 0.35)	(0.35, 0.01)	(0.35, 0.15)	(0.35, 0.35)
$\rho$	0.15	0.45	0.75			
$(HR_1, HR_2)$	(0.5, 0.3)	(0.5, 0.7)	(0.5, 0.9)	(0.6, 0.3)	(0.6, 0.7)	(0.6, 0.9)
$(HR_1, HR_2)$	(0.7, 0.3)	(0.7, 0.7)	(0.7, 0.9)	(0.8, 0.3)	(0.8, 0.7)	
Total number						
of cases	624					

## 4.2. Results

We have set  $\Pi = 0.9$  and  $\alpha = 0.05$  (other values would not provide additional information). We have chosen meaningful values for  $(\beta_1, \beta_2, p_1^{(0)}, p_2^{(0)}, HR_1, HR_2, \rho)$ , based on those arising in cardiovascular clinical trials (Gómez, Gómez-Mateu, Dafni, 2014) (see Table 1). We restrict our simulation study to 624 scenarios corresponding to  $ARE(Z_*, Z) \leq 10$  and sample sizes smaller than 1100 patients per group. These scenarios yield  $ARE(Z_*, Z)$  values between 0.20 and 9.93, sample sizes,  $n$ , for the relevant endpoint between 142 and 1081, and,  $n_*$ , for the composite endpoint between 53 and 1077 (see Table 2). Similar results were obtained for Case 1, when neither the relevant nor the additional endpoint includes a terminating event, and for Case 3 when the relevant endpoint includes a terminating event, and we only discuss here Case 1.

**Table 2:** Computed values of  $n$ ,  $n_*$  and  $ARE(Z_*, Z)$  in step 3 of the simulation based on the parameter values given in Table 1.

	min	median	max
$n$	142	509	1081
$n_*$	53	398	1077
$ARE(Z_*, Z)$	0.2	1.04	9.93

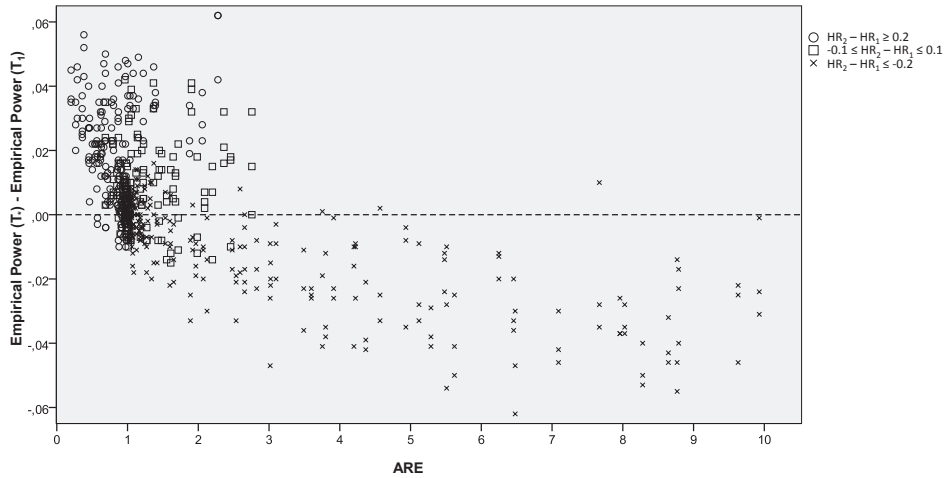
The empirical powers  $\hat{\Pi}_1$  in our simulation study resulted in powers between 0.87 and 0.94, with a median of 0.91. A slightly higher median was found for scenarios with low hazard ratios. This finding is acknowledged as well by Freedman (1982).

Table 3 provides the percentiles for the absolute value differences between  $\hat{\Pi}_*$  and  $\hat{\Pi}_1$ . We observe that in 75% of the cases the difference is smaller than 2.3%, and among cases with  $ARE$  as large as 3 the difference shrinks to 1.9%. There are, however, few instances, where this difference can be as large as 6%, and they deserve a closer look.

