The impact of the violation of the proportional hazards assumption on confirmatory analysis of survival data using delayed entry

Minkyu Kim

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2015

Reading Committee:

Susanne May

Robyn McClelland

Program Authorized to Offer Degree:
School of Public Health

University of Washington

Abstract

The impact of the violation of the proportional hazards assumption on confirmatory analysis of survival data using delayed entry

Minkyu Kim

In many clinical trials with staggered entry, interim analyses are conducted for monitoring purposes. If some unexpected results are found during the interim analyses, performing confirmatory analyses may be desirable. However, a completely new study with different patients in different locations requires a lot of additional resources such as time and money. Furthermore, there may be ethical concerns regarding recruitment of patients for a new study if a harmful treatment effect was found in a previous study. In this situation, the result illustrated in Keiding et al. (1987) may be helpful. It suggests use of recurrence-free survivors among the initial cohort as delayed entry when performing confirmatory analyses. Therefore, we can increase the sample size and power of confirmatory analyses by combining recurrence-free survivors with patients enrolled who have not contributed to the interim analyses. The result from Keiding et al. (1987) requires several assumptions but, our study focuses on the proportional hazards assumption.

In simulations the validity of the proportional hazards assumption depends on the way we generate the datasets and the models we fit. We simulate two binary covariates, Treatment and Status: one indicating either the treatment group or the control group and the other indicating whether patients developed their disease a long or short time before enrollment. Patients who have had their disease for a long time are called "Early patients" because, in our simulations, we assume that they are enrolled during the earlier part of hypothetical studies. Patients who developed the disease shortly before enrollment are called "Late patients" because we assume that they are enrolled during the later part of the hypothetical studies.

Therefore, there are four types of patients, early patients in either the treatment or the control group, and late patients in either the treatment or the control group. We use an interaction effect between the two factors and a change in patient mix (Early and Late patients) over time to create a time dependent treatment effect in a model that does not account for the interaction effect.

We focus on investigating the power of confirmatory analyses that follow an interim analysis which detected a significant harmful effect. We compare the power of confirmatory analyses that only use participants enrolled after the interim analysis to the power of confirmatory analyses that use participants enrolled after the interim analysis combined with recurrence-free survivors from the initial cohort. For recurrence-free survivors only the time of observation after the interim analysis is used and it is used as delayed entry. We fit three models: (1) a model that includes only a treatment effect, (2) a model that includes a treatment effect and a status effect and (3) a model that includes a treatment effect, a status effect and an interaction effect. The likelihood ratio test with one-degree of freedom is performed to test for the treatment effect in (1) and (2) and the likelihood ratio test with two degrees of freedom (main and interaction effect) is performed to test for a treatment effect in (3). For each of these three models we ensure (separately) that the model has a significant treatment effect at the interim analysis. We choose time frames of enrollment and follow-up and sample sizes that are similar to the setting of Keiding et al. (1987). We use exponentially distributed hazard functions for each of the four patient groups. For each of the three models, the treatment effect is estimated using four different analysis cohorts: (i) the initial cohort prior to the interim analysis only, (ii) the cohort that does not contribute to the interim analysis (only), (iii) the cohort that does not contribute to the interim analysis combined with recurrence-free survivors but each contributing only up to one year of data, (iv) the latter, but contributing up to four years of data.

For models 1 and 2, the power for the confirmatory analysis that uses the recurrence-free

survivors can be higher, about the same or lower than the power of the analysis that does not use them depending on the strength of the interaction effect. The use of recurrence-free survivors can typically not overcome issues of fitting an incorrect model. In contrast, when fitting model 3, the power of the analysis that includes the recurrence-free survivors is always higher than the power of the analysis not using the recurrence-free survivors regardless of the size of the interaction effect and the treatment effect.

As expected, in simulations where the proportional hazards assumption holds (either because there is no time trend or the time trend is included in the model) using recurrence-free survivors as delayed entry always improves power regardless of the size of the treatment and interaction effect.

TABLE OF CONTENTS

DEDICATION

This thesis is dedicated to my beloved wife who always trust and support me through thick and thin, my honorable parents who guide me along the right paths and my two passionate brothers who keep challenging themselves and achieving their goals.

Chapter 1

INTRODUCTION

In the real world, we face an unlimited number of scientific questions. To answer some of these scientific questions, we collect data, develop hypotheses and test them with the use of statistics. Conducting confirmatory analyses with well-designed procedures is a statistical technique to verify hypotheses.

In clinical trials with staggered entry, interim analyses are often performed to compare the treatment group to the control group at pre-planned time points before maximum enrollment is achieved. The trials can be terminated to protect subjects if the interim analyses show that the treatment effects are harmful or to not withhold a beneficial treatment from patients when the interim analyses show that the treatment effects are considered to be highly beneficial. However, often, trials reach maximum enrollment and results are confirmed by performing additional trials with independent patients.

Often, a confirmatory study uses different patients in different locations. However, if a study is stopped because interim analyses show a harmful effect there might be concerns regarding initiating a completely new study. At the time point where interim analysis are performed, there are usually patients who have been enrolled who are not contributing to such interim analysis, because there might not have been enough time to collect outcome data. Those patients can be used as an independent cohort, but the number of patients for whom this is the case is typically small. And if these patients are used for confirmatory analyses there might not be sufficient power to either support or reject the results from the interim analyses. In such a case, the result of Keiding et al. (1987) may be helpful. Keiding et al. (1987) point out that we can combine recurrence-free survivors among the initial cohort with the cohort that did not contribute to the interim analysis. The patients who are enrolled

prior to the interim analysis and who contribute observation time to the interim analysis are considered the initial cohort. The patients from the initial cohort who do not died or had recurrence of their cancer at the time of the interim analysis are considered recurrence-free survivors. In Keiding et al. (1987), recurrence-free survivors are patients who have had adjuvant treatment of breast cancer after primary mastectomy and do not experience either recurrences or deaths. The time of observation of the recurrence-free survivors that occurs after the time of observation that is included in the interim analysis is used as delayed entry in Keiding et al. (1987) when they are combined with the cohort that do not contribute to the interim analysis. Delayed entry refers to patients who come under observation only after they start being at risk. In the current setting, the patient from the initial cohort are under observation the for the entire time of the study, but their observation time is split into the time of observation that contributes to the interim analysis and the time of observation that does not contribute to the interim analysis. When using only the time period of observation that follows the interim analysis, then the starting time point for that observation time happens some time after the started being at risk (time of the mastectomy). In Keiding et al. (1987), the amount of the time that recurrence-free survivors contribute to the study as delayed entry can be obtained by subtracting the amount of time they spend before the interim analysis from the total amount of the time they are followed up. Suppose a recurrence-free survivor contributes four years to the study and interim analysis is performed one year after the survivor enters the study. In this situation, the survivor contributes three years to the confirmatory analysis. Using the observation time from recurrence free survivors that follows the interim analysis is efficient especially when we would otherwise not be able to enroll a sufficient number of patients for confirmatory analyses.

The Cox proportional hazards regression model is used in Keiding et al. (1987). The idea described in Keiding et al. (1987) relies on the proportional hazards assumption to hold. In Keiding et al. (1987), they describe a study where the investigators obtain a significant harmful treatment effect at the interim analysis, but the effect is not statistically significant at the confirmatory analysis that uses only patients who do not contribute to the interim analysis.

However, they find a statistically significant harmful treatment effect at the confirmatory analysis that combines recurrence-free survivors with the cohort that does not contribute to the interim analysis. For this analysis, the observation times of recurrence-free survivors following the interim analysis are used as delayed entry.

Chapter 2 illustrates the Cox proportional hazards model, and the background of Keiding et al. (1987). In Chapter 3 we describe how we designed the simulations, what models were fit, the choice of coefficients and how power was calculated for the confirmatory analyses. Chapter 4 provides the results of the simulations and plots showing the power of the three models depending on the strength of the interaction effect. Chapter 5 and 6 describe the limitations of this study and a discussion.

Chapter 2

BACKGROUND

2.1 The Cox proportional hazards model

Semi-parametric model

Keiding et al. (1987) used the Cox proportional hazards (PH) model to obtain an estimate of the treatment effect. Parner & Keiding (2001) showed that the idea from Keiding et al. (1987) can be applied to the Cox PH model, combining the observation time from recurrence-free survivors that follows the interim analysis (used as delayed entry) with the cohort that does not contribute to the interim analysis. Therefore, we will briefly discuss the Cox model.

