MASTER'S DEGREE THESIS

MSc in Statistics and Operations Research

Title: Statistical Efficiency of Survival Endpoints in Breast Cancer Clinical Trials

Author: Jordi Ferrer Playán

Advisor: Guadalupe Gómez Melis

Co-advisor: Urania Dafni

Department: Statistics and Operations Research

Academic year: 2011/2012

Facultat de Matemàtiques i Estadística

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Statistical Efficiency of Survival Endpoints in Breast Cancer Clinical Trials

Author:

JORDI FERRER PLAYÁN

Main Advisor:

GUADALUPE GÓMEZ MELIS

Department of Statistics and Operations Research
Universitat Politècnica de Catalunya

Co-advisor:

URANIA DAFNI

Professor of Biostatistics and Director of the Laboratory of Biostatistics
Faculty of Nursing. School of Health Sciences. University of Athens

Acknowledgements

I would like to thank Professor Lupe Gómez, this Master's Thesis director, for her dedication and support, her thorough reviews and her valuable feedback; and also for introducing me to the fascinating world of the survival analysis.

I would like to thank also Professor Urania Dafni for her support and comments throughout the duration of this Master Thesis.

And finally I would like to thanks Moisés Gómez from the Universitat Politècnica de Catalunya for sharing with the the *R* programming he developed and giving me support when I needed.

Summary

The drug development and marketing process is long and runs through different phases where the tested drugs must show how effective and how safe they are. This process may last up to 12 years, and only a very low percentage of the drugs that start the first phases of development succeed and reaches the market.

Time is not the only concern. The investment in development programs is significant, not only to cover the expenses of an R&D department dedicated to monitor and survey the clinical development, but also to cover those drugs that are stuck on its way to the market. Increasing the efficiency in our development process is key, thus, to optimize the pharmaceutical industry investment in this *savings and cuts* era.

This work contributes to this end and is based on Gomez and Lagakos paper[1], which has been recently accepted in the *Statistics in Medicine* journal. This paper describes a methodology based on the Asymptotic Relative Efficiency (ARE) to test efficiencies of clinical trial endpoints in survival analysis, and allow the investigator to choose among single endpoints such as death or progression, or composite endpoints such as progression-free survival, which includes death and progression as we will explain further, provided certain information that shall be obtained from investigators.

We focus on breast cancer data, although the methodology could be broadly applied to survival data. Through a bibliographic research of breast cancer clinical trials papers, that allowed us working with data as realistic as possible, this work runs several ARE calculations and pretends to identify trends and potential guidelines for investigator and pharmaceutical industry on the use of the most efficient clinical trial endpoint in each scenario given certain conditions, and how potentially this may derive in sample size reductions.

We have studied the behavior of the ARE in each case based on real data coming from published papers.

Table of contents

Table of tables

Table of figures

1. The war on cancer

1.1. Background

> "I will also ask for an appropriation of an extra $100 million to launch an intensive campaign to find a cure for cancer, and I will ask later for whatever additional funds can effectively be used. The time has come in America when the same kind of concentrated effort that split the atom and took man to the moon should be turned toward conquering this dreaded disease. Let us make a total national commitment to achieve this goal."
>
> President Richard M. Nixon, January 1971 State of the Union Address [2]

Four decades later, cancer is still the second-leading cause of death in the US, accounting for 25% of all deaths (roughly 556,000). Worldwide, spending for drugs associated with cancer increased a rate of 15% per year over the last 15 years, and today it is estimated to cost approximately $40 billion per year, growing at record rates in the last decade driven in part by the cost of new medical interventions.[3-6]

The authorities spend time, money and resources in comparative effectiveness research assuming that this information relative to costs and benefits will improve the decision making. But the actual "first-world countries" economic situation and the fact that there are currently more than 100 new molecules in phase III oncology trials, makes necessary to streamline the cancer treatment and research spending.

One of the objectives of the scientific community objective is thus, to identify the most efficient treatments for oncology patients guaranteeing the best response to the treatment, and prolonging patient survival.

This objective should count with the support of the pharmaceutical industry, who try to optimize the spending on research, that would impact at the end with the treatment cost.

1.2. Drug development

In clinical trials, the utility of quantifying cancer outcomes is based on their potential to predict patient outcomes. Unfortunately, an active drug is not necessarily an efficacious drug; and the shrinkage or disappearance of the tumoral mass are not sufficient to obtain an increase in the survival or improvement in the quality of life of cancer patients. [7]

The discovery of a new drug starts at the pharmaceutical labs, with the identification of molecules that are active against a disease or pathology, and the posterior development with the in vitro and in vivo tests. But obviously this evidence does not prove the safety or the efficacy of new treatments, so the next step for the selected molecules is to be tested in clinical trials. Small clinical trials first, called phase I and phase II trials, with few patients and highly monitored to avoid any unknown and undesirable patient reaction; and big ones later to ensure the safety and efficacy of the investigational drug, phase III and phase IV clinical trials. A summary of the different clinical trials phases is summarized in Table 1.

This time consuming and expensive development process is meant to prevent disasters like the newborn defects caused by the Thalidomide, a drug introduced in the 1950s to treat morning sickness, and responsible that 10,000 children in 46 countries were born with deformities, as showed in figure 1.

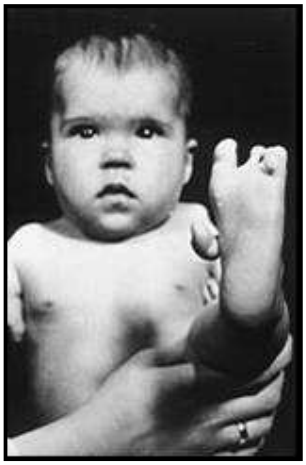


Figure 1 - 1962 born baby to a mother who had taken thalidomide while pregnant

Only if the results mainly from phase I to phase III trials – phase IV trials are usually post-marketing to ensure the product is safe after authorization - are significantly positive, the sponsor can claim the product approval to Health Authorities approval – each country has one regulatory agency evaluating the data -, and then, after years of development the new product is supposedly granted with market access.

This, however, does not guarantee the new drug stays for long in the market, because the product marketing is the "largest trial" it will undergo, reaching a far bigger amount of patients, and is at this time point when the side effects with lower frequency may appear. The ongoing control of the mentioned agencies and sponsors, and the obligation to report any untoward drug reaction to the health authorities, ensure that if any adverse drug reaction of a marketed drug unbalance the risk – benefit ratio, it is removed immediately from the market.

Phase	Purpose	Number of people who take part
Phase I	To find a safe dose To decide how the new treatment should be given To see how the new treatment affects the human body	15-30 people
Phase II	To determine if the new treatment has an effect on a certain cancer To see how the new treatment affects the human body	Less than 100 patients
Phase III	To compare the new treatment (or new use of a treatment) with the current standard treatment	From 100 to several thousand patients
Phase IV	To further assess the long-term safety and effectiveness of a new treatment	Several hundred to several thousand patients

Table 1- Clinical Trial Development Phases

1.3. Clinical trial objectives

One of the reasons of the high cost of the oncology treatments is the cost for the pharmaceutical companies to develop and market a new drug. And choosing the right clinical trial endpoint is the first hurdle the scientists and physicians have to overcome when designing a clinical trial. This question is as relevant as if the molecule they want to test is effective or not, and is around the answer, using what we already know from this molecule – how it behaves in small population groups - and what we expect to see in our target population, that we will define the clinical trials sample size, the trial duration and safety parameters to take into consideration.

There are different types of clinical trial endpoints, based either on objective response rates, quality of life or survival, as we will describe later in section 1.4, but the selected endpoint needs to be consistent throughout the product development period - which may take several years -. If a product completes the phase III trials, this would mean that this new drug is the one between 5.000 or 10.000 that started the early development phases, and that will be used, if the results are favorable, to claim for the indication it has been tested for many years.

It exists an outstanding question that will not be approached in our study, whether investigators should prioritize quality of life indicators rather than efficacy, as to which costs (side effects, long treatments) are some treatments enlarging patients' life. Kozmininski et al [8], analyzed data from 786 surveys sent to US oncologists. In this study, the authors described two different scenarios where the oncologists were supposed to prioritize between survival prolonging or quality of life enhancing criteria to choose the patient's treatment in each scenario. The results of this study were that there was a wide variation across oncologists when they have to choose between survival and quality of life.

Among all clinical trial endpoints, the study we are presenting here has the objective to help the scientific community in the process of selection of survival endpoints. This study pretends to be a statistical tool to help in the early phases of the drug development, which would optimize the number of patients that needs to be dosed in a clinical trial, and that gives us guarantees that is the best one among all different survival options.

1.4. Clinical trial endpoints in oncology

As mentioned before, the discussion on the selection of clinical trial endpoints is linked to the trial design, in the early phases of its development. Like any other disease, cancer can be approached from different points of view, offering different types of clinical trial endpoints. The endpoints discussed here are distributed in three groups, as described by the FDA Guidance for Industry of May 2007[9] , and include Overall Survival, Endpoints based on Tumor Assessment, and Endpoints Involving Symptom Assessment.

A. **Overall Survival (OS)**: defined as the time from patient randomization (allocation to a one treatment arm) until death from any cause. Is considered as the most reliable cancer endpoint, and is the preferred to assess patient survival because it is precise and easy to measure. Demonstration of statistically significant improvement in overall survival can be considered to be clinically significant if the toxicity profile is acceptable, and has often supported a new drug approval.

B. Endpoints based on Tumor Assessments: The collection and analysis of data on these time-dependent endpoints are based on indirect assessments, calculations and estimates.

1. **Disease-Free Survival (DFS):** time from randomization until recurrence of tumor or death from any cause. Frequently used in the adjuvant setting after definitive surgery or radiotherapy.
2. **Objective Response Rate (ORR):** proportion of patients with tumor size reduction of a predefined amount or for a minimum time period.
3. **Progression Free Survival (PFS):** time from randomization until objective tumor progression or death from any cause, whichever comes first.

   We may assume that PFS and DFS only differ on the disease phase, and therefore they are somehow equivalent. In early cancer stages, neo-adjuvant and/or surgery interventions may completely remove the tumoral cells, and medical doctor may consider the patient as "clean" or "disease-free", whilst in advanced patients treatment may help only reducing the tumoral mass or slow the tumor growth, and therefore the endpoint tries as objectively as possible to define time until disease progression, i.e. worsening of cancer.

4. **Time to Progression (TTP):** time from randomization until objective tumor progression. TTP does not include deaths. In TTP analysis, deaths are censored, either at the time of death or at an earlier visit representing informative censoring (non-random pattern of loss from the study).
5. **Time to treatment failure (TTF):** time from randomization to discontinuation of treatment for any reason, including disease progression, treatment toxicity and death. TTF is not recommended by the FDA for drug approval, because does not distinguish the efficacy of the drug from toxicity.

C. **Endpoints involving Symptom Assessment and Biomarkers:** These two last groups do not serve as primary efficacy endpoints in oncology drug approvals. Symptom improvement can simply indicate less toxicity, and further research for biomarkers needs to be done to establish its validity.

The United States Food and Drug Administration (FDA) recommends meeting his specialists before submitting the clinical trial protocols design, to confirm the Agency accepts the design and the endpoint chosen to show improvement in the selected disease.

1.5. Choosing the best clinical trial endpoint

In the early 1980s, the FDA approved oncology drugs based on tumor Response Rate alone. In the mid-1980s, on the advice of the Oncologic Drugs Advisory Committee (ODAC), the FDA determined that response rate generally should not be the sole basis for approval. The potential benefit associated with a partial response did not necessarily outweigh the substantial toxicity of oncology drugs, and the correlation between response rate and survival or clinical benefit was not well established. The new FDA position called for an improvement in survival or patient symptoms for regular approval [10]. Today, and given its objectivity and the unquestionable benefit derived by patients, overall survival has been the most important endpoint in medical oncology and has been broadly recommended from health authorities and a survival cornerstone.

However, overall survival has been shown to be an elusive endpoint. Although objective and simple to measure, it has the disadvantage of requiring long follow-up periods, increasing the costs and duration of these trials; may be confounded by causes of mortality unrelated to cancer; and the effects of a drug in overall survival may be diluted by the effects of crossover and subsequent therapies. And still one argument on patient wellness: prolonging the patient survival does not translate into an improvement on the quality of life[11].

Because of the number of new promising treatments arriving in phase III trials for advanced breast cancer, there is an urgent need to identify endpoints that are more sensitive and more rapidly observed than OS. In recent years there has been an increase in the use and regulatory acceptance of PFS as primary endpoint in cancer clinical trials: as explained already, PFS looks forward to either death - as Overall Survival does - or disease progression, whichever comes first. A key advantage of the progression-free survival end-point is that progression occurs months or years before death, and therefore the time required to get the number of events statistically required is shorter for PFS than for OS. However, there are several points that have been criticized concerning the use of PFS as primary endpoint in oncology trials:

PFS is not an entirely objective end-point – definition of progression may vary among trials -, and thus can be subject to bias. This is partially resolved blinding the patient and/or the investigators, but can still be a concern if there are differences in the toxicity of different arms, which may partially unblind the trial. Another concern is that it is complicate to demonstrate

that effect on PFS does correlate in effect to OS[12]; and finally, informative censoring: those patients that at the cut-off time for analysis are still alive or did not showed progression, shall be balanced in both groups.

This type of endpoints - such as PFS or TTF - are known as composite endpoints (CE), and are defined as the occurrence of any event among a given set of events after a certain follow up time. A patient that experiences any of the CE events (components) will be considered to have experienced the event of interest. In the PFS case, it means that a patient that experiences either disease relapse or death is counted as if he has experienced the event. In other words, we are adding disease progression to the overall survival.

1.6. Composite endpoints

CE are defined thus as the occurrence of any event from among a set events after a certain time of follow-up. This allows adopting several endpoints as a single one.

Its use was approved and described in the International Conference for Harmonization (ICH) of Technical requirements for Registration of Pharmaceuticals for Human Use – Efficacy Guidelines part 9 – , finalized on February 1998[13] as follows:

> "If a single primary variable cannot be selected from multiple measurements associated with the primary objective, another useful strategy is to integrate or combine the multiple measurements into a single or 'composite' variable, using a pre-defined algorithm (...). This approach addresses the multiplicity problem without requiring adjustment to the type I error."

It is also recommended in these guidelines that its use shall be transparent. The construct of this composite endpoint shall be pre-specified in the protocol; the sample size shall be calculated taking into account the overall number of events of all endpoint components, and any statistical significant result is attributed to the CE, and not to any of its component, unless it is clinically meaningful and has been validated. This prevents of investigators constructing the composite endpoints based on ongoing trial results.

At first glance it may seem that using composite endpoints would make easier the investigator's job, as man can add any endpoint of interest without the need of adjusting for multiplicity, but this may be not always the case. The main purpose of composite endpoints is to include as many clinically relevant endpoints as possible in the efficacy assessment of a treatment without having to increase the sample size to an unacceptable level[14].

Ferreira-González et al[15], performed a systematic review of the literature looking for controlled clinical trials were investigators considered using composite endpoints. Among the selected trials, they summarized the rational, the interpretation, the advantages, limitations and recommendations. The main advantage they came up with was that investigators used composite endpoints because it increases the statistical power when assessing intervention differences, and therefore allows a lower sample size. This may be translated in money savings, reduction of follow-up time, and more ethically accepted, understanding that less

number of patients will be recruited on a clinical trial. Some authors highlighted, however, that if treatment effect is not similar across all of the components, and as consequence, the global risk reduction for the composite is reduced, the required sample size may increase.

Freemantle et al[16] put the example on his paper of the CAPRICORN trial[17], trial investigated the effects of carvedilol, a β-blocker, in 1959 patients with left ventricular dysfunction following myocardial infarction. The primary endpoint stated in the protocol was all-cause death, but while the study was ongoing, the data and safety monitoring board, which has the key role of protecting the interest of patients[18], communicated to the investigator's steering committee (SC) that the overall rate of accrual of death was lower than predicted. This lower number of death was threatening the statistical power to detect differences. The SC took the decision of modifying the primary endpoint and turned it into a composite endpoint (all-cause mortality or cardiovascular hospital admissions). The original primary endpoint achieved a p-value of 0.03, whereas the alternative primary endpoint had a p-value of 0.30. Thus, the trial did not met the primary endpoint so could not demonstrate the benefits of carvedilol in those patients. This example demonstrates the consequences of using a composite endpoint, where one component may dilute the measure of treatment effect of other components that may show evidence of efficacy, and highlights the importance of carefully choosing the components of CE.

On the other hand, back to Ferreira-González's paper, the most widely considered disadvantage of composite endpoints is the fact that all components are considered as equivalent in the analysis, irrespective of its relevance for patients. It could be particularly misleading if an overall effect is driven by a less important component of the endpoint. A substantive risk associated with the reporting of composite outcomes is that the benefits described may be presumed to relate to all of the components. Likewise, the results interpretation may be problematic if the magnitude of the effect is very dissimilar on each of the components, reaching its maximum if any of the components shows opposite directions.

We could easily conclude that by adding new clinical outcomes it may increase the probability to detect these events, and reduce time, costs and resources. Unfortunately this assumption is not correct. Before adding such an added endpoint, some questions should be addressed:

- Is this new event clinically relevant? In other words, does the addition of this extra event provide an added value?

- May this new event be as frequent in the control group as in the group we are investigating, which may mask our treatment effect?
- Are we able to develop a statistical method that may help us taking this decision

We will try with this paper to answer the last question.

1.7. Breast cancer

It is not realistic today treating all cancers as a single and same disease, because we know there are many types of cancer:

- **Carcinoma** - cancer that begins in the skin or in tissues that line or cover internal organs.
- **Sarcoma** - cancer that begins in bone, cartilage, fat, muscle, blood vessels, or other connective or supportive tissue.
- **Leukemia** - cancer that starts in blood-forming tissue such as the bone marrow and causes large numbers of abnormal blood cells to be produced and enter the blood.
- **Lymphoma and myeloma** - cancers that begin in the cells of the immune system.
- **Central nervous system cancers** - cancers that begin in the tissues of the brain and spinal cord

To reduce any potential source of bias, and with data as comparable as possible, this study will focus on one single type of cancer disease. Among all different types, and seen the differences on incidence, detection methods, number of treatment lines and prognostic, this study focused on **breast cancer** patients because of the following figures. According to the last information available (June 2011):

- There are 370,000 new cases per year diagnosed in Europe.
- The incidence in North European countries is 82.5 /100,000 inhabitants / year; while in Sothern European countries the rate is 62.4 / 100,000 inhabitants / year.
- In Spain, there are about 22,000 new cases per year which represents almost a 30% of all cancers.
- In Catalunya there are 83.9 cases /100,000 inhabitants, while the national mean is 50.9 cases / 100,000 inhabitants. (last update from June 2011)
- And with regards to mortality, in 2004 the number of breast cancer deaths in Catalunya was 1,029, while at a country in Spain, data from 2009, there were 6,129 deaths [19].

Beyond Europe, breast cancer is an also an actual concern in the United States:

- In 2011, an estimated of 230,480 (female) and 2,140 (male) new cases of invasive breast cancer were expected to be diagnosed in women in the US, along with 57,650 new cases of non-invasive (in situ) breast cancer.
- These will cause 39,520 female and 450 male deaths.
- About 12% of US Women will develop invasive breast cancer over the course of her lifetime.
- Breast cancer is, together with lung cancer, the two diseases that kill more women than any other disease in the US[20] .

Unfortunately, there has not been any revolution neither in the diagnose or in the treatment of breast cancer, which may mean that these figures will be similar ones to what registries will show from 2011 and will be in 2012.

It's worthwhile then to look into this disease and see how statistics can we contribute to optimize the clinical trial endpoint in new therapies being investigated.

1.8. Progression free survival as surrogate for overall survival

As discussed in section 1.6, how to validate the "true" endpoints has caused considerable controversy in the past decades. There are many discussions and studies trying to defend the use of PFS as a surrogate endpoint of OS, but showing correlation between both endpoints does not make a surrogate. PFS needs validation and must reasonably likely to predict clinical benefit.