**Table 3:** Percentiles of  $|\hat{\Pi}_* - \hat{\Pi}_1|$  as a function of ARE values, where  $w_i$  indicates the corresponding percentile.

	min	$w_{0.1}$	$w_{0.25}$	$w_{0.5}$	$w_{0.75}$	$w_{0.9}$	max
For all ARE	0	0.002	0.004	0.010	0.023	0.036	0.062
$ARE(Z_*, Z) \leq 3$	0	0.002	0.004	0.008	0.019	0.033	0.062
$ARE(Z_*, Z) > 3$	0.001	0.009	0.016	0.026	0.038	0.046	0.062

Figure 2 plots the differences  $\hat{\Pi}_* - \hat{\Pi}_1$  as a function of the  $ARE(Z_*, Z)$  values. The behaviour is remarkably different when  $ARE(Z_*, Z) \leq 3$  or  $ARE(Z_*, Z) > 3$ . Whenever  $ARE(Z_*, Z) \leq 3$ ,  $\hat{\Pi}_*$  fluctuates around  $\hat{\Pi}_1$ , within a range of 4%. However, when  $ARE(Z_*, Z) > 3$ , corresponding mostly to scenarios where treatment has a stronger effect on the additional endpoint than on the relevant endpoint ( $HR_2 \leq HR_1 - 0.2$ ) and the anticipated number of events in the control group is larger for the additional endpoint than for the relevant ( $p_2^{(0)} \geq p_1^{(0)}$ ), the empirical power  $\hat{\Pi}_*$  of the logrank test based on the CE never achieves the same power as the logrank test for the relevant endpoint would get. In these cases the interpretation of the  $ARE(Z_*, Z)$  as the ratio of the sample sizes,  $n/n_*$ , is not as straightforward. Nevertheless, this does not mean that the recommendation of using the CE does not have to be followed since larger values for  $n_*$  needed to attain the same power as  $n$  does, would reduce the ARE value but not as much as to cross the “1” border that would imply to use the relevant endpoint instead of the CE.



**Figure 2:** Differences between empirical powers  $\hat{\Pi}_* - \hat{\Pi}_1$  as function of  $ARE(Z_*, Z)$  and in terms of  $HR_2 - HR_1$ .

If we analyze the differences between  $\hat{\Pi}_*$  and  $\hat{\Pi}_1$  as a function of the differences between the two hazard ratios ( $HR_2 - HR_1$ ), we observe that when the two hazard ratios are very close, the two empirical powers are as well very close. Whenever  $HR_2 - HR_1 \leq -0.2$ , not only  $ARE(Z_*, Z)$  values tend to be higher, but also  $\hat{\Pi}_* < \hat{\Pi}_1$ . (see Figure 2).

Taking into account that absolute differences between powers smaller than 5% could be considered irrelevant, we conclude that the asymptotic relationship  $ARE(Z_*, Z) = n/n_*$  is valid in the majority of scenarios.

All computations in this paper have been implemented in R and are available on request to either author.

## 5. Discussion

Pitman's relative efficiency is defined as the limiting ratio of sample sizes to give the same asymptotic power under sequences of local alternatives. Given two asymptotically standard normal tests  $S_n$  and  $T_m$  under the same null and alternative hypotheses, the alternative definition  $ARE = (\mu_S/\mu_T)^2$  where  $\sqrt{n}\mu_S$  and  $\sqrt{m}\mu_T$  are the respective means under local alternatives, can be used because the equality of the powers holds if  $\frac{m}{n} = \left(\frac{\mu_S}{\mu_T}\right)^2$ .

Gómez and Lagakos' method uses the alternative definition of ARE to develop all the computations for the two corresponding logrank tests. Our goal has been to check that the relationship between  $(\mu_S/\mu_T)^2$  and the ratio of sample sizes still held when the two hypotheses under test were not the same ( $H_0$  versus  $H_a$  and  $H_0^*$  versus  $H_a^*$ ).

It is important to keep in mind that these two hypotheses tests are by no means equivalent, for instance, to check whether treatment has a beneficial effect, we might use  $\mathcal{E}_1$  or we might add endpoint  $\mathcal{E}_2$  and use  $\mathcal{E}_*$ . As it is shown in Gómez (2011), even if we assume that the times to  $\mathcal{E}_1$  and to  $\mathcal{E}_2$  are independent, a beneficial effect on  $\mathcal{E}_*$  can occur simultaneously with a beneficial effect on  $\mathcal{E}_1$  and a harmful effect on  $\mathcal{E}_2$  and not finding a beneficial effect on the composite event  $\mathcal{E}_*$  is no guarantee of not having some effect on the individual events  $\mathcal{E}_1$  or  $\mathcal{E}_2$ .