In 1972, Cox proposed a model that is referred to as the *proportional hazards model* in Cox (1972). Cox also proposed a new method which is called *partial likelihood* to estimate parameters in the proportional hazards model. His model has become the most common method to analyze data with the outcome of time to event in health research. The model is written as

$$\lambda(t) = \lambda_0(t) \cdot \exp(\mathbf{X}_i\boldsymbol{\beta}) \tag{2.1}$$

where $\mathbf{X}_i$ indicates a row vector of covariates for individual i from an $n \times p$ matrix $\mathbf{X}$. n and p indicate the number of observations and covariates, respectively. $\boldsymbol{\beta}$ is a $p \times 1$ vector of coefficients.

In the Cox model, the hazard function is the product of $\lambda_0(t)$ and a function of the covariates. The function $\lambda_0(t)$ is called *baseline hazard* function, which is an unspecified non-negative function. It corresponds to the hazard for individuals who have all values of the covariates equal to zero. The Cox model is considered to be a semi-parametric model because no parametric assumptions are made about the shape of the baseline hazard function. Since

the model assumes that the hazard ratio between two groups of patients with different values of a covariate is constant over time, it is called the proportional hazards model. The hazard ratio comparing two groups of patients such as treatment and control groups is obtained such that

$$\begin{aligned} \text{HR(Treat vs Control)} &= \frac{\lambda(\text{t}, \text{X}_\text{T} = 1, \beta)}{\lambda(\text{t}, \text{X}_\text{T} = 0, \beta)} \\ &= \frac{\lambda_0(t) \cdot \exp(\beta \cdot 1)}{\lambda_0(t) \cdot \exp(\beta \cdot 0)} \\ &= \exp(\beta) \end{aligned} \tag{2.2}$$

where HR represents a hazard ratio and X_T is a binary variable indicating whether a patient is in either the treatment (X_T=1) or the control group (X_T=0).

The HR is the ratio of hazards of death or event occurrence comparing subjects in the treatment group to subjects in the control group. The baseline hazard function $\lambda_0(t)$ cancels out in (2.2). Because the remaining terms are independent of time, the hazard ratio is assumed to be constant over time. To estimate β in (2.2), the *partial likelihood* method is used.

Partial likelihood method

The partial likelihood(Cox (1975)) allows us to obtain estimates of the $\boldsymbol{\beta}$ coefficients without specifying the baseline hazard function $\lambda_0(t)$. We can write

$$\begin{aligned} \text{PL} &= \prod_{k=1}^{m} \frac{e^{\mathbf{X}_{(k)}^T \boldsymbol{\beta}}}{\sum_{l \in R(t_{(k)})} e^{\mathbf{X}_l^T \boldsymbol{\beta}}} \\ \text{LogPL} &= \sum_{k=1}^{m} \Big\{ \mathbf{X}_{(k)}^T \boldsymbol{\beta} - \ln \Big[\sum_{l \in R(t_{(k)})} e^{\mathbf{X}_l^T \boldsymbol{\beta}} \Big] \Big\} \end{aligned}$$

where m is the number of distinct ordered survival or event times and $t_{(k)}$ is the k^{th} ordered failure time and $\mathbf{X}_{(k)} = (x_{(k)1}, x_{(k)2}, \ldots, x_{(k)p})$ is a covariate vector for the subject who fails

at $t_{(k)}$. $R(t_{(k)})$ indicates the risk set at $t_{(k)}$.

In the partial likelihood (Cox (1975)) covariates of patients who fail at $t_{(k)}$ are compared to covariates of patients who survive beyond $t_{(k)}$. We can obtain $\hat{\boldsymbol{\beta}}$ by taking the derivative of LogPL with respect to $\boldsymbol{\beta}$ and set it equal to zero.

2.2 The idea of Keiding et al

Our simulation works are based on Keiding et al. (1987). Keiding et al. (1987) propose that we can combine the recurrence-free survivors among the initial cohort with the cohort from the confirmatory analysis if they are independent of each other. When they are combined, recurrence-free survivors will be used as delayed entry. Figure 2.1 illustrates this idea:

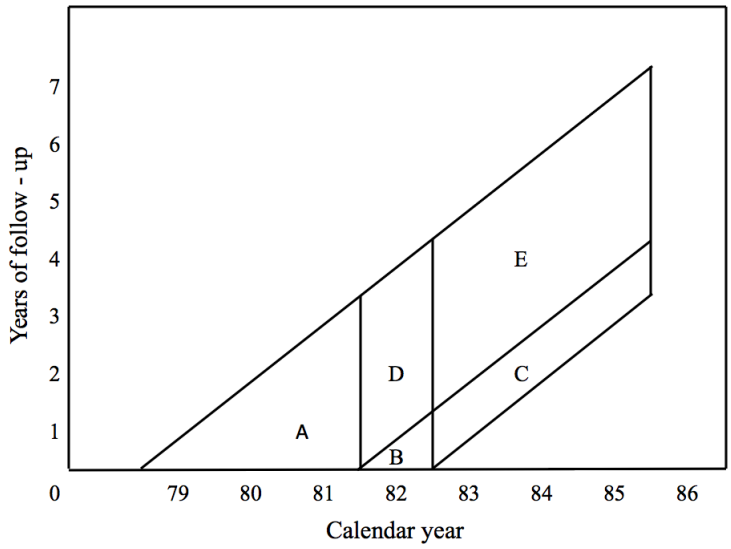


Figure 2.1: Lexis diagram of the Danish clinical trial, 1979 - 1985

This Figure will be referred to throughout this work as a graphical illustration of the different subsets of patients. Area A in Figure 2.1 includes the patients who are enrolled before and contribute to the interim analysis. Area B and C include the patients who do not contribute to the interim analysis and who are followed up for up to four years. Area D and

E include the recurrence-free survivors who are enrolled before the interim analysis and who contribute some observation times to area A and the interim analysis and who are followed for up to four years beyond the time period that they contribute to the interim analysis.

The idea in Keiding et al. (1987) was demonstrated using an example from a clinical trial conducted in Denmark. It was the clinical trial of adjuvant treatment of breast cancer after primary mastectomy. The clinical trial started in 1978 and an unexpected harmful prognostic effect of the presence of residual cancer tissue after diagnostic biopsy was found at the interim analysis (Figure 2.1: Area A), which was conducted in 1981.

Keiding et al. (1987) performed the confirmatory analysis after the interim analysis with the cohort that did not contribute any observation time to the interim analysis (Figure 2.1: Areas B and C). They were interested in checking the finding from the interim analysis by conducting the confirmatory analysis. The initial trial was designed to enroll patients for four years and they performed the interim analysis after three years from the beginning of the trial. Therefore, the amount of time for them to recruit new patients for the confirmatory analysis was only one year. Both this limited time and an ethical issue prevented researchers from enrolling enough patients for the confirmatory analysis (Figure 2.1: Area B).

Data from follow-up of four years of patients who did not contribute any observation time to the interim analysis (Areas B and C) did not provide sufficient evidence to either support or reject the finding from the interim analysis. However, the result was statistically significant when the combined cohort, the cohort that did not contribute to the interim analysis and recurrence-free survivors, were followed for up to four years (Figure 2.1: Areas B, C, D and E). It was also interesting that an analysis that included observation time for only up to one year remained statistically significant (Figure 2.1: Areas B and D).

Chapter 3

SIMULATION STUDY

3.1 Characteristics

First, we describe some characteristics of the four exponentially distributed hazard functions. Some characteristics are chosen based on the setting in Keiding et al. (1987) and others are used for generating a time trend. Characteristics used for our simulations are as follows:

- The hazard ratio comparing the treatment versus control group among those enrolled before the interim analysis time is larger than 1 as in Keiding et al. (1987)

- The proportion of early patients decreases over time, but that of late patients increases over time

- The hazard ratio comparing the treatment versus control group among early patients is larger than 1

The datasets we generate include two binary covariates: Treatment and Status. Treatment indicates either the treatment group (Trt=1) or control group (Trt=0), and Status indicates either late patients (Status=1) or early patients (Status=0). Patients who have had their disease for a long time are called "early patients" because we assume for this setting that they are more likely to be enrolled during the earlier part of the simulated studies. In contrast, patients who developed the disease shortly before enrollment are called "late patients" because they tend to be enrolled during the later part of the simulated studies. This terminology, early patients and late patients, will be used throughout this work. Thus, in the simulated studies, there are four groups of patients: early patients in either treatment or control group, and late patients in either treatment or control group.