Several authors tried to show with examples how OS was not the most appropriate endpoint[21-22-23] in different indications such as colorectal or breast cancer, highlighting clinical trials' results that did not show statistical differences on the OS when this was the primary endpoint, but when this data was analyzed through meta-analysis, some of these trials showed that there were statistical differences on other endpoints such as progression or objective response rate. Although the surrogacy of OS has not been firmly established in breast cancer, PFS has been chosen as the primary endpoint in the majority of recent trials as the basis for approval as we will see as we advance in our study.

From 1990 until 2002 there was a single FDA approval relying primarily on Time to Progression or Progression Free Survival without also relying on a survival benefit. But, over the past seven years the FDA has approved at least nineteen drug applications that were primarily based on a PFS endpoint [24].

If we look into the PFS or DFS definition, we see that we are able to split these outcomes into two different ones. For instance, with PFS we are creating a single endpoint made of disease progression and death. This means that we are observing those patients that may have been considered as well if we would take care only of deaths (OS), but we are adding here progression, and thus patients that may be alive at the end of the trial, but that showed disease progression during the course of the trial. Therefore, we increase the probability to see events in our clinical trials.

Graphically, to describe all potential scenarios one can find when trying to identify surrogate endpoints to OS, we summarized them all on Figure 1:

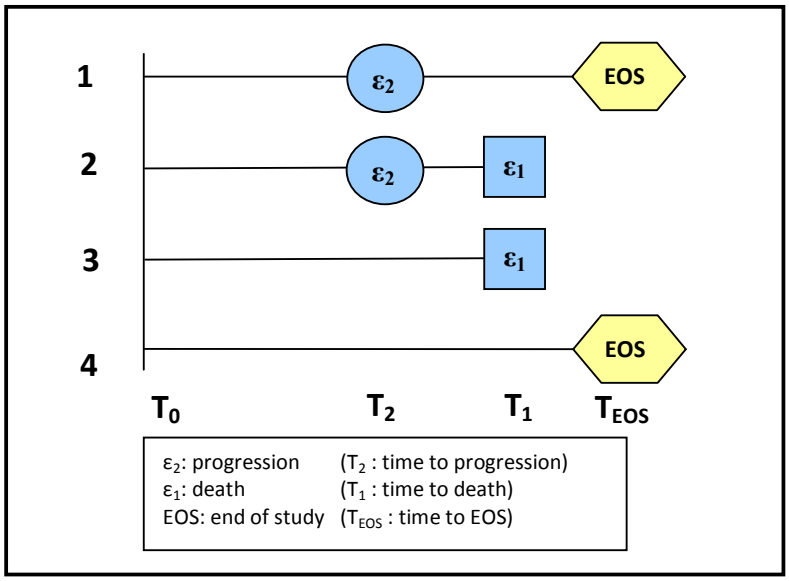


Figure 2: survival scenarios

Since FDA guidelines recommend the use of overall survival as a survival endpoint, and to ease future explanations, we have named death as *E1* (event 1) and will be our event of interest; and *E2* the alternative event (ie. progression, treatment failure…). T_1 is the time from randomization - time when we allocate the patient to one of the treatment arms - to reach our event of interest (death); and T_2 the time to reach the alternate endpoint. Logically, If both *E1* and *E2* are observed, T_2 will always be inferior to T_1, because "death prevents us to see the progression" (T2 < T1). In other words, *E1* is a terminating event.

According to this, we may be able to observe 4 different scenarios:

- **Patient #1**: a patient randomized at T_0 (allocated to one treatment arm), and that experienced *E2* only (progression), but not death at the time of the end of the study. From this patient we are able to get T_2 only.
- **Patient #2**: patient randomized at T_0 that presents *E2* (progression), and *E1* afterwards, so we can get T_2 and T_1 information.
- **Patient #3:** patient that shows *E1* only, so we will only get T_1.
- **Patient #4**: is a patient that after the treatment and throughout the trial does not show either *E2* or *E1* (i.e., the patient was disease-free and alive during the time of the trial)

As described in section 1.6, a key advantage of a composite endpoint is that it allows investigators who are having difficulty deciding on a single endpoint combining two or more outcomes that describe a disease process. In our case, we consider Death as the main event – we will call it *relevant* endpoint – and the one we expect to observe in our clinical trial. And we introduce Disease Progression as an additional endpoint, which we think it may reduce the time of patient follow up, and thus shorten our timelines.

As said above, progression may be observed much before death. We will call this composite endpoint *E**, and will be defined as:

$$E^* = E1 \cup E2$$

And the time (T*) to this composite endpoint will be the minimum between progression and death.

$$T^* = min\{T_2, T_1\};$$

Note that, as we deduced from Figure 2: **survival scenarios**, *E** will only be observed if any of the events happens before the EOS, ie:

$$T^* < EOS$$

By adding a new endpoint we are increasing the probability of detecting an event in our patients. If we consider α as the probability of finding each patient group in our population, we can state that:

$$P1 = \alpha_2 + \alpha_3$$

$$P2 = \alpha_1 + \alpha_2$$

$$P_* = \alpha_1 + \alpha_2 + \alpha_3;$$

where *P1* is the probability of observing death, *P2* probability of observing progression and P* probability of observing the composite endpoint.

2. Objective of this master's thesis

As described in the introduction, overall survival (OS) is the authorities' preferred clinical trial endpoint in oncology trials, given its objectivity and that it is easy to measure. However, it may require in some cases long time to get the required number of events to perform the statistical analysis of OS, which impacts directly to the final cost of the product. Furthermore, increasing patient's life is not the only objective of the scientific community, but also increasing patient's quality of their remaining life, i.e. for instance reducing or delaying the patient's progression rate. By measuring only the overall patient's survival it is not possible to have any clue of patient's quality of live during the clinical trial. Time to progression may provide us broader information on patient's status during the trial period.

The current situation is thus, that the investigators' community and the economic situation require the acceptance of other clinical endpoints which would be able to shorten the clinical trial duration, thus clinical development costs, and create endpoints that may reflect patient's quality of live. These might be the so called composite endpoints.

Since the objective of composite endpoints is to increase the efficiency of the clinical trials, the purpose of this study will be to evaluate in which cases the use of single endpoints (overall survival) will be more statistically efficient to detect differences in clinical trials, and in which ones the addition of a second endpoint (progression), building a composite endpoint, will be beneficial to show differences between the treatment groups.

This study is based on breast cancer clinical trials, for the reasons given at section 1.7, and the information has been obtained from a bibliographic review of breast cancer clinical trials published between 2009 and 2010 as will be described in section 4. It was limited to those two years to try to obtain data as homogenous as possible, and to limit the bias that might have caused changes on the standard of care of breast cancer disease.

Among other composite endpoints, we will analyze those that are based on survival outcomes. Namely, we extracted information from trials that used single endpoints such as Overall Survival or Time to Progression, and information from trials that used composite endpoints such as Disease-free survival or Progression-free survival.

The statistical methodology that will be used to analyze this data is based on the asymptotic relative efficiency (ARE), as defined in *Gomez and Lagakos*[1] paper, that will be better described on the following sections. Through the calculation of the ARE, we pretend to be able to decide, having defined a baseline conditions of what an investigator expects to observe in terms of hazard ratio, probability of observing and event, and some statistical parameters that will be later introduced, when the use of OS would be more efficient to detect statistical differences between two treatment arms, and when, by adding the detection of progression we increase the efficiency of clinical trials.

A subsequent objective of this study would be, and due to the fact that the same information than the one needed to calculate the ARE, is used to calculate the clinical trial sample size, we may be able to use the information obtained to calculate the clinical trial sample size with the less number of patients treated as possible maintaining the desired power (1-β) and level of significance (α).

In summary, increasing the efficiency of clinical trials designs is not only a benefit from an economic point of view because it optimizes the trial duration; but also from an ethical point of view, we may be able to statistically reduce the number of patients to be recruited in a clinical trial, and therefore expose less patients to investigational research.

3. Methods

3.1. Statistical methodology

The methodology described in this thesis is based on the paper "*Statistical considerations when using a composite endpoint for comparing treatment groups*" developed by Guadalupe Gómez and Stephen Lagakos [1].

The above paper developed a statistical methodology for deciding whether to expand a primary endpoint (*E1*) to the composite (*E*=E1 U E2*), by means of the Asymptotic Relative Efficiency of two logrank tests.

The mentioned paper describes 4 different scenarios, depending on the type of endpoints selected for comparison:

- **Case 1:** none of the endpoints include a terminating event. In this case both *E1* and *E2* are observed if they occur before the end of the study (EOS), i.e.: *T1 < EOS* and *T2 < EOS*.
- **Case 2:** the endpoint of interest (*E1*) does not include a terminating endpoint but the additional (*E2*) does. In such case *E1* can only be observed if it occurs before the additional endpoint *E2, T1 < min {T2, C}* while *T2 < EOS*.
- **Case 3 (the case we will consider):** the endpoint of interest is a terminating endpoint and the additional one does not. In this case *E1* is observed while it happens within the study time (T*1* < EOS), while *E2* can only be observed if it occurs before the endpoint of interest *E1* or the end of the study, $T_2 < min\{T1, C\}$.
- **Case 4:** Both events include a terminating event, meaning that each event would be observed if *T1 < min{T2, C}* and $T_2 < min\{T1, C\}$.

These scenarios can be graphically described as showed in Figure 3.

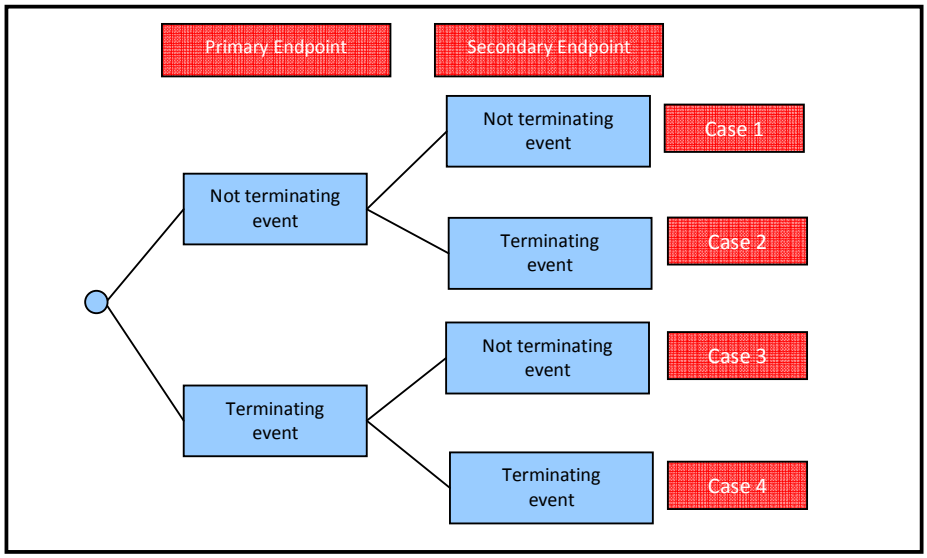


Figure 3 - Graphical distribution of case scenarios described in *Gomez and Lagakos* paper

In our case, seen that overall survival is the Health Authorities' preferred endpoint, we will set *death* as our relevant endpoint (*E1*). It is a terminating endpoint which prevents us to observe any other posterior endpoint, so we will base our study in the **case 3 scenario**: death will be our event of interest, and we will explore if by adding progression and creating a composite endpoint the results improve.

Therefore *Progression* will be our additional endpoint (*E2*), which is a non-terminating event that may allow us to observe death after it.

1. Since *E1* is a terminating event, *T1* will be observed as long as it occurs before the EOS;

$$T1 < EOS;$$

2. And *T2* will only be observed if occurs before *T1* and EOS

$$T2 < min\ \{T1, EOS\}$$

3. The observed endpoint will be denoted $U = min\ \{T1, EOS\ \}$, while for the composite endpoint, T^* will be observed if it occurs before the EOS, i.e. $T^* < EOS$ and we will name the outcome observed as $U^* = min\ \{T^*, EOS\}$

Apart from this notation, this procedure has as well some assumptions, based on the experience of previous clinical trials. Efficiency calculations will be evaluated through a sequence of contiguous alternative to the null hypothesis, and based on the following assumptions:

I. **Assumption 1:** End-of-study censoring at time τ (without loss of generality we take τ = 1) is the only non-informative censoring cause, that is,

$$Pr\{EOS > t\} = 1\{[0, \tau]\}(t) = 1\{[0; 1]\}(t)$$

Declaring that EOS is the only non-informative censoring, we are stating that a patient that completes the clinical trial follow-up period and is still alive, has the same risk of death than a patient that is still ongoing, and we do not have any added information on this patient future survival. Hence, any other censored data: i.e.: discontinuation from clinical trial due to toxicity, will be considered informative.

It is important to highlight that these assumptions are not only important and significant from a statistical point of view.

Non-informative censoring prevents us from getting information from the recruited patients. When the sample size calculation is performed at the set up phases of a clinical trial, most often what it is calculated is the number of events needed to show statistical differences between treatment arms. From this number of events, investigators assume a number of patients from which they will not be able to observe the event, and therefore the real sample size that will be recruited.

In clinical trials, sometimes we have the risk to have patients who are lost to follow-up, especially in those trials where the follow-up period is long (may be of several years). If a patient changes his/her residence during the trial duration and is not able to attend the appointments scheduled along the trial duration, this prevents us to get the information from the patient, and are forced to censor the patient unless the reason for discontinuation is related to the trial endpoints, this censoring is non-informative. If there is no informative censoring other than the EOS, we may be at risk of having underestimated the number of patients needed, and therefore being unpowered to make any statistical assessment.

We will discuss further on this issue in the limitations section of this study.

II. **Assumption 2:** End-of-study censoring is identical across groups, that is, $Pr\{EOS > t \mid X = 0\} = Pr\{EOS > t \mid X = 1\} = Pr\{EOS > t\} = 1\{[0, \tau]\}(t)$. This assumption facilitates computations and derivations although the general expressions could be analogously stated without it.

Statistical considerations aside, and since the purpose of the clinical trials in our study is to compare survival data, at the eyes of the authorities it would not be acceptable to have different follow up times depending on the treatment arm a patient is allocated to. Nowadays some clinical trials are blinded, and those which are not define very well in the protocol that the follow up time is independent of the treatment assigned to the patient. To provide evidence of improvement, both treatment groups have the same follow up time. All clinical trials considered in this study are randomized clinical trials, so what has been observed throughout the study is that most of the clinical trials designs have the same trial duration, and the treatment is randomly assigned to a patient. This means that is at the baseline visit (time of randomization), the patient is allocated to a treatment arm, but the trial duration or the follow up is independent to the treatment arm the patient has been assigned.

III. **Assumption 3:** Treatment groups have proportional hazards. The risk of an event in one group relative to the other does not change with time. This is a key assumption and states that the proportionality assumption is given by the hazard ratios: $HR_1 = \frac{\lambda_1^{(1)}(t)}{\lambda_1^{(0)}(t)_1}$, which is the hazard ratio of the main endpoint (death) in both treatment arms; and $HR_2 = \frac{\lambda_2^{(1)}(t)}{\lambda_2^{(0)}(t)_2}$: hazard ration of the additional endpoint, progression in our case, for all *t*.

It is important to mention that all trials where the method to calculate the hazard ratios was disclosed, it was always assumed the proportionality of the hazards.

3.2. Logrank test

In survival analysis, we often want to compare different treatment effects on two (or more) patient groups. Experiments in which observations are the time to occurrence of an event are common source of right-censored data. In these cases, what we would like to know is the incidence of the event over time[25].

Considering that in our trials patients have been allocated to one of two treatment arms, when we compare survival data we establish:

$$H_0 : S_1(.) = S_2(.) \text{ , or similarly } H_0 : \lambda_1(.) = \lambda_2(.) ,$$

where S_1 and S_2 are the survival functions of *T1* for group *1* and *2*, and λ_1 and λ_2 are the hazard functions for each group.

The most widely used method to compare two survival curves is the Logrank test, a nonparametric test appropriate to use when the data is censored. The null hypothesis of the logrank test is that there's no difference on between the populations in the probability of an event at any time point[26]. Under this null hypothesis, the logrank test is asymptotically *N(0,1)*.

To calculate the logrank test we need to sort chronologically the combined events observed from two treatment groups, as if we had only one group. At each time point where we have an event, say one patient that progressed, we calculate the number of patients that are still alive in each group and calculate the probability that the observed event belongs to that treatment group.

It is important to highlight that patients who are censored are not considered as an event in this method; it only contributes with a reduction of patients at risk at the time of the following event. This information is collected during all trial duration, so that at the end of the trial we will know the total number of patients that we expected from each treatment arm for which we expected to observe an event.

The data can be summarized in a 2x2 table. In a two arms trial, where we want to observe the event *d* (death for example), we could summarize our data at a given time point (t_i) as:

Arm	Events	Survivors	Total at risk
1	d_{i1}	R_{i1} - d_{i1}	R_{i1}
2	d_{i2}	R_{i2} - d_{i2}	R_{i2}
Total	d_{ij}	R_i - d_i	R_i

where d_{ij} is the number of deaths from group *j* at a time t_i, treatments are T= j (1,2); R_i is the total number of subjects at risk, and we can take $\frac{d_{i1}}{R_{i1}}$ as an estimator of $\lambda_1(t_i)$.

Under the null hypothesis, the hazard functions are equal in both treatment groups over time, the estimator of the probability of experiencing an event p_{ij} could be given by $R_{ij}\frac{d_i}{R_i}$, j=1, 2.

We define as well its Expectation and Variance under H_0 as:

$$\mathrm{E}(\mathrm{d_{ij}} \mid \mathrm{R_i}, \mathrm{R_{ij}}, \mathrm{d_i}) = \frac{R_{ij} d_i}{R_i}$$

$$Var(\mathrm{d_{ij}} \mid \mathrm{R_i}, \mathrm{R_{ij}}, \mathrm{d_i}) = \frac{R_{i1} R_{i2} (R_i - d_i) d_i}{(R_i)^2 (R_i - 1)},$$

The statistic defined as:

$$\frac{d_i - \frac{R_{ij} d_i}{R_i}}{\sqrt{\frac{R_{i1} R_{i2} (R_i - d_i) d_i}{(R_i)^2 (R_i - 1)}}}$$

is asymptotically *N(0,1)*, and equivalently:

$$LR_i = \frac{(d_i - \frac{R_{ij} d_i}{R_i})^2}{\frac{R_{i1} R_{i2} (R_i - d_i) d_i}{(R_i)^2 (R_i - 1)}},$$

holds a χ^2 with one degree of freedom and rejects the null hypothesis for extreme values of d_{i1}.

Generally, in our survival curves, we will be interested to test that this null hypothesis holds at each time point considered *i = (1,…D)* so we define the **logrank test** as:

$$LR = \frac{\sum_{i=1}^{D}(d_i - \frac{R_{ij}d_i}{R_i})^2}{Var\left(\sum_{i=1}^{D}(d_i - \frac{R_{ij}d_i}{R_i})\right)}$$

In other words, the usual expression of observed minus expected holds for the logrank test, that is:

$$LR = \frac{\sum_{i=1}^{D}(O_i - E_i)}{\sqrt{\sum_{i=1}^{D}V_i}}$$

where O_i is the number of events observed at time t_i, and E_i is the expected number of events to be observed at time t_i.

Under the H_0 of no treatment differences, the logrank test is asymptotically N(0,1), but when the H_0 does not hold, and if we consider a sequence of contiguous alternatives to H_0, for a finite sample size (n), and considering λ_1 as fixed and λ_2 vary with *n*, the log rank Z results to be asymptotically normal with unit variance and mean μ.

This centrality parameter (μ) is function of $p(i) = \Pr_{H0}\{X = 1 \mid U \geq i\}$; the null probability that someone at risk at time t_i, is in treatment 1, $\Pr_{H0}\{U \geq i\}$ is the null probability that someone is still at risk at time t_i, and $\Pr_{H0}\{U \geq t\}\lambda_1(i)$ is the probability under H_0 of observing *E1* by time *t*.