The main result of this paper proves that  $ARE(Z_*, Z)$  coincides with  $n/n_*$ , being  $n$  and  $n_*$  the sample sizes needed to detect specific alternatives  $HR_1$  and  $HR_2$  to attain power  $\Pi$  and for the same significance level  $\alpha$ . Therefore, we can use and interpret ARE in its usual way.

The simulation study has been conducted in such a way that for fixed values  $n$  and  $ARE(Z_*, Z)$ , the sample size  $n_*$  is calculated as  $n_* = n/ARE(Z_*, Z)$ . Hence an approximate equality of the empirical powers  $\hat{\Pi}_1$ , of logrank test  $Z$  for  $H_0$  versus  $H_{a,n}$ , and of  $\hat{\Pi}_*$  of logrank test  $Z_*$  for  $H_0^*$  versus  $H_{a,n}^*$ , indicates that the relationship  $ARE(Z_*, Z) = n/n_*$  holds. Main results from our simulations show that the absolute differences between  $\hat{\Pi}_1$  and  $\hat{\Pi}_*$  are most of the times less than 2.5%, hence the usual interpretation between  $(n, n_*)$  and  $ARE(Z_*, Z)$  holds for finite sample sizes.

For those scenarios under which  $ARE(Z_*, Z) > 3$ , we observe that the empirical power of the test based on  $\mathcal{E}_*$  never achieves the empirical power that the logrank test based on  $\mathcal{E}_1$  would get. Consequently, larger values of  $n_*$  would be needed to attain the same power as  $n$  does. In these instances, even though the relationship  $ARE(Z_*, Z) = n/n_*$  is not necessarily true, the recommendation to use the composite endpoint  $\mathcal{E}_*$  instead of the relevant endpoint  $\mathcal{E}_1$  will still be valid because very rarely a value of  $ARE(Z_*, Z) > 3$  would go down to less than 1. However, caution will be needed if one wants to use the relationship  $ARE(Z_*, Z) = n/n_*$  to compute the required sample size  $n_*$  if  $ARE(Z_*, Z) > 3$ . In these cases, a different formulation should be seek.

## References

- Ferreira-González, I., Permanyer-Miralda, G., Busse, J.W., Bryant, D.M., Montori, V.M., Alonso-Coello, P., Walter, S.D. and Guyatt, G.H. (2007). Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *Journal of Clinical Epidemiology*, 60, 651–657.
- Freedman, L.S. (1982). Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine*, 1, 121–129.
- Freemantle, N., Calvert, M., Wood, J., Eastaugh, J. and Griffin, C. (2003). Composite outcomes in Randomized Trials. Greater precision but with greater uncertainty? *Journal of the American Medical Association*, 289, 2554–2559.
- Gómez, G. (2011). Some theoretical thoughts when using a composite endpoint to prove the efficacy of a treatment. *International Workshop on Statistical Modelling. Proceedings of the 26th International Workshop on Statistical Modelling*, València 14–21. <http://hdl.handle.net/2117/22571>. Last accessed 19 May 2014.
- Gómez, G., Gómez-Mateu, M. and Dafni, U. (2014). Informed Choice of Composite End Points in Cardiovascular Trials. *Circulation. Cardiovascular Quality and Outcomes*, 7, 170–178.
- Gómez, G. and Lagakos, S.W. (2013). Statistical considerations when using a composite endpoint for comparing treatment groups. *Statistics in Medicine*, 32, 719–738.
- Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, 3rd Ed. Springer.
- Lindner, A.M. and Szimayer, A. (2005). A limit theorem for copulas. *Sonderforschungsbereich 386, Paper 433*. <http://epub.ub.uni-muenchen.de/1802>. Last accessed 19 May 2014.
- Montori, V.M., Permanyer-Miralda, G., Ferreira-González, I., Busse, J.W., Pacheco-Huergo, V., Bryant, D., Alonso, J., Akl, E.A., Domingo-Salvany, A., Mills, E., Wu, P., Schnemann, H.J., Jaeschke, R. and Guyatt, G.H. (2005). Validity of composite end points in clinical trials. *British Medical Journal*, 330, 594–596.
- Plana-Ripoll, O. and Gómez, G. (2014). Extension of the ARE method to select the Primary Endpoint in a Randomized Clinical Trial. *Submitted*.
- Schoenfeld, D. (1981). The Asymptotic Properties of Nonparametric Tests for Comparing Survival Distributions. *Biometrika*, 68, 316–319.
- Wittkop, L., Smith, C., Fox, Z., Sabin, C., Richert, L., Aboulker, J.P., Phillips, A., Chêne, G., Babiker, A. and Thiébaud, R. on behalf of NEAT-WP4. (2010). Methodological issues in the use of composite endpoints in clinical trials: examples from the HIV field. *Clinical Trials*, 7, 19–35.