We use the two covariates to generate a interaction effects depending on the values of four hazards and how the models are fit. When we set the hazard ratio comparing the treatment group to the control group among early patients to differ from the hazard ratio among late patients, the treatment effect changes over time if we do not adjust for the interaction term in a fitted model. Such a model with exhibiting non-proportionality, the extent of non-proportionality will depend on the strength of the interaction and the extent of the differences in patient mix (early and late patients) over time.

We assume that the early patients have higher hazards of death in the treatment group than late patients. We simulate the data such that most patients enrolled at the beginning in our hypothetical studies (simulations) are early patients and that the proportion of early patients decreases over time. At the time of the interim analysis the proportion of early patients is larger than the proportion of late patients.

3.2 Generate datasets

To simulate the datasets, we first choose specific values for the coefficients of the exponentially distributed failure times in the four groups: the hazard of death of early patients in the treatment group (λ_{ET}), of early patients in the control group (λ_{EC}), of late patients in the treatment group (λ_{LT}) and of late patients in the control group (λ_{LC}). Depending on the choice of coefficients and model, the treatment effect changes over time.

Each hazard can be re-parameterized to the $\exp(\beta)$ form because all hazards have to be greater than zero. With those four re-parameterized hazards, hazard ratios comparing the treatment to the control groups among the early patients and the late patients can be expressed as follows:

$$\text{HR}(\text{treatment vs control} \mid \text{early patients}) = \frac{\lambda_{ET}}{\lambda_{EC}} = \exp(\beta_{ET} - \beta_{EC})$$
$$\text{HR}(\text{treatment vs control} \mid \text{late patients}) = \frac{\lambda_{LT}}{\lambda_{LC}} = \exp(\beta_{LT} - \beta_{LC}).$$

We divide a total amount of time used for both the interim analysis (Figure 2.1: Area

A) and the confirmatory analysis (Figure 2.1: Areas B, C, D and E) into several sets of time intervals. In each time interval, the proportion of the early patients and the late patients varies. Therefore, in the setting where the two hazard ratios above are different, a time trend exists if we fit amodel that does not include an interaction term. In contrary, in the same setting, a residual time trend does not exist if we fit a model that does include an interaction term. Therefore, among the models we fit in our simulations, the proportional hazards assumption will hold when using the model that includes the time trend. Conversely, when using models that do not account for a time trend, the proportional hazards assumption will (theoretically) not hold. Nevertheless, the impact of such non-proportionality on the power to detect a treatment effect in confirmatory analyses will depend on the extent of the non-proportionality.

After choosing coefficients for the four hazards, we generate failure times for each group based on exponential distributions with the chosen coefficients.

$$Y_{ET} \sim \text{Exponential}(\exp(\beta_{ET}))$$
$$Y_{EC} \sim \text{Exponential}(\exp(\beta_{EC}))$$
$$Y_{LT} \sim \text{Exponential}(\exp(\beta_{LT}))$$
$$Y_{LC} \sim \text{Exponential}(\exp(\beta_{LC}))$$

where Y_{ET}, Y_{EC}, Y_{LT}, and Y_{LC} are the failure times for the treatment group of the early patients, the control group of the early patients, the treatment group of the late patients, and the control group of the late patients, respectively.

In our simulations, no patients are at risk at the very beginning of the study. We allow for each patient in our simulations to have different times at which they enter the study. Besides the coefficients of four hazards, there are a number of study characteristics that we need to choose to generate the datasets, such as sample sizes for the interim (Figure 2.1: Area A) and confirmatory analysis (Figure 2.1: Areas B and C) and the follow-up time for the interim (Figure 2.1: Area A) and confirmatory analysis (Figure 2.1: Areas B and C).

When we choose sample sizes and follow-up times for both the interim analysis and the confirmatory analysis to generate the datasets, we take into account Keiding et al. (1987). In Keiding et al. (1987), the amount of time used for accrual before the interim analysis is three years, and the number of patients who contribute to the interim analysis is around 1,200 (Figure 2.1: Area A). Two confirmatory analyses are conducted in Keiding et al. (1987). One is conducted using four years of observation time after the interim analysis with the recurrence-free survivors (Figure 2.1: Areas B, C, D and E). This analysis is the one of primary interest. Another analysis is performed using one year (only) of observation time after the interim analysis with the recurrence-free survivors (Figure 2.1: Areas B and D). The number of patients who contribute to the confirmatory analysis, but do not contribute any observation time to the interim analysis 300.

With the information above, we generate times of enrollment for the initial cohort and the cohort that does not contribute to the interim analysis, separately. The time points at which patients are enrolled into the two cohorts in our simulations are generated as follows:

$$I_p \sim \text{Uniform}(0, t_1) \qquad 1 \leq p \leq N_1$$
$$C_{p'} \sim \text{Uniform}(0, t_2) \qquad 1 \leq p' \leq N_2$$

where I_p is the time when the p^{th} patient among the initial cohort is enrolled and $C_{p'}$ is the time when the p'^{th} patient among the cohort that does not contribute to the interim analysis is enrolled. N_1 is the total number of patients enrolled before the interim analysis and N_2 is the total number of patients enrolled after the interim analysis. t_1 is the amount of time to enroll the initial cohort and t_2 is the amount of time to enroll the patients who do not contribute to the interim analysis.

After generating times for entry into the study for the two different cohorts, we divide the time period before the interim analysis into 10 intervals, and the period after the interim analysis into 3 intervals. The proportion of early patients and late patients is generated to be different in each interval. The variability of the proportions in each interval influences

(in conjunction with an interaction effect) the extent of non-proportionality of the hazard ratio comparing the treatment to control groups in models that do not include an interaction term.

The times for entry into the study in each interval are expressed as

$$I_{ij} = \{\, I_p \mid \frac{1}{10} \times t_1 \times (i-1) < \; I_p \; \leq \; \frac{1}{10} \times t_1 \times i\} \qquad 1 \leq i \leq 10,\; i \in \mathbb{N},\; j \leq \; p$$
$$C_{i'j'} = \{\, C_{p'} \mid \frac{1}{3} \times t_2 \times (i'-1) < \; C_{p'} \; \leq \; \frac{1}{3} \times t_2 \times i\} \qquad 1 \leq i' \leq 3,\; i' \in \mathbb{N},\; j' \leq \; p'$$

where I_{ij} is the time for the j^{th} patient in the i^{th} interval to enter the study before the interim analysis and $C_{i'j'}$ is the time for the j'^{th} patient in the i'^{th} interval to enter the study after the interim analysis.

Patients who are generated to be enrolled in these time intervals are generated to belong to one of two groups: early patients and late patients. We generate the data such that the proportion of the early patients decreases over time and the proportion of late patients increases over time.

Both, the hazard ratio comparing the treatment to the control group among early patients and the hazard ratio comparing the treatment to the control group among late patients stay constant during the trial. However, combining the two groups of patients (early and late patients) with different proportions at each time interval causes the hazard ratio comparing the treatment to the control groups to change over time if an interaction is generated but a model is fit without an interaction term. Figure 3.1 below illustrates the proportion of the two patient groups within each time interval.

The upper-left region indicates the proportion of the early patients and the lower-right region indicates the proportion of the late patients. We only have early patients in the first time interval in our simulation. However, the proportion of late patients increases over time. In Chapter 3.1, one of our characteristics when generating datasets is that a hazard ratio comparing the treatment group to the control group among early patients is larger than that among late patients. Therefore, the increase in the proportion of the late patients causes a

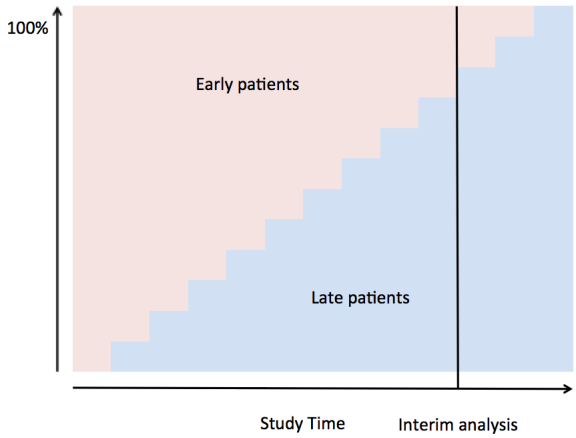


Figure 3.1: Changes of proportion of the early patients and the late patients over time

log hazard ratio comparing the treatment versus control group to decrease over time. Figure 3.2 below illustrates the log hazard ratios in each time interval when we do not adjust for the interaction in our model.