Likewise, and as introduced above, we called T_2 the time to the secondary endpoint *E2*, and $T_* = min\{T1,T2\}$ to the composite endpoint. Likewise, to test the null hypothesis of no treatment difference based on the CE: $H_0^* : \lambda_*^{(0)}(.) = \lambda_*^{(1)}(.)$, we can use the logrank test. Under the H_0, Z_* is asymptotically N(0,1), and when the null hypothesis does not hold, Z_* is asymptotically normal with unit variance and mean μ^*, that is function of $p_*(i) = \Pr_{H0}\{X = 1 \mid U_* \geq i\}$: the null probability that someone at risk at time t_i, is in treatment 1, and $\Pr_{H_0^*}\{U_* \geq t\}$: the probability under H_0 that someone is still at risk at time t_i.

Function *survdiff* from the library *survival* in R[27], runs through a given file where it has been previously indicated the treatment arm, the times where the events occurred, and the existence of censoring. The parameter *rho = 0* corresponds to the logrank method, and the output of this method provides information on the statistical differences between survival curves.

3.3. Asymptotic relative efficiency

One way to compare two statistical tests is to compare their efficiencies. In our study, we will check the differences between using the test Z, based on *E1* (relevant endpoint), and using the test Z_*, based on the composite endpoint *E**.

As we have seen in the previous section, both tests under the null hypothesis are asymptotically N(0,1), and under a sequence of contiguous alternatives to the null hypothesis they are asymptotically normal with variance 1 and mean as per the conditions introduced in section 3.2. Hence, the ARE (Asymptotic Relative Efficiency) is given by:

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

To turn this expression and the procedure behind into something useful for the design of clinical trials and interpretable for investigators, *Gómez and Lagakos* derived this expression to a much interpretable one based on some measurable parameters such as:

- Frequencies *P1* and *P2* of observing endpoints *E1* and *E2* in treatment 0 (control);
- The relative treatment effect on T_1 and T_2 given by *HR1* and *HR2* respectively;
- The degree of dependence between *T1* and *T2* given by the Spearman's rank correlation coefficient ρ.

To come to that simplification the *Weibull distribution* has been chosen to represent the survival distributions, as it is widely used due to its flexibility, which allows hazard functions to increase, decrease, or remain constant, and mimic the behavior normal or exponential distributions. Taking into consideration assumptions described in section 3.1, the mentioned paper reaches the following expression:

$$ARE(Z_*, Z) = \left(\frac{\mu_*}{\mu}\right)^2 = \frac{\left(\int_0^1 \log\{HR_*(u)\} f_*^{(0)}(t)du\right)}{(\log HR_1)^2 \left(\int_0^1 f_*^{(0)}(t)di\right)\left(\int_0^1 f_1^{(0)}(t)du\right)} \qquad (1)$$

From expression (1) we note that, to calculate the ARE (Z*,Z), we need the following information:

- $f_1^{(0)}(t)$: the marginal density function for *T1* in group 0;
- $f_*^{(0)}(t)$: the marginal density function for T_* in group 0;
- $HR_1 = \frac{\lambda_1^{(1)}(t)}{\lambda_1^{(0)}(t)}$: the hazard ratio of the relevant endpoint;
- $HR_* = \frac{\lambda_*^{(1)}(t)}{\lambda_*^{(0)}(t)}$: the hazard ration of the composite endpoint

But this expression does not seem yet as comprehensive as an investigator would expect, so it needs to be further developed and work out relations with parameters that could be obtained.

As introduced in this section, the Weibull distribution has been chosen for the marginal laws of T_1 and T_2. For both treatment arms (*j=1,2*) the survival function is given by:

$$S_k^{(j)}(t) = \exp\left\{-\left(\frac{t}{b_k^{(j)}}\right)^{\beta_k^{(j)}}\right\}$$ (k=1,2 depending if we refer to *T1* or *T2*)

where $b_k^{(j)}$ is the scale parameter and $\beta_k^{(j)}$ is the shape parameter. Again, as stated in assumption 3, we assume to have proportional hazards, meaning that both groups will have the same shape parameters, namely $\beta_1^{(0)} = \beta_1^{(1)} = \beta_1$ and $\beta_2^{(0)} = \beta_2^{(1)} = \beta_2$.

Back to the information needed to compute the ARE, we are able now to say that both the density function and the survival function, $f_1^{(0)}(u) \Leftrightarrow S_1^{(0)}(u)$, can be described by the parameters $(b_1^{(0)}, \beta_1)$.

Regarding $f_*^{(0)}(t)$, that describes the behavior of the composite endpoint in the T_* in the control group, we know that $T_*^{(0)} = \min\{T_1^{(0)}, T_2^{(0)}\}$ so that it requires information about $S_1^{(0)}(t)$ and $S_2^{(0)}(t)$. Similar to what we have done for $S_1^{(0)}(t)$, we can say that $S_2^{(0)}(t)$ can be expressed in terms of a Weibull function with parameters $(b_2^{(0)}, \beta_2^{(0)})$.

With the information gathered so far we would be able to know the information for the first three bullet point we need to calculate the ARE: $f_1^{(0)}(t)$, $f_*^{(0)}(u)$ and $HR_1 = \frac{\lambda_1^{(1)}(t)}{\lambda_1^{(0)}(t)}$.

But what about $HR_* = \frac{\lambda_*^{(1)}(t)}{\lambda_*^{(0)}(t)}$? This expression implies the knowledge of $\lambda_*^{(1)}(t) \Leftrightarrow S_*^{(1)}(t)$ which involves $S_1^{(1)}(t) = (b_1^{(1)}, \beta_1)$ and $S_2^{(1)}(t) = (b_2^{(1)}, \beta_2)$.

As it has been already said, and as per the proportional hazards assumption, we will fix β_1 and β_2, so the information we still need to get are the scale parameters: $b_1^{(0)}, b_2^{(0)}, b_1^{(1)}$ and $b_2^{(1)}$. As developed in *Gomez and Lagakos* paper, the scale parameters $b_1^{(0)}$ and $b_2^{(0)}$ are functions of (*P1*, β_1) and (*P2*, β_2) respectively, while the scale parameters $b_1^{(1)}$ and $b_2^{(1)}$ are functions of ($b_1^{(0)}$,*HR1*) and ($b_2^{(0)}$,*HR2*). Hence, the parameters needed to calculate the ARE can be summarised by:

- β_1 : shape of the marginal law of the relevant endpoint;
- β_2 : shape of the marginal law of the additional endpoint;
- *HR1* : hazard ratio of the relevant endpoint in the control group;
- *HR2* : hazard ratio of the additional endpoint in the control group;
- *P1*: probability of observing the main endpoint in the control group;
- *P2*: probability of observing the additional endpoint in the control group;
- ρ: Spearman's correlation between the relevant and the additional endpoint.

We will evaluate the results of the $ARE(Z_*, Z)$, and decide whether it is more beneficial to use overall survival or progression-free survival.

- If $ARE(Z_*, Z) > 1$, the use of composite endpoints is recommended, meaning that the addition of progression (additional endpoint) to death provides an added value and increases the clinical trial efficiency.
- If $ARE(Z_*, Z) \leq 1$, then the relevant endpoint (death in our study) shall be considered as the primary endpoint alone.

It must be said as well that in the boundaries of 1, the benefit of one endpoint over the over may be slight, so other factors rather than statistical shall be considered.

3.4. Statistical tools and programming

All ARE calculations, charts and tables have been generated using R (version 2.11.1) programming. See code attached in

Appendix 4 – R Programming Code Used. Original ARE calculation programming was performed in MAPLE (v12) language and adapted to R by Moisés Gómez Mateu (Universitat Politècnica de Catalunya).

Computations are based upon Gómez and Lagakos paper following the steps below:

- A range of Spearman's rank correlation values (ρ) are introduced and for each one the corresponding Frank's copula association parameter (θ) is calculated.

- Sets of values for (β_1, β_2, *P1*, *P2*, *HR1*, *HR2*) are provided. For each possible combination the ARE is computed following the method described in section 3.3.

- The resulting values of the ARE for each combination of the above mentioned parameters are written in a database and used for the analysis of the convenience of choosing the composite endpoint versus the primary endpoint.

- As mentioned in the previous section, the Weibull distribution has been chosen as it is widely used in survival analysis due to its flexibility. It allows for decreasing, constant and increasing hazard rates and mimics the behavior of other, e.g., normal and exponential, statistical distributions. Weibull distributions are define with two parameters:
    - the scale parameter: $b_{(j)}$
    - the shape parameter: $\beta_{(j)}$

All parameters are combined with each other in order to set up all kind of scenarios. For each scenario the ARE value is calculated, and output saved in a separate database.

4. Bibliographic research

4.1. Search algorithm

As already mentioned, to be able to compile enough data to perform our study, a bibliographic research has been performed in some of most popular medical journals, with higher impact factor:

Journal	Impact Factor in 2010
New England Journal of Medicine (NEJM)	53,480
Journal of Clinical Oncology (JCO)	18,970
British Medical Journal (BMJ)	13,471
Annals of oncology (AoO)	6,452

The first search algorithm was to look in each of these journals the key words **breast** and **cancer** and **randomized** in title or in abstract; and within a limited time period: from January 2009 through December 2010; limited to papers publication date between 2009 and 2010, and only original papers, no reviews nor discussions.

The first results obtained were summarized in the following table:

	AoO	BMJ	JCO	NEJM	Total
Number of randomized breast cancer clinical trials	32	64	57	27	**180**

Among the 180 articles, the following step was to check that the selected papers were clinical trials on breast cancer, either on early stages or advanced, and to discard meta-analysis and any review document selected by mistake.

With this second filter, about half of the articles were discarded. The table below shows the remaining papers.

	AoO	BMJ	JCO	NEJM	Total
Number of original articles	16	2	47	27	**92**

A second review was performed with the remaining papers to select only those ones which had survival endpoints under their objectives: time to progression (TTP), progression-free survival (PFS), disease-free survival (DFS) or Overall Survival (OS) as primary or secondary endpoint; and were phase II and III trials. Phase I trials were discarded since they are two small ones and whose objective is not survival.

Thus, the pre-selected trials are the following ones.

	AoO	BMJ	JCO	NEJM	Total
Number of phase II/III with survival endpoints	**14**	**0**	**26**	**8**	**48**

The final 48 selected were thoroughly reviewed to confirm they were eligible for this analysis. During this final one by one review, some papers were discarded because they did not met the criteria or did not contain enough information for our work.

The main reasons for discarding are summarized in the following table:

Reason to discard previous pre-screened papers:	
Not enough data	5
Biomarkers trials	4
Not a clinical trial	4
Retrospective	2
Substudy	1

After this final selection, the articles selection is from the following journals:

	AoO	BMJ	JCO	NEJM	TOTAL
selected for data analysis	**11**	**0**	**15**	**6**	**32**

With the 32 finally selected trials, a table was created to summarize and capture the following information:

- Primary endpoint considered, and therefore which endpoint the trial is empowered to detect differences.
- Secondary endpoint(s)
- Trial Phase: II / III
- Disease stage: Early / Advanced breast cancer
- Probability of the event in patients in the treatment and control arms
- Hazard ratio obtained for each of the endpoints
- P-value of each of the tests.
- Test used for the analysis

Appendix 1 shows the bibliographic details of the selected papers and Appendix 2 summarizes the data collected from each trial:

4.2. Brief articles' outline

Once compiled all the data from the selected papers, and before any analysis was performed, some considerations were taken:

1. Since Disease-free Survival and Progression-free Survival are equivalent endpoints, i.e. the only difference is the disease phase and this has been categorized as well, both will be considered as the same composite endpoint.
2. ORR information is not relevant for our objective, because it is not a time-to-event endpoint. Four trials (4, 6, 27 and 44) have ORR as its primary endpoint, so only information on primary analysis p-values has been recorded in our table, to distinguish those cases where this data was statistically significant or not.
3. Three papers (5, 4 and 32) showed TTF as the primary endpoint. The FDA guideline on Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics state that TTF is not an appropriate efficacy endpoint because includes discontinuation from trial for any reason: disease progression, death or toxicity. Thus, no survival information from TTF will be used, other than the p-values, as they are set as the primary endpoint in their corresponding papers, similarly to what is being done for ORR.
4. Per definition, Time to Progression is the time from randomization until objective disease progression, and death is censored (non-informative).
5. All trials selected assessed differences between the treatment arms' survival curves by means of the logrank.

Having taken these considerations, the initial observations on the data selected were the following ones:

1. In terms of primary endpoints used in the selected clinical trials, half of the trials had a composite endpoint – either PFS or DFS - as primary endpoint, while only 3 out of the 32 (9.4%) uses OS as primary endpoint. It is worthwhile to recall that the FDA recommended endpoint is the OS.

Endpoint	#	%
Overall Survival	3	9.4%
Progression	6	18.8%
Composite Endpoint	16	50.0%
ORR / TTF	7	21.9%

Table 2 - Primary endpoints used on the selected clinical trials

2. Regarding the secondary endpoints, 26 / 32 (81.25%) trials selected have OS as a secondary endpoint on their clinical protocols, while 9 / 32 (28.12%) have a composite endpoint.

Endpoint	#	%
Overall Survival	26	81.25%
Progression	5	15.63%
Composite Endpoint	9	28.16%
ORR / TTF	1	3.13%

Table 3 - Secondary endpoints used on the selected clinical trials

From these initial observations we can observe that in breast cancer clinical trials there's a tendency to use composite endpoints as a primary endpoint, and to add overall survival as secondary. Likewise, we can see that the majority of the trials use either composite endpoints or overall survival as endpoint – 29 / 32 (90.63%) OS and 25 / 32 (78.13%) CE – either as primary or secondary.

4.3. Setting of statistical parameters

As mentioned in section 3 and in *Gomez and Lagakos*' paper, to obtain the Asymptotic Relative Efficiency score, that will allow us to decide if we will have statistical benefit in our trial with the inclusion of the additional endpoint (progression) to the main endpoint (death) we need to set certain parameters in advance. These parameters are:

- *P1* : probability in the control group to find the main endpoint,
- *HR1*: ratio of the hazard functions of the two treatment groups for the main endpoint,
- *P2*: probability in the control group to find the additional endpoint,
- *HR2*: ratio of the hazard functions of the two treatment groups for the additional endpoint
- ρ: Spearmen's correlation index between both endpoints
- β: shape parameter of the Weibull function.

In the reality, each of the above parameters may adopt different values in each situation, i.e.: we may expect that in our target population only 20% of patients developing disease relapse in the control group, and overall we are optimistic and expect to have only 2% of deaths (by any cause). We may have the same situation with the hazard ratio values. Depending on what we expect to gain with the experimental treatment against the control group, we will adopt different values of the Hazard ratio for our estimations.

It is important to remark at this point, and due to the fact that all clinical trials are designed to show either superiority or non-inferiority of the treatment group against the control, that although in our literature research we found that some HR values >1 (meaning that the control group had better survival or less progression than the experimental group), we set the limit of these hazard ratios to 0.99, meaning that the although minimal, a benefit over the control group is always expected when an investigator set up a clinical trial. The reason for that is that we assumed no investigator will design a clinical trial to show that the control group is better than the experimental one, and even though it can be designed to show non-inferiority, we set as an upper limit (HR≤0.99).

With regards to the Spearmen's correlation index (ρ), we have defined three levels of correlation between endpoints:

- weak correlation: 0.15
- mild correlation: 0.45
- high correlation: 0.75

Although the correlation parameter may adopt values from -1 to 1, we used three positive values only. By this we assume that the relation between death and progression is positive, i.e. when progression increases, we foresee an increase on the number of deaths. Likewise, since we will try to focus the results from an investigator perspective, we categorized the correlation in only three levels, which we would expect that at this extent an investigator would be able to provide information on the expectations about the correlation between the endpoints.

Finally with regards to the Weibull shape parameter, β may take values <1, 1 and >1, depending if we consider that the hazard ratio is decreasing with time, constant and increasing respectively. Since the data considered here is from breast cancer data, it is easy to assume that the risks over that time remains constant or even increase once the cancer is detected. But there is literature supporting the curability of the disease after certain time[28-30], so we have included in our assessment a decreasing risks model for patients at early stages of the disease. It's the work of the investigator to find out the more reliable anticipated values of these parameters before calculations sample size and design the trial.

4.3.1. Potential Covariables

As we have seen, it is of great importance the values of the Probabilities and the Hazard Ratios obtained from the literature. These variables aim to represent the real values an investigator would have or estimate during the design phase of a clinical trial. These estimations condition the number of events needed to see statistically significant efficacy differences, and thus the size of the trial sample needed.

Therefore, and with the objective to estimate as accurate as possible the ranges in which probabilities and hazard ratios may be found, the first step in our exploratory analysis was to

determine if there was any other variable from those recorded that we should take into consideration in our analysis.

Figure 4 – P and HR differences per disease stage and trial phase with pooled data shows P and HR distributions taking into consideration the Disease Stage and the Clinical Trial Phase. It clearly showed that depending on the disease stage of the trial, i.e.: Advanced breast cancer or Early breast cancer, the Probabilities for death and/or progression are different, so it was taken the decision to consider separately, and therefore generate two study lines, for Early and Advanced stages of the disease.

With regards to the Trial Phase, no such big differences were detected neither for Probabilities nor Hazard Ratios, but further plots were performed to explore if differences from a trial phase perspective should be taken into consideration.

It is important to highlight that the P and HR values on Figure 4 does not distinguish on death, progression or composite endpoint, but are the pooled probabilities in the control arm.

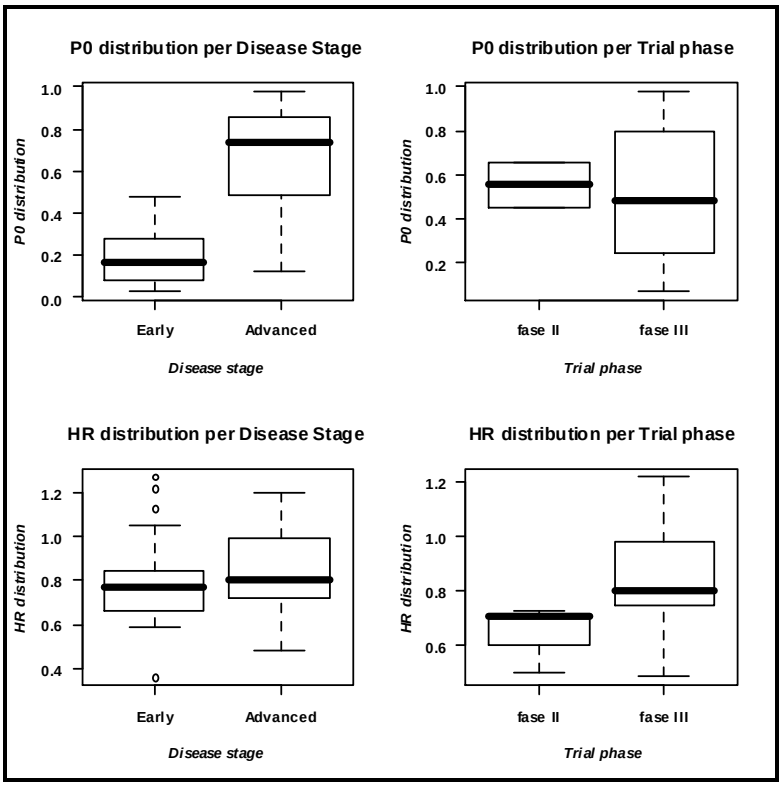


Figure 4 – P and HR differences per disease stage and trial phase with pooled data

To confirm these observations, it was repeated the exploration of Probabilities and Hazard ratios splitting them for all different type of endpoints: death (*P1* and *HR1*), progression (*P2* and *HR2*), and composite endpoints (*P3* and *HR3*).