# Comments on ‘Use of composite endpoints in clinical trials’ by Abdul J. Sankoh, Haihong Li and Ralph B. D’Agostino, Sr

From: *Guadalupe Gómez*  
*Moisés Gómez-Mateu*  
*Department of Statistics and Operations Research,*  
*Universitat Politècnica de Catalunya,*  
*Barcelona, Spain*

## 1. Introduction

The paper by Sankoh *et al.* [1] reviews the use of composite endpoints (CE) in randomized clinical trials and advocates for its use because it leads to higher event rates. They also claim that ‘highly correlated components do not add trial efficiency regarding gain in the overall event rate compared with disparate or independent components’.

The purpose of this letter is to illustrate several situations where, in the presence of highly correlated events, the more efficient primary endpoint (PE) is not always based on one single component. The efficiency of the CE versus one of its components can be quantified by means of the Asymptotic Relative Efficiency (ARE) [2] and with the aid of CompARE [3], a new web-based platform. We will base this illustration on the Losartan Intervention For Endpoint reduction in hypertension study (LIFE) trial [4].

Assume that the PE could be based either on a relevant endpoint (RE) or on a union of a RE and an additional endpoint (AE). Let  $T_R$  and  $T_A$  be the times from randomization to RE and AE, respectively. Under reasonable assumptions, the ARE, equivalent to the ratio of the sample sizes needed to attain the same power for a fixed significance level [5], is expressed as a function of the following: (i) the marginal laws of  $T_R$  and  $T_A$ ; (ii) the probabilities  $p_R$  and  $p_A$  of observing RE and AE in the control group; (iii) the treatment effects given by the hazard ratios  $HR_R$  and  $HR_A$  between the two treatment groups; and (iv) the correlation between  $T_R$  and  $T_A$ . Whenever anticipated values for the aforementioned parameters can be provided, the ARE can be evaluated and can guide the choice of the PE.

In the LIFE trial, the PE was composed of two clinically important outcomes, cardiovascular death (CVD) and myocardial infarction (MI), and a softer outcome, stroke (ST). We use the values in Table I (Table 3 of the original paper [4]) as the feasible anticipated values to study which would have been the most efficient PE. Assume that, within the Atenolol group, the probability of observing CVD or MI is  $p_R = 0.06$  (note that the probability of observing CVD and MI are, respectively, 0.05 and 0.04) and of observing ST is  $p_A = 0.07$ . The plots in Figure 1, which have been computed for the aforementioned parameter values, illustrate how much more efficient it would be to add ST to CVD or MI. The plots assume anticipated hazard ratios for CVD or MI equal to  $HR_R = 0.89$  (left) or  $HR_R = 0.76$  (right), hazard ratios on ST varying from 0.75 to 0.89, and all possible correlations between (CVD or MI) and ST. Observe that the ARE value heavily depends on the correlation between CVD or MI and ST, with highly correlated situations leading to smaller ARE values. When  $HR_R = 0.89$ , the ARE values are always larger than 1, even when stroke is highly correlated with CVD or MI, and the recommendation should have been to include ST into the definition of the PE. If a stronger treatment effect on CVD or MI is anticipated ( $HR_R = 0.76$ ), and the hazard ratio on ST is smaller than 0.80, the ARE is still  $> 1$  and, again, ST should be included as part of the PE. However, for hazard ratios larger than 0.80 and moderate or strong correlation between (CVD or MI) and ST, ARE is smaller than 1 and it would be much wiser not to include ST into the definition of the PE.