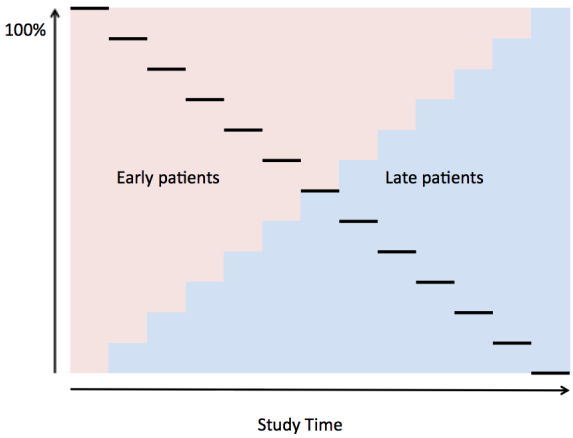


Figure 3.2: Expected log hazard ratios comparing the treatment to the control groups in each time interval when the model is not correct

The horizontal black lines indicate the theoretical log hazard ratios comparing the treatment to the control group in each time interval if an interaction between treatment and status is generated but not included when fitting a model. However, if we fit a model with

an interaction term, the log hazard ratio is constant within early and late patients over time.

Binomial distributions are used to divide patients into early patients and late patients within each time interval. The binomial distributions pertaining to late patients in each time interval are as follows:

$$I_{i\cdot} \sim \text{Binomial}(n_{iI}, \mid 1 - \frac{13-i}{12} \mid) \qquad 1 \leq i \leq 10$$
$$C_{i'\cdot} \sim \text{Binomial}(n_{i'C}, \mid 1 - \frac{3-i'}{12} \mid) \qquad 1 \leq i' \leq 3$$

where n_{iI} is the total number of patients enrolled in the i^{th} time interval before the interim analysis (Figure 2.1: Area A) and $n_{i'C}$ is total number of patients enrolled in the i'^{th} time interval after the interim analysis (Figure 2.1: Area B). $I_{i\cdot}$ represents the number of late patients enrolled in the i^{th} time interval before the interim analysis and $C_{i'\cdot}$ represents the number of late patients who do not contribute to the interim analysis and are enrolled in the i'^{th} time interval.

For each time interval, the ratio of the treatment to the control group is generated to be approximately 1:1 regardless of patients' status. Binomial distributions are used to assign early and late patients into either the treatment or the control group within each interval. The following distributions are used to generate indicator variables representing the treatment and control groups

$$I^k_{i\cdot} \sim \text{Binomial}(n^k_{iI}, 0.5) \qquad 1 \leq i \leq 10,\ k \in 0,\ 1$$
$$C^k_{i'\cdot} \sim \text{Binomial}(n^k_{i'C}, 0.5) \qquad 1 \leq i' \leq 3,\ k \in 0,\ 1$$

where k is a binary indicator equaling one for late patients, and zero for early patients. Therefore, n^1_{iI} represents the number of late patients in the i^{th} time interval before the interim analysis (Figure 2.1: Area A) and $I^1_{i\cdot}$ represents the number of late patients enrolled in i^{th} time interval in the treatment group.

3.3 Fitting models

As described in Chapter 3.2, the proportion of the early and late patients changes over time. This fact plays a key role for a model to exhibit non-proportionality, when the hazard ratio comparing the treatment to the control groups among early patients is not the same as that among late patients and no interaction is fit. Thus, in our simulations, fitting different models in the same generated dataset can yield different results depending on how we choose the values for the coefficients of the exponentially distributed hazards and what we choose to include in the model. However, the proportional hazards assumption holds irrespective of adjusting for the interaction term if the two hazard ratios (comparing treatment to control among early patients and among late patients) are equal.

We fit three different models for 4 different sub-cohorts of the generated dataset:

$$\begin{aligned}
&\text{Model 1}: \ \lambda(t, \text{Trt}, \beta) &&= \lambda_0(t) \cdot \exp(\beta_{i1} \cdot \text{Trt}) \\
&\text{Model 2}: \ \lambda(t, \text{Trt}, \text{Status}, \gamma) &&= \lambda_0(t) \cdot \exp(\gamma_{i1} \cdot \text{Trt} + \gamma_{i2} \cdot \text{Status}) \\
&\text{Model 3}: \ \lambda(t, \text{Trt}, \text{Status}, \theta) &&= \lambda_0(t) \cdot \exp(\theta_{i1} \cdot \text{Trt} + \theta_{i2} \cdot \text{Status} + \theta_{i3} \cdot \text{Trt} \cdot \text{Status})
\end{aligned} \tag{3.1}$$

where i = 1, 2, 3 or 4. The numbers 1, 2, 3 and 4 indicate the interim analysis, Figure 2.1: Area A when i=1 and the confirmatory analyses, Figure 2.1: Areas B and C when i=2, Areas B and D when i=3, and Areas B, C, D and E when i=4.

Regardless of the existence of the interaction effect, fitting model 1 in (3.1) is always incorrect because we choose to generate a status effect. However, whether fitting model 2 is correct or not depends on the four coefficients. If the hazard ratio comparing the treatment to the control group among the early patients is not the same as that among late patients, fitting model 2 is theoretically incorrect. On the other hand, fitting model 3 is correct in that situation because model 3 accounts for the interaction effect. When the two hazard ratios are equal, fitting model 2 is correct because there is no interaction and change in patient mix over time has no impact on the estimated hazard ratio. In this situation, fitting model 3 can potentially be considered overfitting. Therefore, which of the three different models

is fit and whether or not (and the strength of) an interaction effect is generated for any of the simulated datasets plays an important role when investigating the power to detect a treatment effect in confirmatory analyses.

3.4 Power

Investigating differences in power is the main interest of this research. In our simulations, we focus on two powers: the power of confirmatory analyses that do not use recurrence-free survivors (Figure 2.1: Areas B and C) and the power of confirmatory analyses that use recurrence-free survivors as delayed entry (Figure 2.1: Areas B, C, D and E) in addition to patients who were enrolled, but did not contribute any observation time to the interim analysis. Each of these powers is estimated by calculating the probability of finding a significant harmful treatment effect at the confirmatory analysis (with or without the use of recurrence-free survivors, as appropriate) given a significant harmful treatment effect was identified at the interim analysis.

Keiding et al. (1987) identified an unexpected harmful treatment effect at the interim analysis. However, a confirmatory analysis (Figure 2.1: Areas B and C) failed to show a significant harmful treatment effect, because the accrual time for the confirmatory analysis (Figure 2.1: Areas B and C) was not sufficient to enroll enough patients into the trial. In contrast, a strongly significant harmful treatment effect was found after combining recurrence-free survivors with the cohort enrolled after the interim analysis and following them up for up to four years (Figure 2.1: Areas B, C, D and E). The observation time from recurrence-free survivors that followed the interim analysis was used as delayed entry. Keiding et al. (1987) also performed a confirmatory analysis with the same patients/recurrence-free survivors, but using only follow-up data for up to one year (Figure 2.1: Areas B and D) and also found a statistically significant harmful treatment effect in that analysis.

We compare the power at the confirmatory analysis where the observation time from recurrence-free survivors that follows the interim analysis is treated as delayed entry to the power for the analysis without them. We compare the powers in settings when the

proportional hazards assumption holds and when the assumption does not hold. The powers we obtained are conditional on identifying a statistically significant harmful treatment effect at the interim analysis.

We carry out the simulations as follows. First, we perform 1,000 simulations resulting in 1,000 generated datasets. Using each of the 1,000 generated datasets, we fit three different models as specified in (3.1). Then, we perform the likelihood ratio tests to determine whether or not the treatment effect at the interim analysis in each model is statistically significant. Because we only use the datasets that identify a statistically significant treatment effect (at alpha=0.05) at the interim analysis, each model is fit to a different subset of generated datasets based on the p-values from the likelihood ratio tests. After selecting the datasets (with a statistically significant harmful treatment effect at the interim analysis) for each model, we calculate the power of the study with two versions of the confirmatory analysis (Figure 2.1: Areas B, C, D and E and Figure 2.1: Areas B and C) for each model in (3.1).