The result of this analysis is shown in figures Figure 5 **- Probability differences per disease stage with pooled data**, Figure 6, Figure 7 and Figure 8, which confirms that:

- The probability of all of the considered survival endpoints is higher in those trials with Advanced Cancer Patients than in those with Early Breast Cancer. This conclusion does make sense, as the probability of worsening/progression and death is higher with patients in an advanced stage of the disease.
- Perhaps due to that the amount of phase II clinical trials data was less than what was obtained from phase III trials, we could not conclude that the probabilities of each endpoint are different depending on the trial phase. However, we should not be able to detect any difference in this regard, and most of the phase II trial that is positive continues with similar phase III. As previously introduced here, phase II trials are in general "pilot" trials to confirm safety of the treatment being tested. If the safety profile of the experimental treatment is acceptable, then phase III trials are set up with patients of similar characteristics.
- As foreseen in figure 4, no difference is observed in terms of hazard ratios for early or advanced clinical trials.
- Again perhaps due to the reduced number of phase II trials we are not able to observe differences on the hazard ratios between trial phases.

According to this exploratory analysis results, we will consider the disease stage as a covariable in our study, due to the differences observed in the death and progression probabilities, but discard the trial phase, as it would not make sense neither from a clinical perspective.

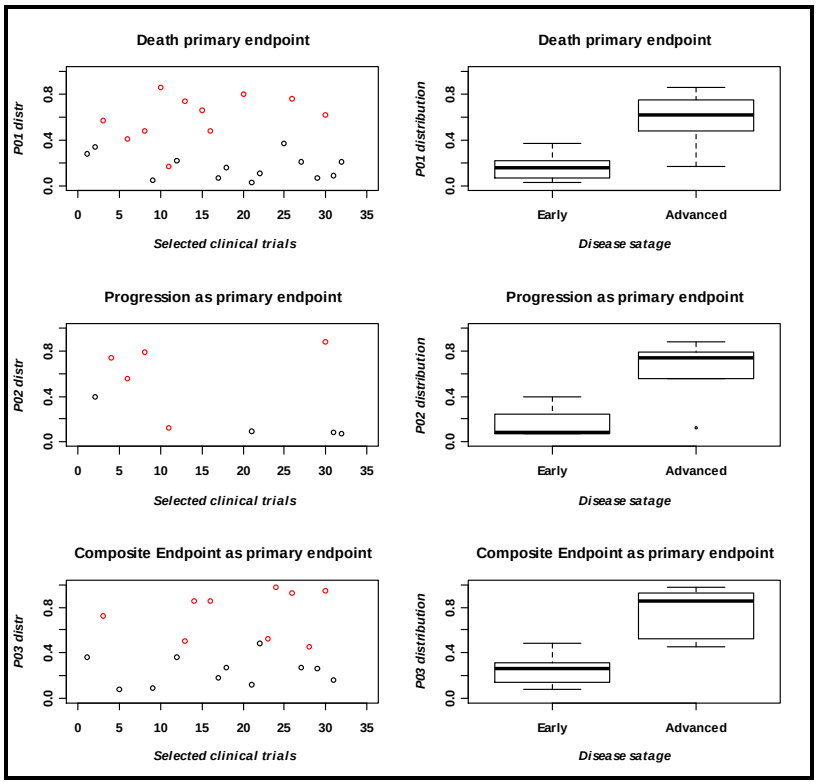


Figure 5 - Probability differences per disease stage with pooled data

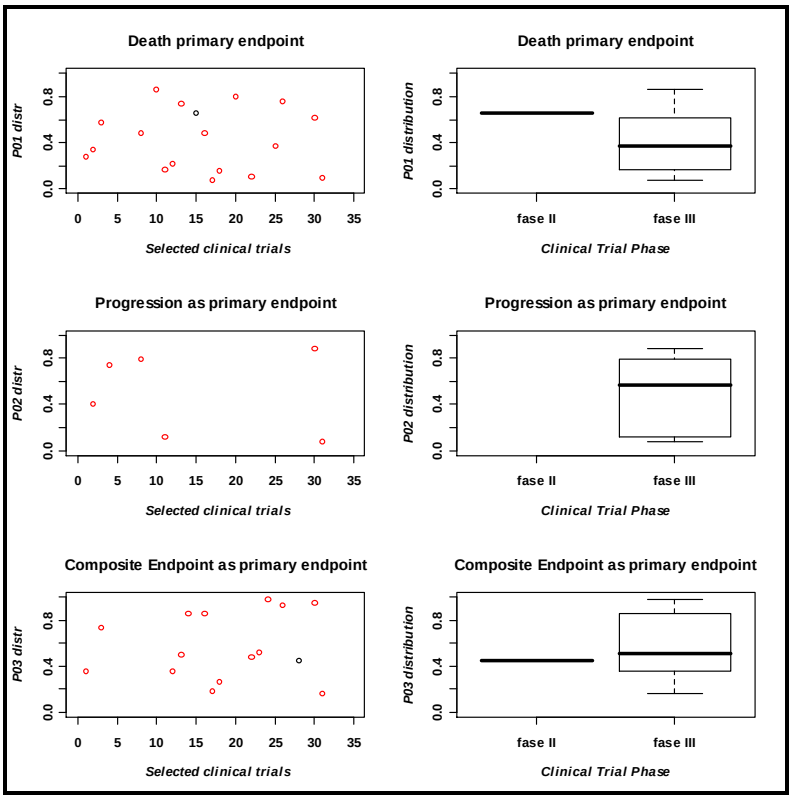


Figure 6 - Probability differences per trial phase with pooled data

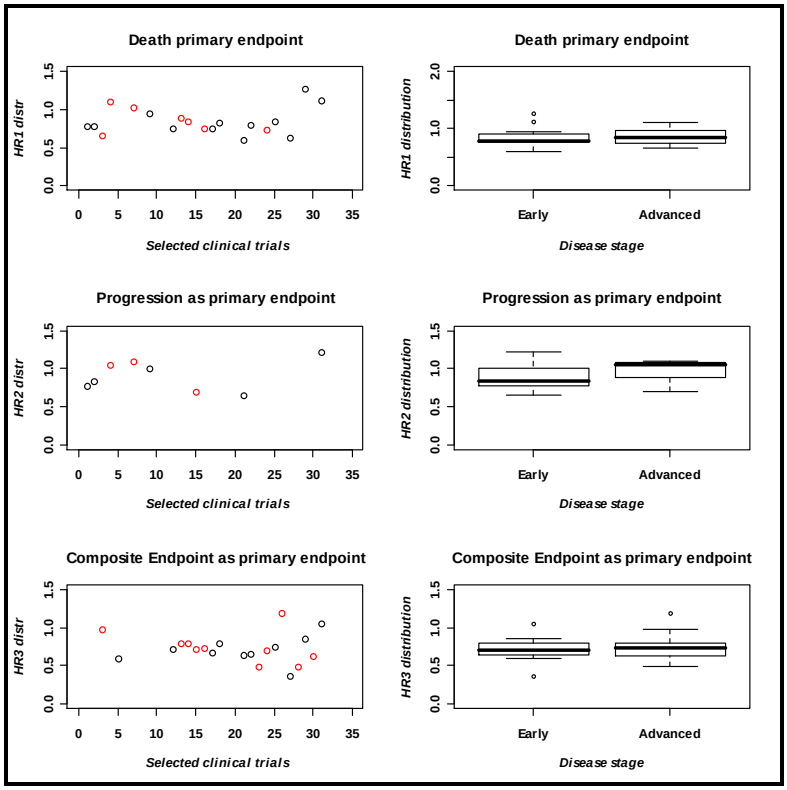


Figure 7 - HR differences per disease stage with pooled data

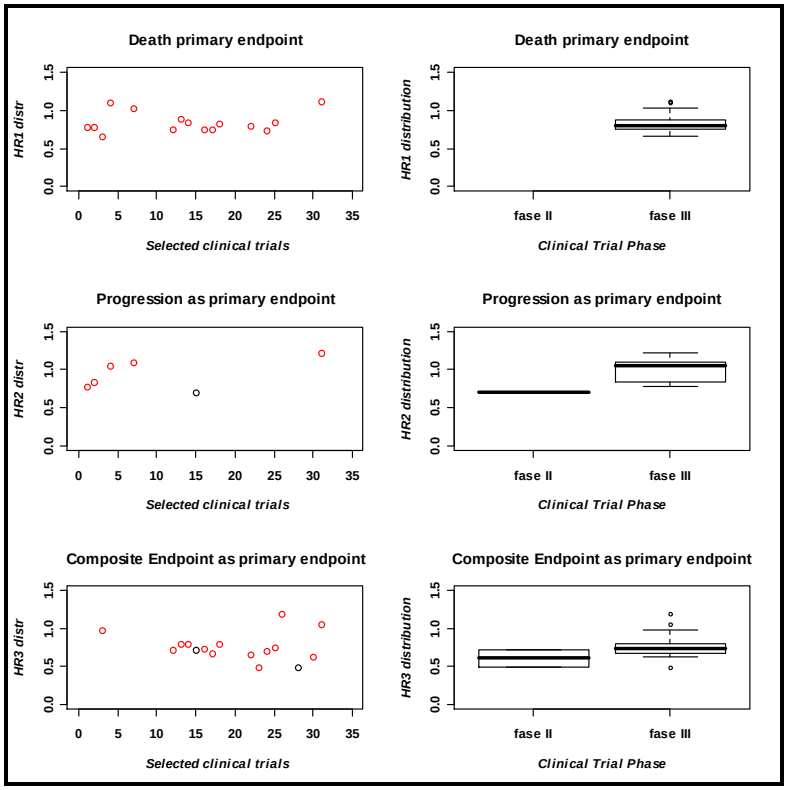


Figure 8 - - HR differences per trial phase with pooled data

4.3.2. Correlation between *T1* and *T2*

Reporting differences on survival events considering different the disease stage has been widely reported in the last decades[28]. As described, our methodology based on *Gómez and Lagakos* paper defines the correlation between the two events of interest. In our study death is the event of interest and progression the additional that combined conform the composite endpoints.

By accepting that patients detected at an early stage of the disease have better prognosis and survival than those patients in advanced breast cancer, either because the type of the disease is more aggressive, or because the disease is consequence of a previous cancer, we accept that progression in the early stage is less likely to happen than in advanced breast cancer. Thus we can assume that progression and death are less correlated in early breast cancer stages than in advanced. Hence, when we established the correlation (Spearman's correlation: ρ) between progression and death, we set it differently for each disease stage:

- For Early Breast Cancer patients we tested weak (0.15) and mild (0.45) correlation,
- For Advanced Breast Cancer patients we tested medium (0.45) and strong correlation (0.75).

4.3.3. β Values for marginal Weibull laws

For the same reason, and we have already explained, a patient who is diagnosed of breast cancer at an early stage may be considered as cured by some authors, although this is actually under discussion. We assumed in this study different hazard function shape parameters for each disease:

- For Early Breast Cancer patients we tested decreasing (0.5), constant (1) and increasing (2) hazard functions,
- For Advanced Breast Cancer patients we tested constant (1) and increasing (2) hazard functions.

4.3.4. Clustering of Probabilities and Hazard Ratios

Apart from deciding the use of Disease Stage as a covariable and analyze our data in two different scenarios: (0) Early breast cancer and (1) Advanced breast cancer, we intended to look for any pattern of association between probabilities and Hazard Ratios in our data, which would allow clustering our data.

As mentioned before, since an investigator willing to test a new treatment against the available one would never consider that the new treatment may be worse than the control, we capped all HR values to 0.99. Namely, even minimal, the investigator will always assume that his treatment is better than the one being tested.

To perform this, we plotted *P1* against *HR1* and *P2* against *HR2*, the values we will need to calculate the ARE, distinguishing per each disease stage 0 and 1.

Note that the *y* axis reads *P01* and *P02*. The *0* right after the *P* refers to the control arm. Remind that only probabilities and hazard ratios from the control group are considered to calculate the Asymptotic Relative Efficiency.

As seen in Figure 9 **- Data plot for cluster identification**, the only pair of data that allows the identification of clusters is *P1* and *HR1* at stage 0 (early Breast Cancer). For the rest of scenarios there was not enough data to identify any potential cluster.

The cluster identification was performed following the K-means clustering method in R, as described by Everitt & Hothorn[31], which calculates distances from each point to a centroid point and defines the clusters. Seen the reduced amount of data available, it was decided that the optimal amount of clusters would never be higher than 3, so it was compared the results for a 2 clusters design against a 3 clusters.

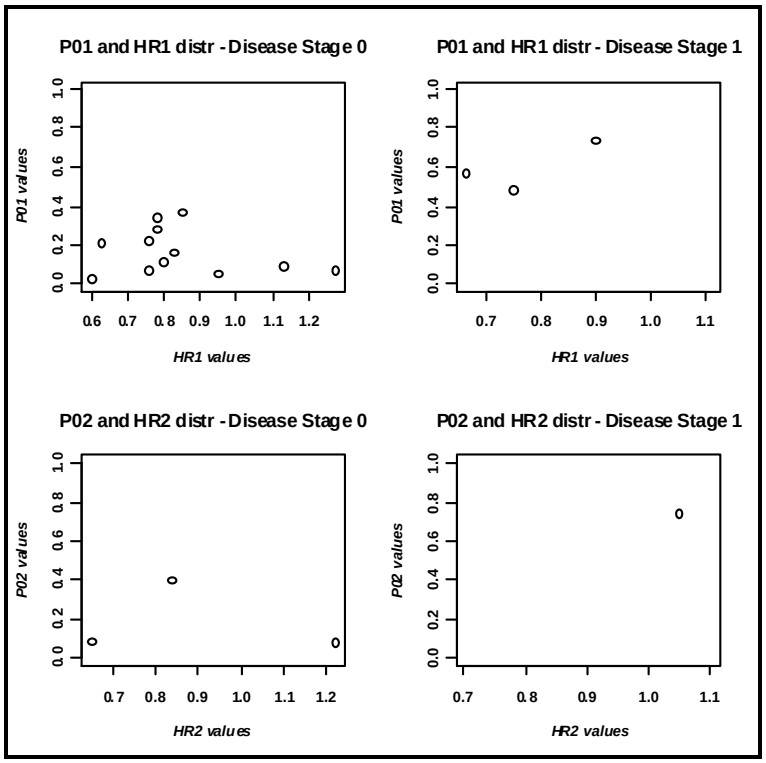


Figure 9 - Data plot for cluster identification

The results can be seen in Figure 10 - Stage 0 *P1* and *HR1* plot of the 2 clusters design and Figure 11 - Stage 0 *P1* and *HR1* plot of the 3 clusters design. These two plots identify, in function of main components, which are the most appropriate clusters. As showed as well in the picture, the 2 clusters model explains more variability, 96.26% from our data, against the 94.38% of the 3 cluster model. However, and given the objective of our study, it was decided to consider the 3 clusters design, as the 2 clusters model embraced a wide number of observations, that could make our post-analysis interpretations too general. The values of these cluster components are, thus:

Cluster	*P1*	*HR1*
1	0.070	0.976
2	0.330	0.803
3	0.133	0.730

Table 4- Cluster values for *P1* and *HR1* in disease stage 0

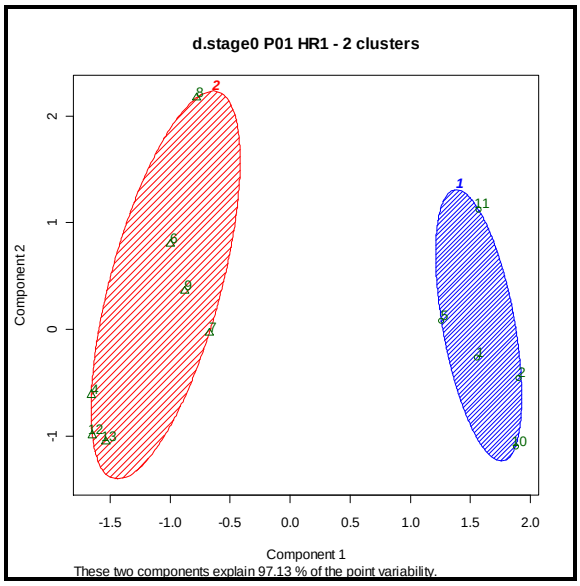


Figure 10 - Stage 0 *P1* and *HR1* plot of the 2 clusters design

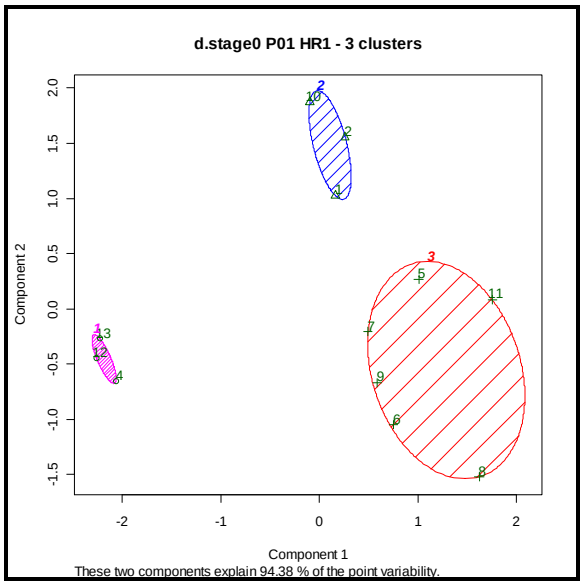


Figure 11 - Stage 0 *P1* and *HR1* plot of the 3 clusters design

4.3.5. Summary of values taken for ARE computation

Having considered the above points, and according to the values obtained from the literature for *P1*, *P2*, *HR1* and *HR2* at each disease stage, these are the values that will be used for the ARE calculation of this study.

- **Disease Stage 0 (Early Breast Cancer)**

(P1,HR1)	*P2*	*HR2*	beta 1	beta 2	*rho*
(0.070 , 0.976)	0.067	0.650	0.5	0.5	0.15
(0.330 , 0.803)	0.163	0.840	1.0	1.0	0.45
(0.133 , 0.730)	0.222	0.990	2.0	2.0	

Table 5- Values for ARE computation for Early Breast Cancer data

Identified in green is the cluster described in section 4.3.4

- **Disease stage 1 (Advanced breast cancer)**

P1	*P2*	*HR1*	*HR2*	beta 1	beta 2	*rho*
0.480	0.560	0.663	0.704	1.0	1.0	0.45
0.615	0.763	0.840	0.800	2.0	2.0	0.75
0.746	0.811	0.966	0.900			
			0.990			

Table 6- Values for ARE computation for Advanced Breast Cancer data

5. Results

5.1. ARE calculations

The objective of this Master Thesis is to study the behavior of the Asymptotic Relative Efficiency (ARE) under different scenarios by means of the methodology described in *Gomez and Lagakos*' paper, applied to breast cancer data. Briefly, the ARE value compares the efficiency of two clinical trial endpoints: a single one (death) and a composite one (death plus progression) and is able to provide us guidance on which is the best endpoint to be used in a potential clinical trial we may consider to set up. We applied this methodology to breast cancer data obtained from randomized clinical trials published during 2009 and 2010, with the intention to define realistic data ranges an investigator might find on their day to day patients. This led to realistic and interpretable results, what we are describing in this section.

We focused on the two main events considered in survival analysis in cancer: death and progression. Overall Survival (time from treatment allocation until death for any reason) is the most recommended endpoint by the regulatory agencies, but quite often, as we have observed in section 4.2, disease progression is added to death (if the investigator or sponsor may foresee long follow-up, for example), and derive to what is known as Progression-free survival: time from randomization until progression or death, whichever comes first.

Based in real data obtained from the articles, and with the perspective and information an investigator would have during the early stages of a protocol design, we describe here the results of all the scenarios created with the data available in section 4.3.5.

As a reminder, the five parameters we need to calculate the ARE are the following ones:

- *P1* and *P2*: the probabilities of death (main endpoint and defined as "1" in our case), and progression (additional endpoint and defined as "2") on the control group.
- *HR1* and *HR2*: hazard ratio of our endpoints of interest.
- β1 and β2: shape of the marginal Weibull laws governing time to death and time to progression, respectively.
- ρ: Spearman's correlation between the main event and the additional endpoint.