To calculate the power at the confirmatory analysis using recurrence-free survivors (Figure 2.1: Areas B, C, D and E), using the first model in (3.1) we count the number of statistically significant $\hat{\beta}_{41}$ when $\hat{\beta}_{11}$ is statistically significant. This number is divided by the number of datasets that have a statistically significant $\hat{\beta}_{11}$. This can be expressed as follows:

$$\frac{\text{The number of cases satisfying p-values of } \hat{\beta}_{11} \text{ and } \hat{\beta}_{41}}{\text{The number of cases satisfying p-values of } \hat{\beta}_{11}} \tag{3.2}$$

The power of the confirmatory analysis that does not recurrence-free survivors (Figure 2.1: Areas B and C) can be calculated as follows:

$$\frac{\text{The number of cases satisfying p-values of } \hat{\beta}_{11} \text{ and } \hat{\beta}_{21}}{\text{The number of cases satisfying p-values of } \hat{\beta}_{11}} \tag{3.3}$$

By using the second and the third models in (3.1), we also calculate the power of the study at the two confirmatory analyses similarly to what is described above. The power for the second model is calculated in the same manner as in (3.2) and (3.3), with the parameter $\hat{\beta}$ being replaced with $\hat{\gamma}$. However, to evaluate the power in the third model, the likelihood

ratio test with two degrees of freedom is used to assess the significance of the treatment effect. This can be expressed as follows:

$$\frac{\text{The number of cases satisfying p-values of } (\hat{\theta}_{11},\ \hat{\theta}_{13}) \text{ and } (\hat{\theta}_{41},\ \hat{\theta}_{43})}{\text{The number of cases satisfying p-values of } (\hat{\theta}_{11},\ \hat{\theta}_{13})}$$

We are interested in comparing the difference in power between the confirmatory analysis using recurrence-free survivors and the one without them, within each model.

3.5 Choosing coefficients

We set N_1 and N_2 to 1,200 and 300, respectively. We also set t_1 and t_2 to 3 and 1, respectively. These numbers are selected to be similar to the setting of Keiding et al. (1987). We generate exponentially distributed censoring times with a rate of $\exp(-2) \approx 0.14$, which allows for the proportion of the recurrence-free survivors to be between 65% to 73%.

In each of the three different models, we are interested in comparing the power of two confirmatory analyses, one that combines observation time from recurrence-free survivors that follows the interim analysis (used as delayed entry) with the cohort that did not contribute any observation time to the interim analysis and the other analysis that does not use any observation time from the recurrence-free survivors. We generate an interaction effect between treatment and status such that a weighted sum of them is constant as the interaction effect is made stronger. The power comparisons allow us to estimate what can be gained by using observation time from recurrence-free survivors across a range of interaction effect and how violation of the proportional hazards assumption affects the difference in power. In our simulations, we use four different weighted sums of treatment effects: 1.2, 1.3, 1.4 and 1.5. The weighted sum of treatment effects is calculated as follows:

$$\text{HR(WeightedSum)} = 0.8 \times \text{HR}_\text{I} + 0.2 \times \text{HR}_\text{C} \tag{3.4}$$

where HR_I is a hazard ratio for patients enrolled prior to the interim analysis and HR_C is a hazard ratio for patients enrolled after the interim analysis. The values 0.8 and 0.2 for the

weighted sum are chosen because they represent the proportion of patients enrolled prior to the interim analysis and after the interim analysis, respectively. Then, the HR_I and HR_C is specified as follows:

$$\text{HR}_\text{I} = 0.625 \times \text{HR(Early)} + 0.375 \times \text{HR(Late)} \tag{3.5}$$

$$\text{HR}_\text{C} = 0.08 \times \text{HR(Early)} + 0.92 \times \text{HR(Late)} \tag{3.6}$$

where HR(Early) indicates the hazard ratio comparing the treatment to control groups among patients who were enrolled into the study early due to having developed the disease long before the start of the study, and HR(Late) indicates the hazard ratio comparing treatment to control groups among patients who enrolled late into the study because they developed the disease recently. In the above equations, each weight in front of the hazard ratios refers to the proportion of patients that are early or late at each analysis. Plugging (3.5) and (3.6) into (3.4) yields:

$$\begin{aligned}\text{HR(WeightedSum)} = {} & 0.8 \times \{0.625 \times \text{HR(Early)} + 0.375 \times \text{HR(Late)}\} + \\ & 0.2 \times \{0.08 \times \text{HR(Early)} + 0.92 \times \text{HR(Late)}\} \\ = {} & 0.516 \times \text{HR(Early)} + 0.484 \times \text{HR(Late)}\end{aligned}$$

While holding this weighted sum constant, we change the strength of the interaction effect. In our simulations, the (exponentiated) interaction effect varies from 0.1 to 1. Thus, after setting the weighted sum of treatment effects and the interaction effect, we can determine the values of HR_I and HR_C for our simulations. Table 3.1 below illustrates the hazard ratios for both early patients and late patients for each combination of the weighted sum of the treatment effects and interaction effect.

Table 3.1: List of hazard ratios comparing the treatment to control group among patients enrolled early and patients enrolled late for each of the chosen constant weighted sums of treatment effects.

HR(WS)	**Interaction**	**HR(Early)**	**HR(Late)**	**HR(WS)**	**HR(Early)**	**HR(Late)**
1.2	0.1	2.18	0.22	1.4	2.55	0.26
	0.2	2.00	0.40		2.33	0.47
	0.3	1.85	0.55		2.16	0.65
	0.4	1.72	0.69		2.00	0.80
	0.5	1.60	0.80		1.87	0.93
	0.6	1.50	0.90		1.75	1.05
	0.7	1.41	0.99		1.65	1.15
	0.8	1.33	1.07		1.56	1.25
	0.9	1.26	1.14		1.47	1.33
	1.0	1.20	1.20		1.40	1.40
1.3	0.1	2.37	0.24	1.5	2.73	0.27
	0.2	2.17	0.43		2.50	0.50
	0.3	2.00	0.60		2.31	0.69
	0.4	1.86	0.74		2.14	0.86
	0.5	1.73	0.87		2.00	1.00
	0.6	1.63	0.98		1.88	1.13
	0.7	1.53	1.07		1.77	1.24
	0.8	1.45	1.16		1.67	1.33
	0.9	1.37	1.23		1.58	1.42
	1.0	1.30	1.30		1.50	1.50

As mentioned in Chapter 3.4, we estimate power while conditioning on having a statistically significant harmful treatment effect at the interim analysis.

Chapter 4

RESULTS

Table 4.1 below illustrates the number of recurrence-free survivors for each combination of the interaction effect and the constant weighted sum (WS) of treatment effects. Among the different combinations, the average proportion of the recurrence-free survivors among the initial cohort varies from 0.65 to 0.7 (see Table 4.1).

Table 4.1: The average number of recurrence-free survivors, with accompanying 5% and 95% percentiles, for each combination of the constant weighted sum of treatment effects and interaction effect. 1,000 replications are conducted for all combinations

HR(WS)	Interaction	# Recurrence-free S	HR(WS)	Interaction	# Recurrence-free S
1.2	0.1	813 (785, 839)	1.4	0.1	831 (802, 857)
	0.3	832 (808, 858)		0.3	835 (810, 861)
	0.5	831 (806, 858)		0.5	839 (814, 864)
	0.7	828 (800, 855)		0.7	841 (813, 867)
	0.9	827 (802, 851)		0.9	837 (812, 864)
1.3	0.1	826 (798, 852)	1.5	0.1	835 (805, 861)
	0.3	831 (805, 856)		0.3	839 (813, 864)
	0.5	835 (810, 859)		0.5	842 (818, 868)
	0.7	837 (808, 863)		0.7	845 (816, 870)
	0.9	833 (807, 860)		0.9	831 (815, 867)

In Table 4.1 above, the column "# Recurrence-free S" means the average number of recurrence-free survivors among the initial cohort from 1,000 simulations. Two numbers inside of parentheses indicate 5% percentile and 95% percentile of recurrence-free survivors from 1,000 simulations.

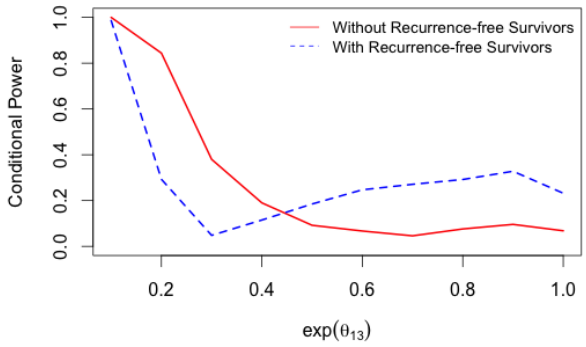
In Table 4.1 above, the number 813 in the first row indicates that among the 1,200 patients at the interim analysis, the number of recurrence-free survivors was 813 on average

with 785 and 839 for 5% and 95% percentiles each, when we set the hazard ratio among early patients to be 2.18 and we set the hazard ratio among late patients to be 0.22 in the 1,000 simulations (See Table 3.1).

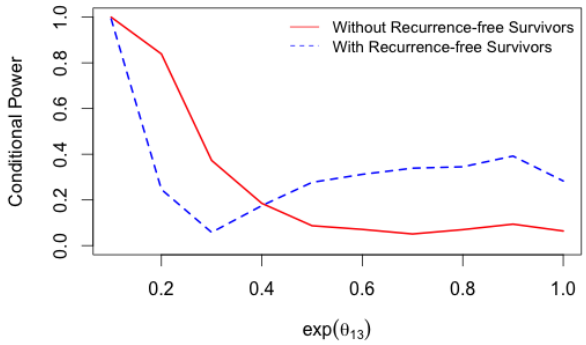
We generate the datasets to have an interaction effect by choosing the coefficients for the exponentially distributed hazards such that the hazard ratio comparing treatment to control group is different among early patients compared to late patients. Although, some of the datasets and models will not show a statistically significant interaction at the interim analysis (by random chance), the interaction effect theoretically exists based on the way we simulate the datasets. When we generate an interaction effect, the proportional hazards assumption (theoretically) does not hold for a model that does not include an interaction term. Thus, (theoretically) the proportional hazards assumption does not hold when we fit either model 1 or model 2 in (3.1). Nevertheless, we might not have sufficient power to detect non-proportionality. On the other hand, when generating an interaction effect and fitting model 3 in (3.1) the proportional hazards assumption holds.

Whether the proportional hazards assumption holds or not depends on the fitted model. As the difference between the two hazard ratios (hazard ratio comparing treatment to control among early patients and the hazard ration comparing treatment to control among late patients) increases, the interaction effect is stronger. Within each model, we evaluate the difference in power between the confirmatory analysis using the observation time from recurrence-free survivors that follows the interim analysis (as delayed entry) and the analysis without using them at each level of strength of the interaction effect. We also compare the difference in powers among the three different models.

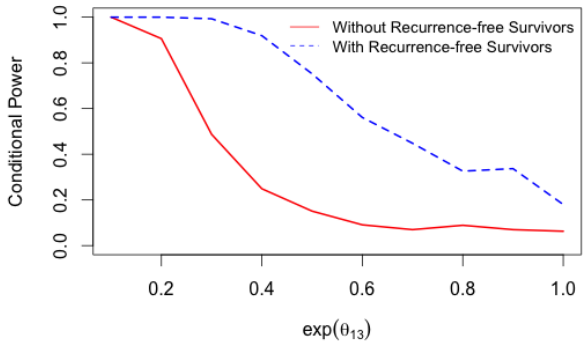
Figures 4.1, 4.2, 4.3 and 4.4 each consist of three subfigures: (a), (b) and (c). All subfigures in Figures 4.1, 4.2, 4.3 and 4.4 include two lines. The lines indicate the conditional power of the (hypothetical/simulated) studies at the confirmatory analysis using the observation time from recurrence-free survivors that follows the interim analysis (as delayed entry) (Dashed blue line) and that of the analysis without using them (Continuous red line), i.e. using only patients who did not contribute any observation time to the interim analysis. For



(a) Model 1

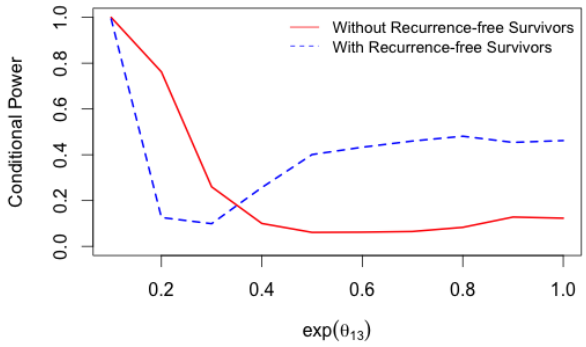


(b) Model 2

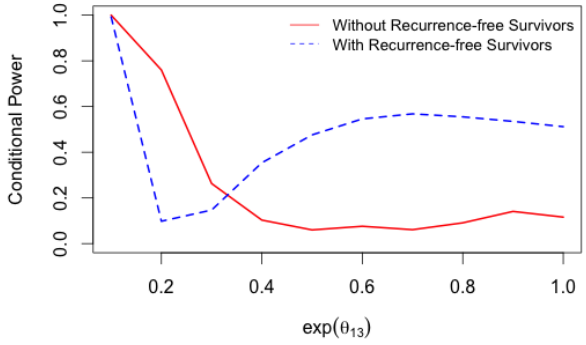


(c) Model 3

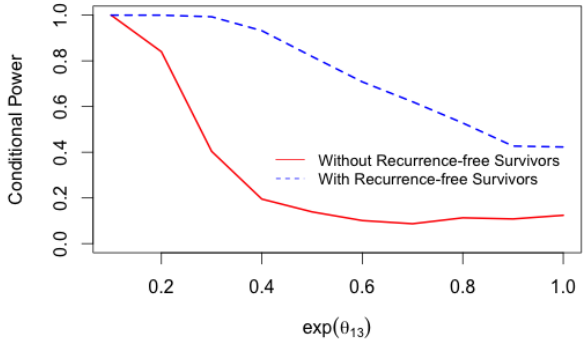
Figure 4.1: The weighted sum of treatment effect is 1.2



(a) Model 1

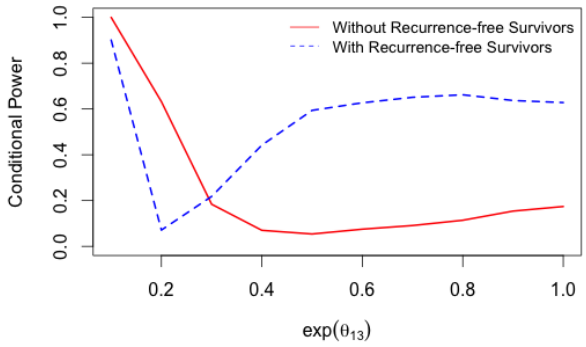


(b) Model 2

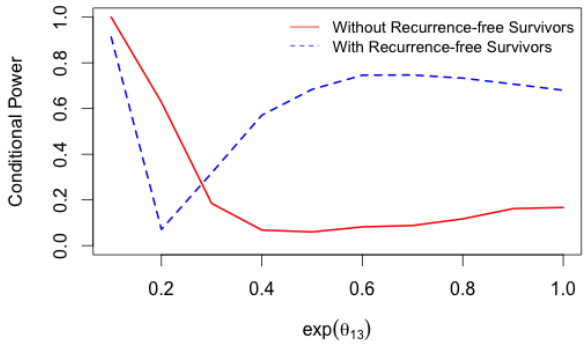


(c) Model 3

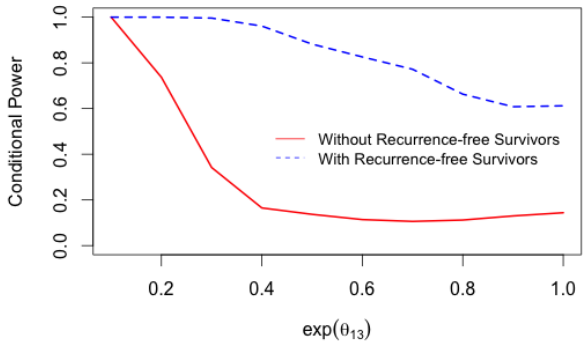
Figure 4.2: The weighted sum of treatment effect is 1.3