In section 4.3.1 we took the decision, based on the results of the exploratory analysis, that we would perform the analysis separately for Early Breast Cancer (stage 0) and Advanced Breast Cancer (stage 1) data. Thus, the number of scenarios, combination of the parameters introduced in section 5.3.5 is:

- Stage 0: 486 different scenarios
- Stage 1: 864 different scenarios

Once calculated the Asymptotic Relative Efficiency score, the output was coded to 0 when the $ARE \leq 1$, so that the recommendation would be NOT adding progression to death, meaning to keep Overall Survival as primary endpoint in our trial; or coded to 1 if $ARE > 1$, where we would recommend the addition of progression to death, and therefore the use of the composite endpoint Progression-free Survival.

As we said we intend to interpret the results from an investigators perspective, bearing in mind the information an investigator may be able to provide during the set up phase of a clinical trial. Most probably, an investigator thinks only in how much benefit from the experimental treatment he/she expects to find:, i.e. reduction on the number of deaths or patients that expects to progress *(hazard ratios*) and he/she might be able to know or expect the probability of death (*P1*) and progression (*P2*) on the control arm.

We are aware that correlation between endpoints is not a concept that an investigator is used to think on, as they might do with hazard ratios and probabilities, so to ease their job and make it as simple as possible we have considered only three levels: weak, moderate and strong correlation, measured here by the Spearman's Correlation index. We assume an investigator may understand the principles of correlation, i.e. understand what the correlation is, and might be able to suggest its level between death and progression.

Results are presented in this perspective, showed in tables with the percentage of cases where the use of the progression-free survival is recommended rather than overall survival.

5.2. Early Breast Cancer Results (Stage 0)

5.2.1. Comparing main endpoint and additional endpoint through their Hazard Ratios

We will discuss in this section the situation where the investigator is able to give us information about the hazard ratio expected to be observed in terms of the main endpoint (death), and could also be able to give us information about the hazard ratio that might be observed if we would focus on disease progression.

Tables 7 and 8 show the percentage of scenarios, given *HR1* (hazard ratio of death), and *HR2* (hazard ratio of progression), with *rho* values 0.15 (weak correlation between them) and 0.45 (moderate correlation).

The first thing we can see when we look at the results of the ARE calculation for those scenarios considering weak correlation, is that there are two different but very clear situations: scenarios where the use of composite endpoint is recommended in 100% of the cases, and those where its use is not recommended at all. For instance, at both correlation levels considered in disease stage 0, when the *HR2* is lower (less number of progressions expected, so more beneficial for patients) than the expected *HR1*, in all cases (100%) the ARE score recommends the use of the composite endpoint. Meaning that in cases where we expect to see a low *HR2* value and this value is even lower than *HR1*, the use of composite endpoint will be in all circumstances recommended.

On the other hand, scenarios where *HR2* is not expected to show much benefit (*HR2*=0.99), and the *HR1* values are ≤0.803, the use of composite endpoints is contraindicated, and is death the only endpoint that shall be used, thus Overall Survival.

It must be highlighted though, that in these scenarios where an investigator may expect very mild benefit on the additional endpoint, where in most cases the investigator would not include the endpoint, we observe that if we expect an *HR1* value of 0.976, in 55.6% of the cases the inclusion of progression would be advisable, and the composite endpoint might be the recommended endpoint. In other words, if we would not expect a big effect in terms of death differences between the treatment groups, we would consider the addition of

progression in roughly half of the situations even if we didn't expect either big difference between groups.

If we focus on *HR1* values now, we have already seen that if *HR2* is lower that that, the use of the composite endpoint is recommended. But for values of *HR1* = 0.73 (the most positive of the values we have used in our calculations, meaning that we expect a high reduction on the number of deaths on our experimental group), we see how for values of *HR2* of 0.84, the use of composite endpoint is still recommended in 77.8% of the cases, and drops to 0% when the *HR2* value is 0.99. As the *HR1* value increase (less effect over the control group), we see how for values of *HR2*=0.84 the percentage of scenarios where we would recommend the use of the composite endpoint would increase to 100%.

Likewise, if instead of weak correlation between death and progression we consider that there's a moderate correlation situation (ρ = 0.45), we observe that the pattern is similar to what we have seen with ρ =0.15. For positive values of *HR2* (0.65), it is always recommended to use the composite endpoint independent of the *HR1* value (at the selected values).

It is worthwhile to highlight how the increase of the dependence between both endpoints affects the ARE score. We can see in (*HR1*,*HR2*) = (0.803,0.84), how when death and progression are almost independent, the percentage of cases where the composite endpoint is recommended is 100%, while when we increase the correlation between them, this percentage drops until 11%; and when (*HR1*,*HR2*) = (0.73,0.84), when ρ = 0.15, the percentage of scenarios where composite endpoints is recommended is 77.8%, while when ρ = 0.45, this value turns to 3.7%. In line with this observation, it's important to highlight how in scenarios where we combine high *HR1* values (0.976) with *HR2*=0.99, the addition of progression is not recommended at all (0% of the cases), while with low correlation we observed that in 55.6% of the cases it was recommended. Looking at scenarios where we to have *HR2* values around 0.84, the use of composite endpoint is recommended only when the *HR1*=0.976.

It is important then to highlight how the effect of the increase of the correlation between both endpoints results in a reduction of the ARE score, and therefore in a reduction on the cases where the composite endpoint is recommended.

HR1* | *HR2	**0.65**	**0.84**	**0.99**
0.73	100.0%	77.8%	0.0%
0.803	100.0%	100.0%	0.0%
0.976	100.0%	100.0%	55.6%

Table 7- stage 0 *HR1*/*HR2*: % of scenarios where CE is recommended with ρ=0.15

HR1 | *HR2*	0.65	0.84	0.99
0.73	100.0%	3.7%	0.0%
0.803	100.0%	11.1%	0.0%
0.976	100.0%	100.0%	0.0%

Table 8 - stage 0 *HR1*/*HR2*: % of scenarios where CE is recommended with ρ=0.45

5.2.2. Comparing main endpoint and additional endpoint Probabilities

We may find the situation where our investigators' team would have an idea on the probability of the death in our control group, and is able to come up with an expected probability to find progression in the same group. We compared those situations where we would need to decide whether we should use the main endpoint only or adding progression based only in the probability of these events. The results are summarized in tables 9 and 10, again taking into account the correlation between progression and death. We see in this case a pattern not as conclusive as we had seen with the hazard ratios.

In all cases, at given *P1* (probability of death in the control group), the percentage of scenarios where the additional endpoint is recommended to be used increases with *P2* (probability of progression in the control group), either when there is a weak correlation or when there's a moderate one between endpoints. We can see how for *P1*=0.07 (lowest level of probability considered), even for *P2*=0.07 values, the addition of progression is recommended in 66.7% of the cases, and how this increase until 100% of the cases when *P2*=0.22.

Likewise, we see how for high *P1* values (*P1*=0.33), for *P2*=0.01 the composite endpoint is recommended in 33% of the cases, and 66.7% when *P2*=0.22 (the highest level considered).

Having said that, we observed as mentioned on section 5.2.1 that any increase of the correlation between endpoints causes a reduction of the ARE results, which derives to a lower percentage where the composite endpoint is recommended. As an example, for a *P1*=0.07 and *P2*=0.16, we have that for a low correlation (ρ=0.15) the percentage of recommendation to use a composite endpoints is 88%, while if we increase the expected correlation to 0.45, the percentage of cases where the use of composite endpoint is recommended is 66.7%.

P1 | *P2*	0.07	0.16	0.22
0.07	66.7%	88.9%	100.0%
0.133	33.3%	66.7%	66.7%

0.33	33.3%	66.7%	66.7%

Table 9 - stage 0 *P1*/*P2*: % of scenarios where CE is recommended with ρ=0.15

P1* | *P2	**0.07**	**0.16**	**0.22**
0.07	66.7%	66.7%	66.7%
0.133	33.3%	33.3%	37.0%
0.33	33.3%	33.3%	44.4%

Table 10 - stage 0 *P1*/*P2*: % of scenarios where CE is recommended with ρ=0.45

At this point we have checked death and progression as endpoints in terms of hazard ratios and probabilities. If we would be asked now based in which parameters we could easily take a decision, it is clear that the analysis of the hazard ratios provided recommendations of either 100% or 0% of cases in 6 or 7 out of the 9 scenarios, while if we check the probabilities, only in 1 case scenario 100% of the cases recommended the addition of progression to death.

5.2.3. Checking (*P1*, *HR1)* against *P2*

To reduce the number of scenarios, and since the data available allowed us building clusters for *P1* and *HR1*, we analyze in the following two sections how these chosen clusters combine with *P2*: the expected probability of progression in the control arm; and *HR2*: the expected hazard ratio for progression.

As described in section 4.3.4, we clustered the data in three groups. We distinguish one where death has a low probability (*P1*=0.07) of being observed in the control group, and where we do not expect either to detect a big benefit in terms of the hazard ratio (*HR1*=0.976). In these situations, the addition of progression, even if the probability of detecting progression is low (*P2*=0.07), the percentage of scenarios where the composite endpoint is recommended in 66.7%, and increases when *P2*=0.16 to 88.9% of the cases, and it is recommended in all cases when *P2*=0.22. It is worthwhile to highlight that in the first case where *P2*=0.07, investigators might have declined the suggestion of adding progression due to the low probability of observing that, but the ARE results show how in almost 67% of the cases its addition is supported.

With regards to the other two clusters, we see how for cluster (*P1*, *HR1)* = (0.330,0.803), the percentage of cases where the composite endpoint is recommended is 66.7% for all values of *P2* considered. And we observe something similar with the third cluster (0.133,0.730): if the

investigator would expect a probability to identify progression of 0.07, its addition would be recommended in 44.4% of the cases, increasing to 66.7% for *P2* values of 0.16 and 0.22.

Similar to what has been observed in sections 5.2.1 and 5.2.2, we observe how these results are affected with the increase in the expected correlation between endpoints (from ρ=0.15 to ρ=0.45). And as happened in scenarios with expected Spearman's correlation index ρ=0.15 we observe that for each cluster, the percentage of scenarios where the composite endpoint is recommended remains barely constant for each of the *P2* values. Namely, we observe again how for cluster: (0.07,0.976) it is recommended the addition of progression in 66.7% of the cases for *P2* values of: 0.07, 0.16 and 0.22; and between 33% and 44% for clusters (0.330,0.803) and (0.133,0.730).

We defined all parameters' ranges based on real data from the literature, but seen the results obtained here, it would have been interesting to observe the behavior of the ARE beyond the upper and lower limits to confirm if it does remain constant and if not, we could have estimated from which values it increases or decreases.

(P1,HR1) | P2	**0,07**	**0,16**	**0,22**
(0.07,0.976)	66.7%	88.9%	100.0%
(0.330,0.803)	66.7%	66.7%	66.7%
(0.133,0.730)	44.4%	66.7%	66.7%

Table 11 - stage 0 (*P1,HR1) /P2*: % of scenarios where CE is recommended with ρ=0.15

(P1,HR1) | P2	**0,07**	**0,16**	**0,22**
(0.07,0.976)	66.7%	66.7%	66.7%
(0.330,0.803)	33.3%	33.3%	44.4%
(0.133,0.730)	33.3%	33.3%	37.0%

Table 12 - stage 0 (*P1,HR1)/P2*: % of scenarios where CE is recommended with ρ=0.45

5.2.4. Checking the (*P1, HR1)* cluster against *HR2*

In line with what has been done for section 5.2.3, we compared the clustered data of (*P1*, *HR1*), with the chosen ranges for *HR2*, the effect on progression that we would expect in our trial. Table 13 - stage 0 (P1,HR1)/HR2: % of scenarios where CE is recommended with ρ=*0.15* and Table 14 - stage 0 (P1,HR1)/HR2: % of scenarios where CE is recommended with ρ=*0.45* summarize the results.

We observe, similar to what we did before, how for *HR2* values of 0.65, the addition of progression is recommended in 100% of the cases for ρ=0.15 and ρ=0.45. On the other hand, and as we could foresee, in most of the cases when *HR2*=0.99 (very mild effect over progression expected), the addition is not recommended. Again, it is worthwhile to highlight that for scenarios where a weak correlation is expected (ρ=0.15), in trials belonging to the (0.07,0.976) cluster and *HR2* expected of 0.99, the addition of progression is recommended in 55.6% of the cases, which would be interesting for investigators.

It is interesting the behavior of the ARE score for HR values of 0.84. We can see how for trials adhered to the first cluster (0.07,0.976), both with weak or medium correlation, the addition of progression is always recommended, while for those in (0.133,0.730) cluster, we see how the ARE results do not recommend its addition, or in very few cases. Finally, we can see the relevant effect of the correlation in trials included in cluster (0.330,0.803), for which at soft correlation scenarios would recommend the use of composite endpoint in 100% of the cases, while if it does increase (ρ=0.45), the percentage of cases falls until 11.1%.

(P1,HR1) | HR2	**0,65**	**0,84**	**0,99**
(0.07,0.976)	100.0%	100.0%	55.6%
(0.330,0.803)	100.0%	100.0%	0.0%
(0.133,0.730)	100.0%	*77.8%**	0.0%

Table 13 - stage 0 (*P1,HR1)*/*HR2*: % of scenarios where CE is recommended with ρ=0.15

*33% of the total number of scenarios had an ARE <1.1

(*P1,HR1*) | *HR2*	**0,65**	**0,84**	**0,99**
(0.07,0.976)	100.0%	100.0%	0.0%
(0.330,0.803)	100.0%	11.1%	0.0%
(0.133,0.730)	100.0%	3.7%	0.0%

Table 14 - stage 0 (*P1,HR1)*/*HR2*: % of scenarios where CE is recommended with ρ=0.45

5.3. Advanced Breast Cancer Results (Stage 1)

Following the same procedure as what has been done for Early Breast Cancer data, we looked at Advanced Breast Cancer ARE scores and determined its results in function of the parameters investigators may use during the set up phase of a clinical trial (probabilities and hazard ratios). The main difference between stage 0 and stage 1 data can be found in the probabilities of observing the events. For stage 0 we found probabilities:

- *P1*: 0.07, 0.133 and 0.330;
- *P2*: 0,07, 0,16 and 0,22

while for stage 1 we have:

- *P1*: 0.48, 0.615 ad 0.746
- *P2*: 0.56, 0.763 and 0.811

We shall bear in mind that a patient detected of breast cancer at an early stage of the disease, during a routine check, has a wide range of treatment possibilities: chemotherapy, radiology, surgery, which provides the patient the possibility to be considered "clean" at the end of the treatment period, and thus very low probability of observing death or progression. On the contrary, patients with an advanced breast cancer are usually patients with metastatic diseases (cancer is widespread in other body locations), and some of them may have undergone already surgery and chemotherapy treatment. In these patients we observed an increased probability of observing death and progression.

As already highlighted, the correlation between the main endpoint (death) and progression) is also different to what we have used for early breast cancer data. Seen this increase on the probabilities, and the status of the patients, we have considered all scenarios with moderate (ρ=0.45) and strong correlation (ρ=0.75).

5.3.1. Comparing the main endpoint's and additional endpoint's Hazard Ratios

The first conclusion that could be drawn from the data summarized in tables Table 15 and Table 16, is that almost in all scenarios where the expected *HR2* value is better that *HR1*, the

ARE recommends the addition of progression and use the composite endpoint in the clinical trials.

From the data analyzed, this conclusion has only two exceptions. At the highest correlation model considered, spearman's correlation index of 0.75, when *HR1*=0.84 and *HR2*=0.80, although *HR2* is slightly better than *HR1*, the ARE does not recommend in any case the inclusion of progression.

The second exception observed may be due to the same effect as the observed in the early breast cancer data: an increase in the correlation between endpoints makes decrease the ARE results. We can see how in scenario with Spearman's correlation index of 0.45 and *HR1*=0.84 and *HR2*=0.80, the percentage of scenarios for which we would recommend the composite endpoint in 100%, while in the scenario with a ρ=0.75, this percentage falls to 0%.

Finally, in scenarios with the lowest *HR1* considered (0.663) and the lowest *HR2* (0.704), i.e. we have the best case scenario where it is expected to observe an important improvement in terms of death and progression, we see how, for both moderate and high correlation scenarios, the ARE recommends the use of composite endpoint in 66.7% and 58.3% of the cases respectively.

HR1* | *HR2	**0.704**	**0.80**	**0.90**	**0.99**
0.663	66.7%	0.0%	0.0%	0.0%
0.84	100.0%	100.0%	0.0%	0.0%
0.966	100.0%	100.0%	100.0%	0.0%

Table 15- stage 1 *HR1*/*HR2*: % of scenarios where CE is recommended with ρ=0.45

HR1* | *HR2	**0.704**	**0.80**	**0.90**	**0.99**
0.663	58.3%	0.0%	0.0%	0.0%
0.84	100.0%	0.0%	0.0%	0.0%
0.966	100.0%	100.0%	100.0%	0.0%

Table 16 - stage 1 *HR1*/*HR2*: % of scenarios where CE is recommended with ρ=0.75

5.3.2. Comparing the main endpoint's and additional endpoint's Probabilities

As we can observe in tables Table 17 and Table 18, for any *P1* and *P2* combinations, the use of composite endpoint is recommended in roughly the 50% of the cases, which we may interpret

as if the frequencies of observing either endpoint are not conclusive and may not be used to determine whether or not we should use the composite endpoint.

Likewise, we can see how the effect of the correlation that we have seen in the sections before, is diluted when comparing *P1* and *P2*. We can see how both tables may be summarized in a singe one.

As a conclusion of the results below, we could say that the comparison between *P1* and *P2* values in the ranges taken from clinical trials with advanced patients does not really give a clear clue on the strategy to follow, and would not be a case to consider in a potential real case.

P1* | *P2	**0.56**	**0.763**	**0.811**
0.48	50.0%	50.0%	50.0%
0.615	47.9%	50.0%	50.0%
0.746	43.8%	41.7%	41.7%

Table 17 - stage 1 *P1*/*P2*: % of scenarios where CE is recommended with ρ=0.45

P1* | *P2	**0.56**	**0.763**	**0.811**
0.48	47.9%	50.0%	50.0%
0.615	43.8%	50.0%	50.0%
0.746	43.8%	41.7%	41.7%

Table 18 - stage 1 *P1*/*P2*: % of scenarios where CE is recommended with ρ=0.75

6. Discussion and Recommendations

In general, we have seen that in those scenarios where the investigator expects (1) low hazard ratio values, what means a reduction on the number of deaths in the experimental group with respect to the control one, or (2) high probability of observing death in our control arm (high *P1*) the addition of the secondary endpoint is only recommended in those cases where *HR2* and *P2* values are similar to *HR1* and *P1*, or *HR2* values are lower than HR1 and *P2* higher than *P1*. In other words, if an investigator expects a high reduction on the number of deaths in the experimental arm (*HR1*~0.7), or the probability of death is high (*P1*=0.3 in early breast cancer and *P1*=0.74 in advanced breast cancer), the most efficient endpoint will be Overall Survival, unless *HR2* is lower and *P2* higher than HR1 and *P1* respectively.

Likewise, if we would consider the addition of progression to our clinical trial endpoint, and we would be expecting low *HR2* values – reduction on number of patients that progressed – of 0.7 or lower, in most of the cases the use of the composite endpoint is recommended irrespective of the values of the rest of parameters. This means that if the difference between the observed number of deaths from both treatment arms is equal or higher than 30%, the addition of progression to the clinical trial endpoint will always increase the statistical efficiency.

We have not seen, though, such a clear effect of *P2* – probability of observing patients that progressed in the control arm – in terms of deciding whether or not the composite endpoint is recommended, although we identified a trend: high *P2* values increase the number of scenarios where it is recommended the addition of progression.