(a) Model 1

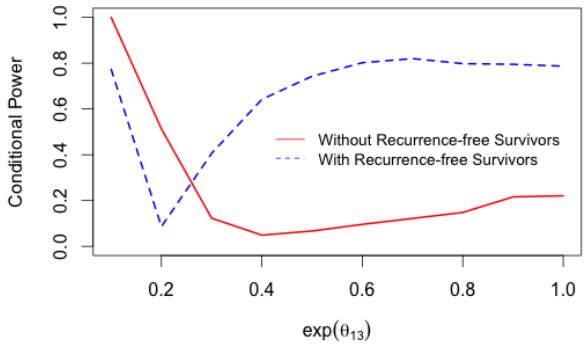


(b) Model 2

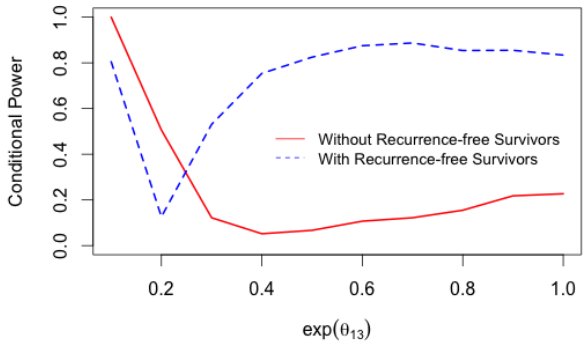


(c) Model 3

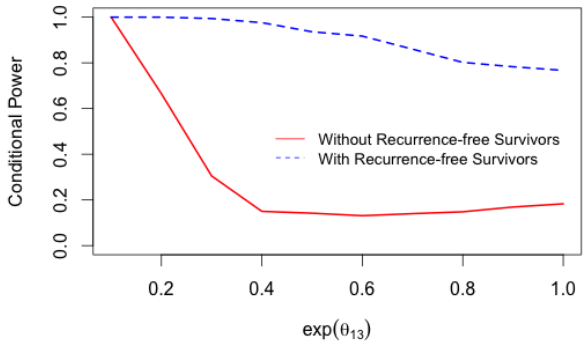
Figure 4.3: The weighted sum of treatment effect is 1.4



(a) Model 1



(b) Model 2



(c) Model 3

Figure 4.4: The weighted sum of treatment effect is 1.5

each of the subfigures, the y-axis and x-axis indicate the conditional power and interaction effect, respectively. The interaction effect is stronger the further away it is from 1.

Subfigures (a), (b) and (c) are for models 1, 2, and 3, respectively. The proportional hazards assumption does not hold (theoretically) in subfigures (a) and (b) except when then exponentiated coefficient for the interaction term is one, because models 1 and 2 do not include an interaction term. Nevertheless, non-proportionality will be more pronounced as the exponentiated coefficient is further away from 1. The proportional hazards assumption holds in subfigure (c) for all choices of the interaction effect.

The difference between the four Figures is the weighted sum of the treatment effects. As we illustrate in Table 3.1, we use four different weighted sums of treatment effects: 1.2, 1.3, 1.4 and 1.5.

Figure 4.1 illustrates that using observation time from recurrence-free survivors that follows the interim analysis (as delayed entry) provides higher power when the interaction effect is weak (through the increase in sample size and a similar treatment effect among both, early and late patients). However, as interaction effect gets stronger, the power of the confirmatory analysis using the recurrence-free survivors as delayed entry decreases for models that do not include the interaction effect (the marginal effect of treatment is a combination of a protective effect among late patients and a harmful effect among early patients). The increase in the interaction effect can also be interpreted as the increase in non-proportionality for models 1 and 2. When the (exponentiated) interaction effect drops below around 0.4 (in Figure 4.1, and below 0.35, 0.3 and 0.25 in Figures 4.2, 4.3 and 4.4 respectively), the power of the analysis without using the recurrence-free survivors is higher compared to the power of analysis using them for models 1 and 2 because they are primarily estimating the effect of treatment among mostly late patients who (in these scenarios) have a strong beneficial effect. In contrast, for model 3, which includes the interaction effect, power increases as the interaction effect increases regardless whether or not the recurrence-free survivors are used. Furthermore, power is always higher for the analysis using the observation time from recurrence-free survivors that follows the interim analysis (as delayed

entry) for model 3 as it correctly models and estimates the treatment effect(s).

Figure 4.2 shows a similar trend as Figure 4.1. Figure 4.2 also illustrates that if the proportional hazards assumption does not hold, using the recurrence-free survivors does not always lead to improved power of analysis. Nevertheless, models 1 and 2 have higher power to detect a marginal effect that is reflective of a combination of effects that were generated for late and early patients. According to subfigures (a) and (b) in Figure 4.2, the power of the analysis with recurrence-free survivors decreases as the interaction effect becomes stronger, and then it increases after the interaction effect drops below around 0.2. In contrast, when we fit the model that adjusts for the interaction, combining the recurrence-free survivors with the cohort that did not contribute to the interim analysis always leads to higher power.

Figures 4.3 and 4.4 also illustrate a similar trend as the other two Figures. As we increase the weighted sum of treatment effects, using recurrence-free survivors as delayed entry improves power even when the interaction effect is weak which is reflective of the increased sample size. For Figures 4.1 and 4.2, which have a low value of the weighted sum of treatment effects, the increase in power of the analysis with recurrence-free survivors is not very large regardless of model fitting when the interaction effect is around 1.0. However, Figures 4.3 and 4.4 illustrate that when the interaction effect is weak (around 1.0), using recurrence-free survivors as delayed entry is beneficial no matter what model we fit. However, in subfigures (a) and (b) in each Figure, the power of the analysis with recurrence-free survivors decreases as the interaction effect increases because the models used for subfigures (a) and (b) do not include the interaction effect and represent a mix of early patients where the treatment effect is harmful and late patients where the treatment effect is beneficial.

As we see in subfigures (a) and (b) for each of the Figures, the power of the analysis using the recurrence-free survivors has an inflection point when the interaction effect is around 0.2. As we illustrate in table 4.1, the recurrence-free survivors are around 830, on average. Among the 830 recurrence-free survivors, 55% of them are early patients and 45% are late patients, on average. Thus, among patients used for the confirmatory analysis (Figures 2.1: Areas B, C, D and E), 460 of them are early patients and 670 of them are late patients (including

the 300 patients who did not contribute to the interim analysis who are almost exclusively late patients), on average. In our simulations, we set the weighted sum of treatment effects and the interaction effect when generating the datasets. Thus, to make the interaction effect strong, the difference in hazard ratios (comparing treatment to control group) for early patients and for late patients increases (See Table 3.1). Because of the imbalance in number between early patients and late patients, the marginal treatment effect (models 1 and 2) for the analysis with the recurrence-free survivors decreases and it is close to 1.0 when the interaction effect is approximately 0.2 (an about doubling of hazard among early patients and about halving of hazard among late patients). Since the marginal treatment effect is further away from 1.0 with the interaction effect of 0.1 compared to the marginal treatment effect when the interaction is 0.2, the power is higher.

It is interesting that Keiding et al. (1987) found a statistically significant treatment effect with a follow-up of only one year (Figure 2.1: Areas B and D). Therefore, we also investigate the differences in powers of confirmatory analyses that only include the cohort that does not contribute to the interim analysis to the confirmatory analysis that combines this cohort with observation time from recurrence-free survivors following the interim analysis but each contributing only up to one year (Figure 2.1: Areas B and D) instead of up to four years.

We repeat the same procedures to find the powers and compare them as described in Chapter 3. However, in these repeated procedures, we instead perform the confirmatory analysis one year after the interim analysis. The results are illustrated in Figures B.1, B.2, B.3 and B.4 (see Appendix). Each Figure in the Appendix has a trend similar as those in Figures 4.1, 4.2, 4.3 and 4.4, with the notable difference that the power is generally lower.