Furthermore, we observed that:

- There is an inverse relation between *HR1* or *P1* and the percentage of scenarios where the use of composite endpoint is recommended. The lower *HR1* is, or the higher the expected *P1* value is, the lower is the ARE value obtained, and hence less cases where the use of composite endpoint is recommended.

- This relation turns to directly proportional when we look at the additional endpoint information (*HR2* or *P2*). In those scenarios where we have low *HR2* (high benefit in terms

of progression) and high *P2* values, the ARE value increases and thus the percentage of scenarios where the composite endpoint is recommended as well.

Apart from this, we should take into consideration the correlation between endpoints, death and progression. We observed that the recommendations may vary depending on the breast cancer stage, as we have summarized in the following sections 6.1.1 and 6.1.2. In cancer, these correlation differences might be linked to the time between both endpoints. In early breast cancer, progression and death are less correlated and thus the time from progression until death - if this is ever observed - is longer than in advanced breast cancer, where progression and death are more correlated and closer in time. Hence, it makes more sense in early breast cancer scenarios to include progression as an endpoint. Taking care of the disease progression concerns patient's quality of life so if we expect to reduce the number of disease progressions with our experimental treatment, we are improving patient's quality of live even if not improving the overall survival and this would still be adding a clinical benefit on patient's life. We should obviously take toxicological considerations to make such statements about quality of life, but we are assuming here that less rate of disease progression is translated in an increase of patients' quality of life.

However, in advanced breast cancer progression and death are more correlated and closer in time, so from a clinical perspective this could mean that a patient that progressed is more likely to die than another one that did not progressed. In this case, observing progression increase the probability of observing death, so in some cases detecting progression would not add any value neither from a statistical perspective nor clinical.

Since the ranges of *HR* and *P* are different for both disease stages, we can not derive the reason of the effect of correlation over the scenarios, but only mention potential causes that should be analyzed in future works:

- Stage 1 *P* ranges were higher than stage 0 ones, so it could be that the effect of the correlation's dilution is due to these higher values. Meaning that the correlation effect is observed mainly for low probability values, while this effect is not observed as we increase them.

- Likewise, we tested only low (ρ=0.15) and medium (ρ=0.45) correlation for early breast cancer data, and medium and high (ρ=0.75), so another reason might be that the effect of

the correlation is high when both endpoints are not correlated, while this correlation softens as we increase the correlation between these.

Apart from identifying trends and see how the Asymptotic Relative Efficiency behaves wit the different parameters considered, we tried to identify rules and define guidelines for future investigators that might consider the use of this procedure to assess.

As a result of our simulations, where we analyzed all possible scenarios with combinations of the parameters required to calculate the ARE: *P1*, *P2, HR1, HR2,* β*1*, β2 *and ρ* we are able to derive the following guidelines or recommendations:

6.1.1. Early Breast Cancer data interpretation

1. If the expected *HR2* value is low - ≤0.65 in our study -, the added value of the additional endpoint is enough to recommend the inclusion of progression in all cases. Therefore the use of composite endpoint is always recommended indistinctly of the rest of the parameters.

A significant benefit over the disease progression may be translated into an increase on the quality of patient's life, which is turning to be an actual objective of oncologists in general. As discussed in the introduction, increasing patients' life should not be the only objective of our investigators. The quality of life an investigator may be able provide to these patients sometimes is as important as the time they live. We should definitely include progression, because by reducing the rate of patients that progress, we increase in most of the cases the quality of our patients' life.

2. In early breast cancer we considered low (ρ=0.15) and medium (ρ=0.45) correlation between death and progression. We observed how whenever the correlation between *T1* and *T2* increases, the ARE value decreases under equal conditions of *HR1*, *HR2*, *P1* or *P2*. This means that the increase of the level of correlation is more conservative and restrictive with regards to the addition of progression and to recommend its inclusion as part of a composite endpoint. In other words, as the correlation increases, the effect of addition of progression to our endpoint provides less information. It is important to highlight that the investigator can not control the level of correlation between death and progression. As it

has already been explained in this study, the investigator is asked during the set up phase of the clinical trial the expected grade of correlation between progression and death, to be able to predict as accurate as possible the efficiency of single endpoint and the composite one.

3. When the expected correlation between death and progression is weak and we expect to have high values of *HR1* (0.976), although the expected *HR2* (effect over progression) may be mild (0.99), in half of the cases the inclusion of progression might be recommended.

Most part of the investigators, in the scenario where they foresee a mild effect over patient's progression would reject to include it as part of a clinical trial endpoint. We are in a scenario where we expect to find mild improvement from the experimental therapy against the control group, but we may find situations where this improvement is worthwhile the effort.

4. In scenarios where the expected *HR2* = 0.84, and we expect low correlation between death and progression, if *HR1* > 0.73, in 100% of the cases it is recommended the use of the composite endpoint. At *HR1*=0.73, though, we find that in 77.8% of the cases we find that the composite endpoint as the most advisable, so we should test the ARE behavior in *HR1* values <0.73 to see if the percentage keeps decreasing. As the correlation increases, and as above highlighted, this relation disappears, and for medium correlation values the addition of the composite endpoint is only suggested when no real benefit is expected for *HR1*. What may even suggest to not using death as an endpoint in our trial.

5. For *P2* values > 0.16 and low correlation between endpoints, at the analyzed *P1* values range, in > 66.7% of the cases the use of composite endpoint was recommended, while as the correlation increases the recommendation falls to around 30%.

6. In terms of the (*P1*,*HR1*) clusters analyzed, we observed that those trials that fell under the (0.07,0.976), in 85.2% and 66.7% of the scenarios recommended the composite endpoint when correlation was low and medium respectively.

6.1.2. Advanced Breast Cancer data

1. At the correlation levels this study was conducted, we can consider that the effects of the correlation can be negligible. We have already mentioned the limitations on this regard in our work, but from a clinical perspective, in advanced patients progression and death are correlated and often not very distant in time from each other. We tested for advanced breast cancer data correlation levels (ρ) of 0.45 and 0.75, and what we have seen is that the effect seen in stage 0 barely disappears. Under same conditions of *HR1, HR2, P1* or *P2,* if we compare the percentage of scenarios where the ARE value recommends the use of composite endpoint between scenarios with moderate correlation and those with high, we observe that the difference between these percentages is never higher than 10%, being in most of the cases the same values.

2. Only in cases where the expected *HR2* is lower than the *HR1*, or *HR2* values are ~0.66, the use of composite endpoint is recommended in about 66% of the cases. Again, death and progression are linked in advanced stages of the disease, so only high improvements on progression may be considered to be added to death in a clinical trial.

3. In line with the previous recommendation, only when *HR2* is ≤ 0.8 and *P2* value is >0.56, more than 65 % of the scenarios recommend the addition of progression to death. Even if the prognosis for a patient with advanced breast cancer is bad, observing improvements of more than 20% in terms of progression provide enough statistical benefit to recommend the use of the composite endpoint.

4. Decisions based on *P1* and *P2* do not provide clear guidance on what is best, as the percentages of use or not of the composite endpoint are evenly distributed.

7. Limitations and future development of the statistical methodology

The first limitation of this study is an advantage as well. The bibliographic research has been limited to papers published between 2009 and 2010. The direct impact of this is that the number of papers is reduced, and although the initial pull of papers was big, the final selection was a reduced sample of all this. At the same time this is a benefit for our data, as taking data from a wider time period may include studies with different standards of care that may had given us a less homogeneous data and one source of bias.

As we already mentioned in section 3, the methodology used in our study assumes that the end of study censoring is the only non-informative one. However, clinical trials often have drop out patients due to other uncontrolled circumstances, i.e. withdrawal of patient's consent, change of patient's residence. In these cases the information provided by patients is censored, and although the trial team does all efforts to try to get the information they may not reach to get it. Therefore, this methodology should be tested as well assuming that there might be situations where apart form the end of study, we have non-informative censoring of data.

With regards to the raw data ranges selection, some of the selected papers do not provide all the information needed, and therefore from some of the endpoints the value of probabilities and hazard ratios that have been estimated were not taken from a sample with the appropriate number of values from the literature.

In terms of the number of points taken, and since the purpose of this study was to be as concrete as possible and to generate scenarios as similar to the reality as possible, we reduced each variable to 3 or 4 points, with even distribution among them. The results showed, though, that the percentage of scenarios where the ARE recommended the use of composite endpoint was changing sometimes from one point to the next from 0% to 100%, what would suggest that it would be worthwhile to consider points between the ones already selected to detect and better characterize the behavior of ARE for that variable. This would mean though the creation of more scenarios, which may generate a higher number of data combinations and therefore more computational power requested.

Regarding future research that could follow this study, we could work to validate the results and the conclusions obtained in this work. We would need to come back to the papers found in the literature and check the primary endpoint chosen by the investigators and their p-

values. We should be able to see if papers used the proper test for the clinical trial, and if the outcome matches with the ARE recommendation.

The real benefit of the statistical methodology introduced by *Gómez and Lagakos* is not checking whether clinical trials that already published their results took the correct decision when they decided to use a composite endpoint in stead of a single one, but to prove that this methodology can be used when an investigator or a company wants to design a clinical trial.

Using this methodology, a user friendly *ARE calculator* might be set up. Investigators would be supposed to introduce the expected values of the parameters they know or they expect, and the ARE results would recommend the most efficient endpoint.

Once defined the most efficient endpoint, we would calculate the sample size according to that clinical trial endpoint, for which calculation we do not need any other information than the one needed for the ARE calculation. We would be able, then to ensure that the sample size obtained is the most efficient and optimal one.

Bibliography

1. G. Gómez SL. Statistical considerations in the use of a composite time-to-event endpoint for comparing treatment groups. Statistics in medicine 2012:(accepted).

2. . (Accessed at http://dtp.nci.nih.gov/timeline/noflash/milestones/M4 Nixon.htm.)

3. Hilner. Efficacy dies not necessarily Translate to cost Effectiveness. Journal of Clinical Oncology 2009;27(13 (May 1)):2111-3.

4. Kolata G PA. Costly Cancer Drug Offers Hope, but Also a Dilemma. The New York Times 2008 July 6.

5. PB. B. Limits on Medicare's ability to control rising spending on cancer drugs. N Engl J Med 2009;360:626–33.

6. Press. A. Docs to weigh costs in choosing cancer treatment. In; 2008.

7. Fausto Roila EC. Quality of life as a primary end point in oncology. Annals of Oncology 2001(12 (Suppl. 3); S3-S6).

8. MA. Kozminski. How long and how well: Oncologists' attitudes toward the relative valie of life-prolonging vs quality of life-enhancing treatments. Medical Decision Making 2010.

9. FDA. Guidance for Industry. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. In; 2007.

10. John R. Johnson GW, and Richard Pazdur. Endpoints and United States Food and Drug Administration. Journal of Clinical Oncology 2003;21(7):1404-11.

11. ED. Saad. Progression-free survival as surrogate and as true endpoint: insights from the breast and colorrectal cancer literature. Annals of Oncology 2010;21:7-12.

12. TR. Fleming. Issues when using progression-free-survival when evaluating oncology products. Journal of Clinical Oncololgy 2009;27:2874-80.

13. ICH E9 - Statistical Principles for Clinical Trials. 1998. (Accessed at http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/statistical-principles-for-clinical-trials.html.)

14. G. Chi. Some issues with composite endpoints in clinical trials. Fundamental & Clinical Pharmacology 2005;19:609-19.

15. I. Ferreira-González. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. Journal of Clinical Epidemiology 2007;60:651 - 7.

16. N. Freemantle. Composite outcomes in randomized trials: greater precision but with greater uncertainty? Jama 2003;289(19):2554-9.

17. H. J. Dargie. Effect of carvedilol on outcome after myocardial infarction in patients with left-ventricular dysfunction: the CAPRICORN randomized trial. Lancet 2001;357(9266):1385-90.

18. J. R Hampton. Clinical trial safety committees: the devil's spoon. British medical journal 2000;320(7229):244-5.

19. Asociación Española Contra el Cáncer. (Accessed at https://www.aecc.es/SobreElCancer/CancerPorLocalizacion/CancerMama/Paginas/incidencia.aspx.)

20. BreastCancer.org. (Accessed at http://www.breastcancer.org/symptoms/understand_bc/statistics.jsp.)

21. E D. Saad. Progression-free survival as surrogate and as true end point: insights from the breast and colorectal cancer literature. Annals of Oncology;21(1):7-12.

22. R. Tuma. Progression-free survival remains debatable endpoint in cancer trials. 2009.

23. A. Di Leo. Overall survival is not a realistic endpoint for clinical trials in advanced solid tumors: A critical assessment based on recently reported phase III trials in colorectal and breast cancer. Journal of Clinical Oncology 2003;21:2045-7.

24. AM. Stone. Research outcomes and recommendations for the assessment of progression in cancer clinical trials from a PhRMA working group. European Journal of Cancer 2011;47:1763- 71.

25. R. Peto. Asymptotically Efficient Rank Invariant Test Procedures. Journal of the Royal Statistical Society 1972;135:185 - 207.

26. J. M. Bland. The logrank test. British medical journal 2004;328(7447):1073.

27. Test Survival Curve Differences (Accessed at http://stat.ethz.ch/R-manual/R-patched/library/survival/html/survdiff.html.)

28. A. O Langlands. Long-term survival of patients with breast cancer: a study of the curability of the disease. British medical journal 1979;2(6200):1247-51.

29. P.H. Zahl. Long-term survival of breast cancer in Norway by age and clinical stage. Statistics in medicine 1997;16(13):1435-49.

30. P. H. Zahl. Frailty modelling for the excess hazard. Statistics in medicine 1997;16(14):1573-85.

31. B. Everitt TH. A Handbook of Statistical Analyses Using R; 206.

Appendix 1 – Clinical Trials selected

Authors	Title	Date and Journal of publication
Bramwell VH, et al.	A randomized placebo-controlled study of tamoxifen after adjuvant chemotherapy in premenopausal women with early breast cancer (National Cancer Institute of Canada--Clinical Trials Group Trial, MA.12).	Annals of Oncology 21: 283–290, 2010
Sirohi B, et al.	A randomized comparative trial of infusional ECisF versus conventional FEC as adjuvant chemotherapy in early breast cancer: the TRAFIC trial.	Annals of Oncology 21: 1623–1629, 2010
Blohmer JU, et al.	Epirubicin and cyclophosphamide versus epirubicin and docetaxel as first-line therapy for women with metastatic breast cancer: final results of a randomized phase III trial	Annals of Oncology 21: 1430–1435, 2010
Joensuu H, et al.	Docetaxel versus docetaxel alternating with gemcitabine as treatments of advanced breast cancer: final analysis of a randomized trial.	Annals of Oncology 21: 968–973, 2010
Eidtmann H, et al.	Efficacy of zoledronic acid in postmenopausal women with early breast cancer receiving adjuvant letrozole: 36-month results of the ZO-FAST Study	Annals of Oncology 21: 2188–2194, 2010
Frasci G, et al.	Preoperative weekly cisplatin, epirubicin, and paclitaxel (PET) improves prognosis in locally advanced breast cancer patients: an update of the Southern Italy Cooperative Oncology Group (SICOG) randomized trial 9908	Annals of Oncology 21: 707–716, 2010
Seidman AD,et al.	Phase III trial of gemcitabine plus docetaxel versus capecitabine plus docetaxel with planned crossover to the alternate single agent in metastatic breast cancer	Annals of Oncology 22:1094-101, 2011
Mavroudis D, et al.	Randomized phase III trial comparing docetaxel plus epirubicin versus docetaxel plus capecitabine as first-line treatment in women with advanced breast cancer	Annals of Oncology 21: 48–54, 2010
Canavese G,et al.	Sentinel node biopsy compared with complete axillary dissection for staging early breast cancer with clinically negative lymph nodes: results of randomized trial	Annals of Oncology 20: 1001–1007, 2009
Katsumata N, et al.	Phase III trial of doxorubicin plus cyclophosphamide (AC), docetaxel, and alternating AC and docetaxel as front-line chemotherapy for metastatic breast cancer: Japan Clinical Oncology Group trial (JCOG9802)	Annals of Oncology 20: 1210–1215, 2009
Stopeck AT, et al.	Denosumab compared with zoledronic acid for the treatment of bone metastases in patients with advanced breast cancer: a randomized, double-blind study	J Clin Oncol. 10;28(35):5132-9, 2010
Moebus V,et al.	Intense dose-dense sequential chemotherapy with epirubicin, paclitaxel, and cyclophosphamide compared with conventionally scheduled chemotherapy in high-risk primary breast cancer: mature results of an AGO phase III study	J Clin Oncol. 10;28(17):2874-80, 2010
Sparano JA, et al.	Randomized phase III trial of ixabepilone plus capecitabine versus capecitabine in patients with metastatic breast cancer previously treated with an anthracycline and a taxane	J Clin Oncol. Jul 10;28(20):3256-63, 2010
Di Leo A, et al.	Results of the CONFIRM phase III trial comparing fulvestrant 250 mg with fulvestrant 500 mg in postmenopausal women with estrogen receptor-positive advanced breast cancer	J Clin Oncol. 20;28(30):4594-600, 2010
Wardley AM, et al.	Randomized phase II trial of first-line trastuzumab plus docetaxel and capecitabine compared with trastuzumab plus docetaxel in HER2-positive metastatic breast cancer	J Clin Oncol. 20;28(6):976-83. 2010

Appendix 1 - cont

Authors	Title	Date and Journal of publication
Blackwell KL, et al.	Randomized study of Lapatinib alone or in combination with trastuzumab in women with ErbB2-positive, trastuzumab-refractory metastatic breast cancer	J Clin Oncol. 1;28(7):1124-30, 2010
Martín M, et al.	Adjuvant docetaxel for high-risk, node-negative breast cancer	N Engl J Med. 2;363(23):2200-10, 2010
Swain SM, et al.	Longer therapy, iatrogenic amenorrhea, and survival in early breast cancer	N Engl J Med, 3;362(22):2053-65, 2010
Hyman B. Muss, et al.	Adjuvant Chemotherapy in Older Women with Early-Stage Breast Cancer	N Engl J Med 2009; 360:2055-2065, 2009
Katsumata N, et al.	Phase III trial of doxorubicin plus cyclophosphamide (AC), docetaxel, and alternating AC and docetaxel as front-line chemotherapy for metastatic breast cancer: Japan Clinical Oncology Group trial	Ann Oncol. Jul;20(7):1210-5, 2009
Gnant M, et al.	Adjuvant endocrine therapy plus zoledronic acid in premenopausal women with early-stage breast cancer: 62-month follow-up from the ABCSG-12 randomized trial	N Engl J Med 2009;360:679-91
Joensuu H,et al.	Fluorouracil, epirubicin, and cyclophosphamide with either docetaxel or vinorelbine, with or without trastuzumab, as adjuvant treatments of breast cancer: final results of the FinHer Trial	J Clin Oncol. 1;27(34):5685-92, 2009
Gray R, et al.	Independent Review of *E2*100: A Phase III Trial of Bevacizumab Plus Paclitaxel Versus Paclitaxel in Women With Metastatic Breast Cancer	J Clin Oncol. 20;27(30):4966-72, 2009
Johnston S, et al.	Lapatinib combined with letrozole versus letrozole and placebo as first-line therapy for postmenopausal hormone receptor-positive metastatic breast cancer	J Clin Oncol. 2009 20;27(33):5538-46
de Azambuja E, et al.	Long-term benefit of high-dose epirubicin in adjuvant chemotherapy for node-positive breast cancer: 15-year efficacy results of the Belgian multicentre study	J Clin Oncol. 2009 10;27(5):720-5
Chan S,et al.	Phase III study of gemcitabine plus docetaxel compared with capecitabine plus docetaxel for anthracycline-pretreated patients with metastatic breast cancer	J Clin Oncol. 2009 10;27(11):1753-60
Holli K, et al.	Radiotherapy after segmental resection of breast cancer with favorable prognostic features: 12-year follow-up results of a randomized trial	J Clin Oncol. 2009 20;27(6):927-32
Gradishar WJ, et al.	Significantly Longer Progression-Free Survival With nab-Paclitaxel Compared With Docetaxel As First-Line Therapy for Metastatic Breast Cancer	J Clin Oncol. 2009 1;27(22):3611-9
Spielmann M, et al.	Trastuzumab for Patients With Axillary-Node–Positive Breast Cancer: Results of the FNCLCC-PACS 04 Trial	J Clin Oncol. 2009 20;27(36):6129-34
Kaufman B, et al.	Trastuzumab Plus Anastrozole Versus Anastrozole Alone for the Treatment of Postmenopausal Women With Human Epidermal Growth Factor Receptor 2–Positive, Hormone Receptor–Positive Metastatic Breast Cancer: Results From the Randomized Phase III TAnDEM Study	J Clin Oncol. 20;27(33):5529-37, 2009
BIG 1-98 Collaborative Group	Letrozole therapy alone or in sequence with tamoxifen in women with breast cancer	N Engl J Med. 2009 20;361(8):766-76
Whelan TJ, et al.	Long-term results of hypofractionated radiation therapy for breast cancer	N Engl J Med. 2010 11;362(6):513-20

Appendix 2 –Selected clinical trials data used

ID	phase	stage	N	primary	*P1*	*P2*	P3	*HR1*	*HR2*	HR3	p-v0	p-v1	p-v2	p-v3
2	3	0	672	1	0,28	NA	0,36	0,78	0,77	NA	NA	0,12	NA	0,056
3	3	0	349	2	0,34	0,40	NA	0,78	0,84	NA	NA	0,19	0,33	NA
4	3	1	236	0	0,57	NA	0,73	0,663	NA	0,98	0,63	0,21	NA	0,38
5	3	1	237	0	NA	0,74	NA	1,11	1,05	NA	0,31	0,6	0,72	NA
6	NA	0	1064	0	NA	NA	0,08	NA	NA	0,59	NA	NA	NA	0,0314
7	NA	1	200	2	0,41	0,56	NA	NA	NA	NA	NA	0,07	0,11	NA
9	3	1	475	2	NA	NA	NA	1,031	1,101	NA	NA	0,785	0,39	NA
10	3	1	272	2	0,48	0,79	NA	NA	NA	NA	NA	0,744	0,735	NA
13	NA	0	248	3	0,05	NA	0,09	0,95	1	NA	NA	0,679	NA	0,715
14	3	1	293	0	0,86	NA	NA	NA	NA	NA	0,13	0,09	NA	NA
15	3	1	2046	2	0,17	0,12	NA	NA	NA	NA	NA	0,49	0,93	NA
16	3	0	1284	3	0,22	NA	0,36	0,76	NA	0,72	NA	0,029	NA	0
18	3	1	1221	1	0,74	NA	0,50	0,9	NA	0,79	NA	0,231	NA	0,0005
19	3	1	736	3	NA	NA	0,86	0,84	NA	0,8	NA	0,091	NA	0,006
27	2	1	222	0	0,66	NA	NA	NA	0,704	0,725	0,717	NA	0,033	0,0449
28	3	1	296	3	0,48	NA	0,86	0,75	NA	0,73	NA	0,106	NA	0,008
29	3	0	1060	3	0,07	NA	0,18	0,76	NA	0.68	NA	0,29	NA	0,01
30	3	0	3506	1	0,16	NA	0,27	0,83	NA	0,8	NA	0,03	NA	0,001
31	NA	NA	633	3	0,07	NA	0,11	1,85	NA	2,09	NA	0,02	NA	0,001
32	3	1	290	0	0,80	NA	NA	NA	NA	NA	0,09	0,13	NA	NA
33	NA	0	1803	3	0,03	0,09	0,12	0,6	0,65	0,64	NA	0,11	0,01	0,01
34	3	0	1009	3	0,11	NA	0,48	0,8	NA	0,66	NA	0,086	NA	0,01
35	3	1	722	3	NA	NA	0,52	NA	NA	0,483	NA	NA	NA	0
37	3	1	219	3	NA	NA	0,98	0,74	NA	0,71	NA	0,113	NA	0,019

Appendix 2 –Selected clinical trials data used (cont)

ID	phase	stage	N	primary	*P1*	*P2*	P3	*HR1*	*HR2*	HR3	p-v0	p-v1	p-v2	p-v3
38	3	0	522	3	0,37	NA	NA	0,85	NA	0,75	NA	0,26	NA	0,02
40	3	1	305	3	0,76	NA	0,93	NA	NA	1,2	NA	0,983	NA	0,121
43	NA	0	264	3	0,21	NA	0,27	0,63	NA	0,36	NA	0,11	NA	0,00071
44	2	1	148	0	NA	NA	0,45	NA	NA	0,495	NA	NA	NA	0,0065
45	NA	0	528	3	0,07	NA	0,26	1,27	NA	0,86	NA	NA	NA	0,41
46	3	1	207	3	0,62	0,88	0,95	NA	NA	0,63	NA	0,325	0,0007	0,0016
53	3	0	3094	3	0,09	0,08	0,16	1,13	1,22	1,05	NA	NA	NA	NA
54	NA	0	1234	2	0,21	0,07	NA	NA	NA	NA	NA	0,79	0,01	NA

ID: order number given to each selected trial.

Phase: clinical trial development phase.

Stage: 0 for early disease and 1 for advanced breast cancer.

Primary: coding if the primary endpoint was 1 = death; 2 = progression; 3 = composite endpoint.

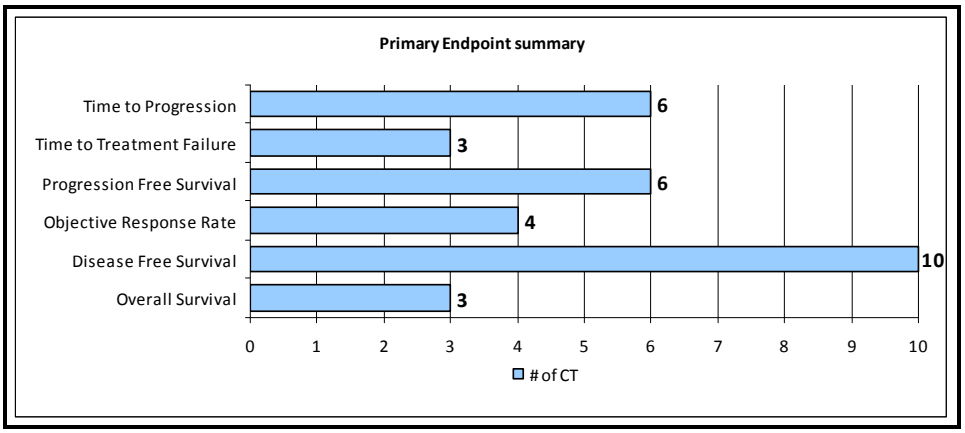
P: probability of the event in the control group depending on the type of endpoint: 1 = death; 2 = progression; 3 = composite endpoint.

HR: hazard ratio in the control group depending on the type of error coded as above

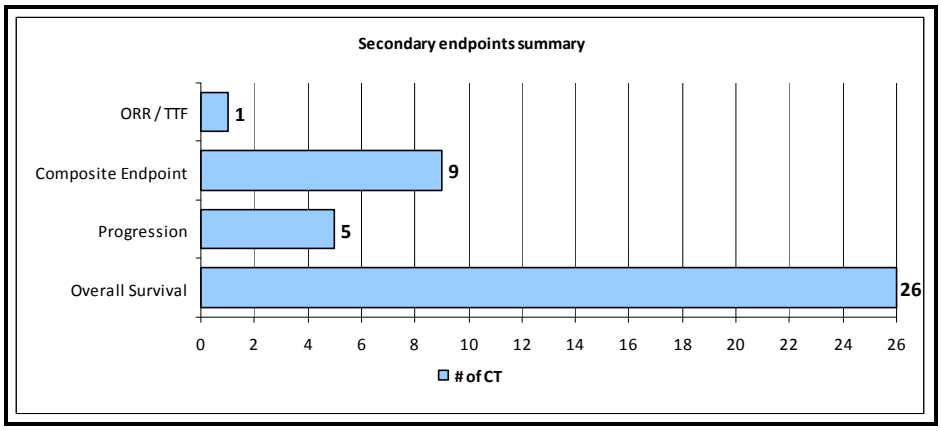
p-v : p-value of each endpoint, coded as above adding code 0, where primary endpoint was not a survival one.

Appendix 3 – Clinical Trials selected summary

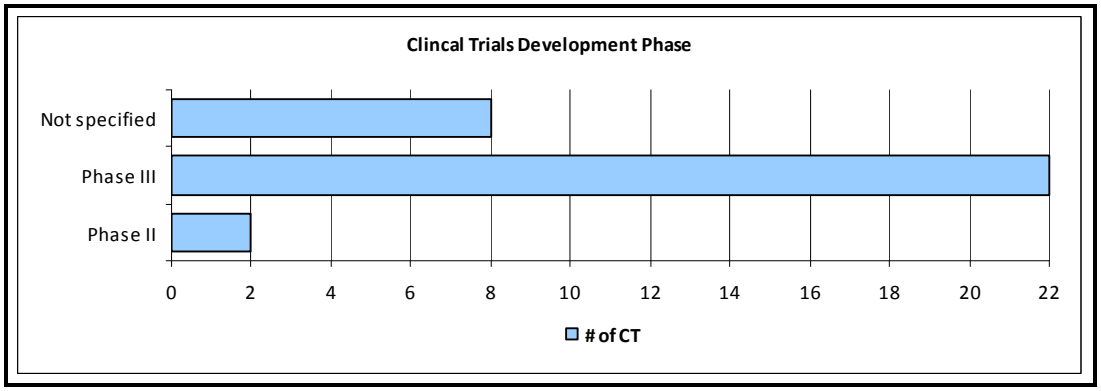
A. Primary Endpoint of the selected Clinical Trials:



B. Secondary Endpoints of the selected clinical Trials



C. Clinical Trials Development Phase



Appendix 4 – R Programming Code Used

```
##################################
#      Descriptive
##################################

#      sample size
median(ce$N)
quantile(ce$N,0.25)
quantile(ce$N,0.75)
max(ce$N)
min(ce$N)

#      P0
median(ce$P0,na.rm="T")
quantile(ce$P0,0.25,na.rm="T")
quantile(ce$P0,0.75,na.rm="T")
max(ce$P0,na.rm="T")
min(ce$P0,na.rm="T")

##################################
#      Rmatrix.txt complete dataset pooled
##################################

ce<-read.table("Rmatrix.txt", header=T)
summary(ce)
str(ce)

ce$ID<-as.factor(ce$ID)
ce$Priority<-factor(ce$Priority,
       levels = c("P","S"),
       labels = c("Primary", "Secondary"))

ce$Phase<-factor(ce$Phase,
       levels = c(2,3),
       labels = c("fase II", "fase III"))
ce$Stage<-factor(ce$Stage,
       levels = c("0","1"),
       labels = c("Early", "Advanced"))


######################
Graphs# plot P0 and HR. [with CE.txt data]
######################

par(mfrow=c(2,2),font=2,font.lab=4,font.ax
is=2,las=1)

boxplot(ce$P0~ce$Stage,xlab='Disease
stage',ylab="P0 distribution", lwd=2)
title("P0 distribution per Disease Stage")

boxplot(ce$P0~ce$Phase,xlab='Trial
phase',ylab="P0 distribution", lwd=2)
title("P0 distribution per Trial phase")

boxplot(ce$HR~ce$Stage,xlab='Disease
stage',ylab="HR distribution", lwd=2)
title("HR distribution per Disease Stage")

boxplot(ce$HR~ce$Phase,xlab='Trial
phase',ylab="HR distribution", lwd=2)
title("HR distribution per Trial phase")


##################################
#Rmatrix2 data.
#differentiates  what  kind  of  primary
endpoint, as well as its Ps, HRs and p-
values
##################################

cE2<-read.table("Rmatrix2.txt", header=T)

cE2$ID<-as.factor(cE2$ID)
cE2$phase<-factor(cE2$phase,
       levels = c(2,3),
       labels = c("fase II", "fase III"))
cE2$stage<-factor(cE2$stage,
       levels = c(0,1),
       labels = c("Early", "Advanced"))
cE2$primary<-factor(cE2$primary,
       levels = c(0,1,2,3),
       labels  =  c("ORR/TTF",  "Death",
"Progression", "Composite Endpoint"))

summary(cE2)
str(cE2)

######################
##  Graphs  #
######################

par(mfrow=c(3,2),font=2,font.lab=4,font.ax
is=2)

plot(cE2$P1,             col=c(cE2$stage),
xlim=c(0,35), ylim=c(0,1), lwd=1, ylab="P1
distr", xlab=" Selected clinical trials")
       title("Death primary endpoint")

boxplot(cE2$P1~cE2$stage,xlab='Disease
stage',ylab="P1             distribution",
ylim=c(0,1), lwd=1)
       title("Death primary endpoint")

plot(cE2$P2,             col=c(cE2$stage),
xlim=c(0,35),  ylim=c(0,1),  lwd=1,ylab="P2
distr",xlab=" Selected clinical trials")
       title("Progression    as    primary
endpoint")
boxplot(cE2$P2~cE2$stage,xlab='Disease
stage',ylim=c(0,1),              ylab="P2
distribution", lwd=1)
       title("Progression    as    primary
endpoint")

plot(cE2$P03,   col=c(cE2$stage),   lwd=1,
xlim=c(0,35),    ylim=c(0,1),    ylab="P03
distr", xlab=" Selected clinical trials")
       title("Composite    Endpoint    as
primary endpoint")
boxplot(cE2$P03~cE2$stage,xlab='Disease
stage',ylim=c(0,1),              ylab="P03
distribution", lwd=1)
       title("Composite    Endpoint    as
primary endpoint")


##################################
#p_hr distributions # to detect potential
clusters
##################################

#      complete dataset
plot(ce$P0~ce$HR,col=1,   lwd=2,   xlab="HR
values", ylab="P0 values")
       title("P0  and  HR  distr  -  Pooled
data")

#      Ps and HRs separate by stage (all
endpoints mixed):

par(mfrow=c(2,1),font=2,font.lab=4,font.ax
is=2,las=4)
plot(ce$P0~ce$HR,col=1,   lwd=2,   xlab="HR
values",         ylab="P0         values",
data=(ce$Stage=0))
       title("P0  and  HR  distr  -  Disease
Stage 0")
```

```
plot(ce$P0~ce$HR,col=2,   lwd=2,   xlab="HR
values",         ylab="P0         values",
data=(ce$Stage=0))
       title("P0  and  HR  distr  -  Disease
Stage 1")

##########################################
##     P and HR correlation per stage and
type of endpoint and stage
##########################################

phr0<-read.table("st0p_hr.txt", header=T)
pHR1<-read.table("st1p_hr.txt", header=T)

par(mfrow=c(2,2),font=2,font.lab=4,font.ax
is=2,las=4)

#stage 0 - P1_HR1
plot(phr0$P1~phr0$HR1,col=1,
ylim=c(0,1),lwd=2,    xlab="HR1    values",
ylab="P1 values", data=(phr0))
       title("P1 and HR1 distr - Disease
Stage 0")
#stage 1 - P1_HR1
plot(pHR1$P1~pHR1$HR1,col=1,
ylim=c(0,1),lwd=2,    xlab="HR1    values",
ylab="P1 values", data=(pHR1))
       title("P1 and HR1 distr - Disease
Stage 1")

#stage 0 - P2_HR2
plot(phr0$P2~phr0$HR2,col=1,  ylim=c(0,1),
lwd=2,    xlab="HR2    values",    ylab="P2
values", data=(phr0))
       title("P2 and HR2 distr - Disease
Stage 0")
#stage 1 - P2_HR2
plot(pHR1$P2~pHR1$HR2,col=1,
ylim=c(0,1),lwd=2,    xlab="HR2    values",
ylab="P2 values", data=(pHR1))
       title("P2 and HR2 distr - Disease
Stage 1")

##########################################
#### (HRs limited to 0.99)

phr0b<-read.table("st0p_hr_b.txt",
header=T)
pHR1b<-read.table("st1p_hr_b.txt",
header=T)

par(mfrow=c(2,2),font=2,font.lab=4,font.ax
is=2,las=3)

#stage 0 - P1_HR1
plot(phr0b$P1~phr0b$HR1,col=1,       lwd=2,
xlab="HR1   values",   ylab="P1   values",
data=(phr0b))
       title("P1 and HR1 distr - Disease
Stage 0")
#stage 1 - P1_HR1
plot(pHR1b$P1~pHR1b$HR1,col=1,       lwd=2,
xlab="HR1   values",   ylab="P1   values",
data=(pHR1b))
       title("P1 and HR1 distr - Disease
Stage 1")

#stage 0 - P2_HR2
plot(phr0b$P2~phr0b$HR2,col=1,       lwd=2,
xlab="HR2   values",   ylab="P2   values",
data=(phr0b))
       title("P2 and HR2 distr - Disease
Stage 0")
#stage 1 - P2_HR2
plot(pHR1b$P2~pHR1b$HR2,col=1,       lwd=2,
xlab="HR2   values",   ylab="P2   values",
data=(pHR1b))
       title("P2 and HR2 distr - Disease
Stage 1")


##########################################
 as we are only using HRs until 0.99, this
data has all HR>1 turned into 0.99
### only for stage 0


## Clustering

library(cluster)

s0P1HR1b<-
read.table("s0_P1_HR1_b.txt",header=T)
s0P1HR1b<-  na.omit(s0P1HR1b)  #  listwise
deletion of missing

## K-Means Cluster Analysis

# 2 cluster solution
fit1 <- kmeans(s0P1HR1b, 2)
 # get cluster means
aggregate(s0P1HR1b,by=list(fit1$cluster),F
UN=mean)
 # append cluster assignment
s0P1HR1c1     <-     data.frame(s0P1HR1b,
fit1$cluster)

# 3 cluster solution
fit2 <- kmeans(s0P1HR1b, 3)
 # get cluster means
aggregate(s0P1HR1b,by=list(fit2$cluster),F
UN=mean)
 # append cluster assignment
s0P1HR1c2     <-     data.frame(s0P1HR1b,
fit2$cluster)

par(mfrow=c(1,2),font=2,font.lab=4,font.ax
is=2,las=4)
clusplot(s0P1HR1c1,          fit1$cluster,
color=TRUE, shade=TRUE,
   labels=2,  lines=0,  main="d.stage0  P1
HR1 - 2 clusters")
clusplot(s0P1HR1c2,          fit2$cluster,
color=TRUE, shade=TRUE,
   labels=2,  lines=0,  main="d.stage0  P1
HR1 - 3 clusters")

# comparing 2 cluster solutions
 library(fpc)
d <- dist(s0P1HR1b)
cluster.stats(d,             fit1$cluster,
fit2$cluster)

cluster(s0P1HR1b)

##################################
### DATA CLUSTERING
##################################

# seen P0 and HR distribution, clustering
will only be performed for DS0 P1 HR1

library(cluster)

s0P1HR1<-
read.table("s0_P1_HR1.txt",header=T)
s0P1HR1<-   na.omit(s0P1HR1)   #   listwise
deletion of missing

phr0b<-read.table("st0p_hr_b.txt",
header=T)


## K-Means Cluster Analysis
```

```
# 2 cluster solution
fit1 <- kmeans(s0P1HR1, 2)
 # get cluster means
aggregate(s0P1HR1,by=list(fit1$cluster),FU
N=mean)
 # append cluster assignment
s0P1HR1c1      <-      data.frame(s0P1HR1,
fit1$cluster)

# 3 cluster solution
fit2 <- kmeans(s0P1HR1, 3)
 # get cluster means
aggregate(s0P1HR1,by=list(fit2$cluster),FU
N=mean)
 # append cluster assignment
s0P1HR1c2      <-      data.frame(s0P1HR1,
fit2$cluster)

par(mfrow=c(1,2),font=2,font.lab=4,font.ax
is=2,las=4)
clusplot(s0P1HR1c1,          fit1$cluster,
color=TRUE, shade=TRUE,
   labels=2,  lines=0,  main="d.stage0  P1
HR1 - 2 clusters")
clusplot(s0P1HR1c2,          fit2$cluster,
color=TRUE, shade=TRUE,
   labels=2,  lines=0,  main="d.stage0  P1
HR1 - 3 clusters")

# comparing 2 cluster solutions
 library(fpc)
d <- dist(s0P1HR1)
cluster.stats(d,             fit1$cluster,
fit2$cluster)

## WE CHOOSE THE 3 CLUSTERS OPTION

########################################
### ARE CALCULATION FOR STAGE 0
### using Moisés Gómez Mateu R programming
########################################

setwd("H:\\MEIO\\TFM      -      composite
endpoints\\R\\ARE calc\\stage 0\\c")

library(reshape)

########################################
# WE SET THE PARAMETER VALUES IN A DATASET
########################################

dataset<- c(0)
dataset<-as.data.frame(dataset)

#values  for  stage  0  from  the  bibl.
research
beta1V <- c(0.5,1,2) # constant risk or
increasing
beta2V <- c(0.5,1,2)
P1V<- c(0.07, 0.330, 0.133)
P2V<- c(0.07,0.16,0.22)
HR1V<-c(0.730, 0.803,0.976)
HR2V<-c(0.65,0.84,0.99)
rhoV<-c(0.1500,  0.4500) #low  and  medium
correlation

###################################
# WE CREATE THE DATASET
###################################
row<-1

for (beta1X in 1:length(beta1V)){
for (beta2X in 1:length(beta2V)){
for (P1X in 1:length(P1V)){
for (P2X in 1:length(P2V)){
for (HR1X in 1:length(HR1V)){
for (HR2X in 1:length(HR2V)){
for (rhoX in 1:length(rhoV)){
       dataset[row,1]<-beta1V[beta1X]
       dataset[row,2]<-beta2V[beta2X]
       dataset[row,3]<-P1V[P1X]
       dataset[row,4]<-P2V[P2X]
       dataset[row,5]<-HR1V[HR1X]
       dataset[row,6]<-HR2V[HR2X]
       dataset[row,7]<-rhoV[rhoX]
       row<-row+1
}}}}}}}


# WE RENAME THE VARIABLES

dataset<-                  rename(dataset,
c(dataset='beta1',V2='beta2',V3='P1',   V4
='P2',
 V5 ='HR1', V6 = 'HR2',    V7='rho'))

View(dataset)


###################################
#  WE  SAVE  THE  DATA  IN  A  .CSV2  FILE
(SEPARATED BY SEMICOLON(;) AND COMMA AS
DECIMALS)
########################################
write.csv2(dataset,'stage_0_dataset_c.csv2
')
#####################################
# WE READ THE DATA
########################################


dataset2<-
read.csv2("stage_0_dataset_c.csv2", header
= TRUE, sep = ";", quote="\"", dec=",")

dataset2<-dataset2[,2:8]  #  only  if  it
creates an extra column

summary(dataset2)
View(dataset2)


##  INTENTO  FER  EL  CLUSTER  AMB  LES
INSTRUCCIONS DEL MOISES DE MÉS ABAIX

dataset2<-subset(dataset2,(
((P1== 0.07   )&(HR1==      0.976  ))|
((P1== 0.33   )&(HR1==      0.803  ))|
((P1== 0.133  )&(HR1==      0.730  ))))

# en principi només ha creuat les P1 y HR1
sel.leccionades i crec que surt bé

View(dataset2)
dataset2<-dataset2[,2:8]  #  only  if  it
creates an extra column

########################################
#    ITERATION WITH THE PATTERN VALUES
########################################

ncol<-ncol(dataset2)
ncol

data_ARE<- c(0)
data_ARE<-as.data.frame(data_ARE)

#is.data.frame(data_ARE)
#data_ARE

#dataset2[,8]=0
View(dataset2)
#for (i in 3580:3590) {
```

```
for (i in 1:nrow(dataset2)) {

# IMPUTING VALUES TO THE PARAMETERS

beta1 <- dataset2[i,1]
beta2 <- dataset2[i,2]
P1 <- dataset2[i,3]
P2 <- dataset2[i,4]
HR1 <-dataset2[i,5]
HR2 <- dataset2[i,6]
rho <- dataset2[i,7]

input <- c(beta1, beta2, P1, P2, HR1, HR2,rho)

########################################
# Functions: f, F
########################################
# Description: It computes Theta from the values of Rho
#
# Arguments:
# t                   Time to event Random variable.
# theta        We calculate theta from the values of Rho
# Rho                 Spearman's coefficient that we set
########################################

f <- function(t,theta) {
  num <- t*theta-2*t^2
  den <- theta*(exp(t)-1)
  return(num/den)
}

F <- function(theta,rho) {
  inte     <-     integrate(f,     lower=0, upper=theta,                  theta=theta, subdivisions=1000)$value
  return(1- 12*inte/theta^2 - rho)
}

limits <- c(0.00001,1000)
theta   <-   uniroot(F,   interval=limits, rho=rho)$root
#print(theta)

####### ASSESSMENT OF THE SCALE PARAMETER VALUES b10, b11, b20, b21

b10 <- 1/((-log(1-P1))^(1/(beta1)))
b11 <- b10/HR1^(1/beta1)
# b20 <- 1/(-log(1-P2))^(1/beta2)

########################################
# Function: Fb20
#####################################
#  Description:  It  computes  b20  values (What is different from the CASE 1)
#
# Arguments:
# b20
#  P2       Probability  of  observing  the additional endpoint
########################################

Fb20<-function(b20,P2){
integral<-integrate(function(y) {
sapply(y, function(y) {
integrate(function(x) (      (theta*(1-exp(-theta))*exp(-theta*(x+y)))/    (exp(-theta)+     exp(-theta*(x+y))     -   exp(-theta*x)-exp(-theta*y))^2
                       )        ,  lower=0, upper=exp(     -(    (((-log(y))^(1/beta2)) * b20) /b10)^beta1    )       )$value
})
},       lower=       exp(-(1/b20)^beta2), upper=1)$value
return(integral-P2)
}

#Fb20(0.00001,P2)
#Fb20(10000,P2)

limits <- c(0.00001,100000)
b20           <-           uniroot(Fb20, interval=limits,P2=P2)$root
#b20

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


########################################
# Function: numerador
#####################################
# Description: It computes the numerador of  the  ARE  expression,  with  Weibull distribution
# Arguments:
########################################

numerador                                  <- function(t,b10,b11,b20,b21,beta1,beta2,theta,fT10,fT11,fT20,fT21,

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



#numerador1<- integrate(numerador,lower=0,upper=1, b10,b11,b20,b21,beta1,beta2,
#
theta,fT10,fT11,fT20,fT21,ST10,ST11,ST20,ST21,Sstar0,fstar0,Lstar0,Sstar1,
```

```
#           fstar1,Lstar1,HRstar,logHRstar
,subdivisions=1000,      stop.on.error   =
FALSE)

check_numerador1<-try(numerador1<-
integrate(numerador,lower=0,upper=1,
b10,b11,b20,b21,beta1,beta2,

theta,fT10,fT11,fT20,fT21,ST10,ST11,ST20,S
T21,Sstar0,fstar0,Lstar0,Sstar1,
 fstar1,Lstar1,HRstar,logHRstar
,subdivisions=1000,      stop.on.error   =
FALSE),silent=TRUE)

check2_numerador1<-
check_numerador1[1]=="Error             in
integrate(numerador, lower = 0, upper = 1,
b10,  b11,  b20,  b21,   : \n   non-finite
function value\n"

if(check2_numerador1=="TRUE")
{dataset2[i,8] <- NA ; next(i)}

########################################
#using sums
#h<-seq(0.0001,1,by=0.0001)
#temp<-h*0
#for (i in 1:length(h) ) {
#temp[i]<-
0.0001*numerador(t=h[i],b10,b11,b20,b21,be
ta1,beta2,theta,fT10,fT11,fT20,fT21,
#
ST10,ST11,ST20,ST21,Sstar0,fstar0,Lstar0,S
star1,fstar1,Lstar1,HRstar,logHRstar)
#}
#numerador1<-sum(temp)
########################################

numerador2<-(numerador1$value)^2

ST10_1 <- exp(-(1/b10)^beta1)
ST20_1 <- exp(-(1/b20)^beta2)
Sstar0_1  <-  (-log(1+(exp(-theta*ST10_1)-
1)*(exp(-theta*ST20_1)-1)/(exp(-theta)-
1))/theta)
ST10_1 <- exp(-(1/b10)^beta1)

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

AREstarT <- (numerador2/denominador)


#  IF  THE  VALUE  THE  NUMERATOR  IS  NOT
COMPUTED, THEN WE ASSIGN A MISSING IN THE
ARE VALUE

if(numerador1$message!="OK") {AREstarT <-
NA}
dataset2[i,8] <- AREstarT

}

########################################
# END OF ITERATIONS
########################################

dataset2<- rename(dataset2, c(V8='ARE'))

View(dataset2)
summary(dataset2)

########################################
# SAVING DATA IN A .CSV2 FILE (SEPARATED
BY SEMICOLON(;) AND COMMA AS DECIMALS)

write.csv2(dataset2,'stage_0_ARE_c2.csv2')
```

```
########################################
####################################
# READING THE DATASET

dataset2<-read.csv2("stage_0_ARE_c.csv2",
header  =  TRUE,  sep  =  ";",  quote="\"",
dec=",")
View(dataset2)
dataset2<-dataset2[,2:9]

########################################
### ARE CALCULATION FOR STAGE 1
### using Moisés Gómez Mateu R programming

setwd("H:\\MEIO\\TFM      -      composite
endpoints\\R\\ARE calc\\stage 1\\d")


library(reshape)

########################################
# WE SET THE PARAMETER VALUES IN A DATASET
#
#   Note:   for   a   large   number   of
combinations, it is better to compute it
step by step.
# That is, with the larger vector, compute
the dataset for each value separatelly. It
is
# much faster.
#
########################################

dataset<- c(0)
dataset<-as.data.frame(dataset)

#values  for  stage  0  from  the  bibl.
research
beta1V <- c(1,2)
beta2V <- c(1,2)
P1V<- c(0.48, 0.615, 0.746)
P2V<- c(0.56,0.763,0.811)
HR1V<-c(0.663,0.840, 0.966)
HR2V<-c(0.704,0.80,0.90,0.99)
rhoV<-c(0.4500, 0.7500)

####################################
# WE CREATE THE DATASET
####################################
row<-1

for (beta1X in 1:length(beta1V)){
for (beta2X in 1:length(beta2V)){
for (P1X in 1:length(P1V)){
for (P2X in 1:length(P2V)){
for (HR1X in 1:length(HR1V)){
for (HR2X in 1:length(HR2V)){
for (rhoX in 1:length(rhoV)){
       dataset[row,1]<-beta1V[beta1X]
       dataset[row,2]<-beta2V[beta2X]
       dataset[row,3]<-P1V[P1X]
       dataset[row,4]<-P2V[P2X]
       dataset[row,5]<-HR1V[HR1X]
       dataset[row,6]<-HR2V[HR2X]
       dataset[row,7]<-rhoV[rhoX]
       row<-row+1
}}}}}}}


# WE RENAME THE VARIABLES

dataset<-                  rename(dataset,
c(dataset='beta1',V2='beta2',V3='P1',   V4
='P2',
 V5 ='HR1', V6 = 'HR2',    V7='rho'))

View(dataset)
```

```
########################################
# WE SAVE THE DATA IN A .CSV2 FILE (SEPARATED BY SEMICOLON(;) AND COMMA AS DECIMALS)
#####################################

write.csv2(dataset,'stage_1_dataset.csv2')

#####################################
# WE READ THE DATA
#####################################

dataset2<-
read.csv2("stage_1_dataset.csv2", header = TRUE, sep = ";", quote="\"", dec=",")

dataset2<-dataset2[,2:8]  #  only  if  it creates an extra column

summary(dataset2)
View(dataset2)

########################################
#    ITERATION WITH THE PATTERN VALUES
########################################

ncol<-ncol(dataset2)
ncol

data_ARE<- c(0)
data_ARE<-as.data.frame(data_ARE)

#is.data.frame(data_ARE)
#data_ARE

#dataset2[,8]=0
View(dataset2)
#for (i in 3580:3590) {

for (i in 1:nrow(dataset2)) {

# IMPUTING VALUES TO THE PARAMETERS

beta1 <- dataset2[i,1]
beta2 <- dataset2[i,2]
P1 <- dataset2[i,3]
P2 <- dataset2[i,4]
HR1 <-dataset2[i,5]
HR2 <- dataset2[i,6]
rho <- dataset2[i,7]

input <- c(beta1, beta2, P1, P2, HR1, HR2,rho)

########################################
# Functions: f, F
#####################################
# Descricption: It computes Theta from the values of Rho
#
# Arguments:
# t                 Time to event Random variable.
# theta       We calculate theta from the values of Rho
# Rho               Spearman's coefficient that we set
#####################################

f <- function(t,theta) {
  num <- t*theta-2*t^2
  den <- theta*(exp(t)-1)
  return(num/den)
}

F <- function(theta,rho) {
  inte    <-    integrate(f,    lower=0, upper=theta,                theta=theta, subdivisions=1000)$value
  return(1- 12*inte/theta^2 - rho)
}

limits <- c(0.00001,1000)
theta   <-   uniroot(F,   interval=limits, rho=rho)$root
#print(theta)

####### ASSESSMENT OF THE SCALE PARAMETER VALUES b10, b11, b20, b21

b10 <- 1/((-log(1-P1))^(1/(beta1)))
b11 <- b10/HR1^(1/beta1)
# b20 <- 1/(-log(1-P2))^(1/beta2)

########################################
# Function: Fb20
#  Description:  It  computes  b20  values (What is different from the CASE 1)
# Arguments:
# b20
#  P2       Probability  of  observing  the additional endpoint
#####################################

Fb20<-function(b20,P2){
integral<-integrate(function(y) {
sapply(y, function(y) {
integrate(function(x) (      (theta*(1-exp(-theta))*exp(-theta*(x+y)))/    (exp(-theta)+     exp(-theta*(x+y))     -   exp(-theta*x)-exp(-theta*y))^2
                       )        ,  lower=0, upper=exp(     -(    (((-log(y))^(1/beta2)) * b20) /b10)^beta1     )       )$value
})
},       lower=       exp(-(1/b20)^beta2), upper=1)$value
return(integral-P2)
}

#Fb20(0.00001,P2)
#Fb20(10000,P2)

limits <- c(0.00001,100000)
b20            <-            uniroot(Fb20, interval=limits,P2=P2)$root
#b20

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


# Function: numerador
#####################################
# Description: It computes the numerador of  the  ARE  expression,  with  Weibull distribution
# Arguments:
#####################################

numerador                                  <- function(t,b10,b11,b20,b21,beta1,beta2,theta,fT10,fT11,fT20,fT21,

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

```
  ST10 <- exp(-(t/b10)^beta1)
  ST11 <- exp(-(t/b11)^beta1)
  ST20 <- exp(-(t/b20)^beta2)
  ST21 <- exp(-(t/b21)^beta2)
  Sstar0   <-   (-log(1+(exp(-theta*ST10)-
1)*(exp(-theta*ST20)-1)/(exp(-theta)-
1))/theta)
  fstar0    <-    (exp(-theta*ST10)*(exp(-
theta*ST20)-1)*fT10+exp(-
theta*ST20)*(exp(-theta*ST10)-
1)*fT20)/(exp(-theta*Sstar0)*(exp(-theta)-
1))
  Lstar0 <- (fstar0/Sstar0)
  Sstar1   <-   (-log(1+(exp(-theta*ST11)-
1)*(exp(-theta*ST21)-1)/(exp(-theta)-
1))/theta)
  fstar1    <-    (exp(-theta*ST11)*(exp(-
theta*ST21)-1)*fT11+exp(-
theta*ST21)*(exp(-theta*ST11)-
1)*fT21)/(exp(-theta*Sstar1)*(exp(-theta)-
1))
  Lstar1 <- (fstar1/Sstar1)
  HRstar <- (Lstar1/Lstar0)
  logHRstar <- log(HRstar)
  return(logHRstar*fstar0)
}

# WITH BETAS = 0.5, THERE ARE SOME CASES
WHERE THE INTEGRAL DIVERGES AND STOPS
# THE LOOP. HENCE, I ADD "stop.on.error =
FALSE" INSTRUCTION.

#  AND  THERE  ARE  ALSO  SOME  COMBINATIONS
WHERE THE FUNCTION funcion() DOESN'T
# HAVE VALUES FOR CERTAIN VALUES OF T (EX.
T>0.9). HENCE, WE FIRST EVALUATE THE
# INTEGRATE BEFORE CONTINUING. IF THERE IS
ANY PROBLEM, THEN R ASSIGNS A MISSING
# TO THE ARE VALUE AND IT CONTINUES WITH
THE NEXT ITERATION OF THE LOOP.

#numerador1<-
integrate(numerador,lower=0,upper=1,
b10,b11,b20,b21,beta1,beta2,
#
theta,fT10,fT11,fT20,fT21,ST10,ST11,ST20,S
T21,Sstar0,fstar0,Lstar0,Sstar1,
#           fstar1,Lstar1,HRstar,logHRstar
,subdivisions=1000,      stop.on.error   =
FALSE)

check_numerador1<-try(numerador1<-
integrate(numerador,lower=0,upper=1,
b10,b11,b20,b21,beta1,beta2,

theta,fT10,fT11,fT20,fT21,ST10,ST11,ST20,S
T21,Sstar0,fstar0,Lstar0,Sstar1,
 fstar1,Lstar1,HRstar,logHRstar
,subdivisions=1000,      stop.on.error   =
FALSE),silent=TRUE)

check2_numerador1<-
check_numerador1[1]=="Error             in
integrate(numerador, lower = 0, upper = 1,
b10,  b11,  b20,  b21,   : \n   non-finite
function value\n"

if(check2_numerador1=="TRUE")
{dataset2[i,8] <- NA ; next(i)}

########################################
# Integrating using sums
#h<-seq(0.0001,1,by=0.0001)
#temp<-h*0
#for (i in 1:length(h) ) {
#temp[i]<-
0.0001*numerador(t=h[i],b10,b11,b20,b21,be
ta1,beta2,theta,fT10,fT11,fT20,fT21,
#
ST10,ST11,ST20,ST21,Sstar0,fstar0,Lstar0,S
star1,fstar1,Lstar1,HRstar,logHRstar)
#}
#numerador1<-sum(temp)
########################################

numerador2<-(numerador1$value)^2

ST10_1 <- exp(-(1/b10)^beta1)
ST20_1 <- exp(-(1/b20)^beta2)
Sstar0_1  <-  (-log(1+(exp(-theta*ST10_1)-
1)*(exp(-theta*ST20_1)-1)/(exp(-theta)-
1))/theta)
ST10_1 <- exp(-(1/b10)^beta1)

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

AREstarT <- (numerador2/denominador)

#  IF  THE  VALUE  THE  NUMERATOR  IS  NOT
COMPUTED, THEN WE ASSIGN A MISSING IN THE
ARE VALUE

if(numerador1$message!="OK") {AREstarT  <-
NA}
dataset2[i,8] <- AREstarT
}

########################################
# END OF ITERATIONS
########################################

dataset2<- rename(dataset2, c(V8='ARE'))

View(dataset2)
summary(dataset2)

#####################################
# SAVING DATA IN A .CSV2 FILE (SEPARATED
BY SEMICOLON(;) AND COMMA AS DECIMALS)

write.csv2(dataset2,'stage_1_ARE.csv2')


#####################################
# READING THE DATASET
dataset2<-read.csv2("stage_1_ARE.csv2",
header  =  TRUE,  sep  =  ";",  quote="\"",
dec=",")
View(dataset2)
dataset2<-dataset2[,2:9]
#####################################
```