Chapter 5

LIMITATIONS

When generating the datasets, the sample sizes for the initial cohort and the cohort that does not contribute to the interim analysis, as well as the amount of time for enrollment, are approximately equal to those stated in Keiding et al. (1987). The hazard ratios for early patients and late patients are obtained after fixing a weighted sum of treatment effects and the interaction effect. The calculation for determining the two hazard ratios that conform with these choices is illustrated in Section 3.4. However, different sample sizes for the two cohorts, a different length of enrollment time of both analyses, and different distributions of failure times in the four groups can be chosen. We expect that using different sample sizes for the two cohorts can change the results, however, we believe that the setting (particularly regarding the limited number of patients enrolled that do not contribute to the interim analysis) mentioned by Keiding et al. (1987) is a reasonable setting for the scenario where a harmful treatment effect might be found at an interim analysis and there are concerns regarding enrollment of patients for a study that aims to confirm a harmful effect. Also, changing the weighted sum of treatment effects can yield very different results. For example, from our simulations, we only show the four different weighted sums of treatment effects: 1.2, 1.3, 1.4 and 1.5. However, if the weighted sum of treatment effects is extremely high, the hazard ratios for early patients and late patients will always be away from 1. For example, if we set the weighted sum of treatment effects to be 10, then the hazard ratios for early patients and late patients are 18.2 and 1.8, respectively. Both numbers are far way from 1, so the power is always high regardless of model fitting. Thus, among the three models, there would be no large difference in power for the analyses using recurrence-free survivors if we set the weighted sum of treatment effects to be extremely high. However, such large

treatment effects are extremely unlikely from a practical health research perspective.

Chapter 6

DISCUSSION

In this work, we investigated the power of analyses that combine the observation time from recurrence-free survivors that follows the interim analysis with a cohort that did not contribute to the interim analysis to the power analyses that do not make use of any observation time from the initial cohort. We investigate the difference in powers for the case where the proportional hazards assumption holds and where it does not hold. In addition, within each model, we estimate how the power of the analysis using recurrence-free survivors and the analysis without them changes as the interaction effect changes. As the interaction effect gets stronger, non-proportionality is also stronger if we fit a model that does not includes an interaction term. Thus the proportional hazards assumption holds when we fit the model including the interaction term (model 3), but it does not hold in the other models (models 1 and 2).

In our simulations, if the proportional hazards assumption holds (model 3), we observe that the power is improved when combining the recurrence-free survivors with the cohort that did not contribute to the interim analysis, for each chosen weighted sum of treatment effects. As the weighted sum of treatment effects increases, the difference in power between the analysis using recurrence-free survivors and the analysis without them increases. This indicates that using recurrence-free survivors is more effective as the weighted sum of treatment effects increases in cases where the proportional hazards assumption holds.

In contrast, if the proportional hazards assumption does not hold (models 1 and 2), the difference in powers depends on the interaction effect and the weighted sum treatment effects. For instance, when weighted sum of treatment effects is either 1.2 or 1.3, we obtain higher power by including the recurrence-free survivors, (although power is still very low, less than

60%) if the interaction effect is close to 1. The power decreases as the interaction effect is stronger, which is reasonable because the models (1 and 2) estimate a treatment effect that is a combination of a harmful and a beneficial effect. In addition, if the interaction effect is very strong (around 0.2), using the recurrence-free survivors provides the lowest estimated power because the harmful and beneficial effects approximately cancel each other out. However, we obtain reasonably higher power by using the recurrence-free survivors for our analysis even when the interaction effect is close to 1 if the weighted sum of treatment effects is high and the interaction effect is not strong.

In summary, if the proportional hazards assumption holds (either because there is no time trend, or if the time trend is accommodated by the model), we are able to improve the power of the study at the confirmatory analysis (Figure 2.1: Areas B and C) by using the observation time from recurrence-free survivors after the interim analysis as delayed entry and combining it with the cohort that did not contribute to the interim analysis. However, if the proportional hazards assumption does not hold (there is a time trend and we do not model it correctly), using the recurrence-free survivors might not overcome issues arising from using an incorrect model.

BIBLIOGRAPHY

Cox, D. R. (1972), 'Regression models and life-tables', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 187–220.

Cox, D. R. (1975), 'Partial likelihood', *Biometrika* **62**(2), 269–276.

Grambsch, P. M. & Therneau, T. M. (1994), 'Proportional hazards tests and diagnostics based on weighted residuals', *Biometrika* **81**(3), 515–526.

Keiding, N. (2006), 'Event history analysis and the cross-section', *Statistics in medicine* **25**(14), 2343–2364.

Keiding, N., Bayer, T. & Watt-Boolsen, S. (1987), 'Confirmatory analysis of survival data using left truncation of the life times of primary survivors', *Statistics in medicine* **6**(8), 939–944.

Parner, E. T. & Keiding, N. (2001), 'Misspecified proportional hazard models and confirmatory analysis of survival data', *Biometrika* **88**(2), 459–468.

Appendix A

SIMULATIONS

The details of the time points for patients enrolled in the study

The time points for patients recruited in each time interval are expressed as

$$\begin{aligned} I_{ij} &= \{\, I_p \mid \frac{1}{10} \times t_1 \times (i-1) < \; I_p \; \leq \; \frac{1}{10} \times t_1 \times i\} \qquad 1 \leq i \leq 10,\; i \in \mathbb{N},\; j \leq \; p \\ C_{i'j'} &= \{\, C_{p'} \mid \frac{1}{3} \times t_2 \times (i'-1) < \; C_{p'} \; \leq \; \frac{1}{3} \times t_2 \times i\} \qquad 1 \leq i' \leq 3,\; i' \in \mathbb{N},\; j' \leq \; p' \end{aligned} \tag{A.1}$$

where I_{ij} is the time for the j^{th} patient in the i^{th} interval to enter the study before the interim analysis and $C_{i'j'}$ is the time for j'^{th} patient in the i'^{th} interval after the interim analysis. Suppose $t_1 = 3$ and $I_{i\cdot 50} = 2.3$. for the 50^{th} patient. This means that the time for the 50^{th} patient to be enrolled is 2.3 and that person would be in 8^{th} interval because 2.3 lies between 2.1 ($= \frac{3}{10} \times (8-1)$) and 2.4 ($= \frac{3}{10} \times (8)$).

The details of patients who are enrolled in late patients

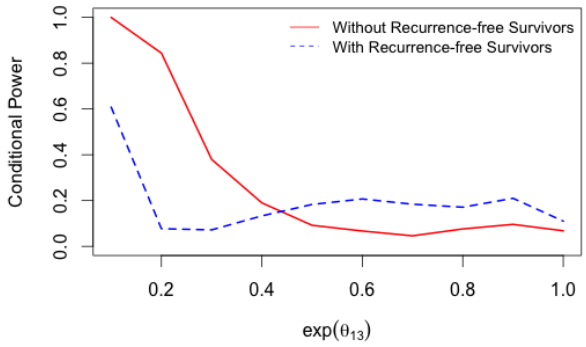
For instance, suppose the total number of patients in 5^{th} time interval before the interim analysis is 96 (n_{5I}=96). In 5^{th} time interval, the probability of a late patient is estimated to be $\frac{1}{3}$ ($= \; 1 - \frac{13-5}{12}$). The expected number of late patients is 32 ($= 96 \times (1 - \frac{13-5}{12})$).

The details of late patients who are enrolled in the treatment group

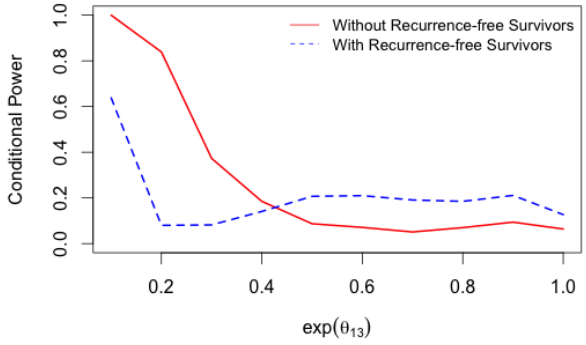
For example, $I^1_{5\cdot}$ is the number of late patients in the 5^{th} time interval. Therefore, continuing with the example used for (3.2), $n^1_{5\cdot}$ would be 32. Then, the expected number of late patients pertaining to the treatment group before the interim analysis is 16 ($=32 \times \frac{1}{2}$).

Appendix B

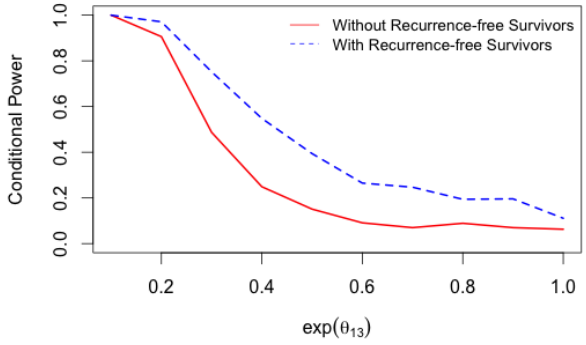
ADDITIONAL FIGURES



(a) Model 1

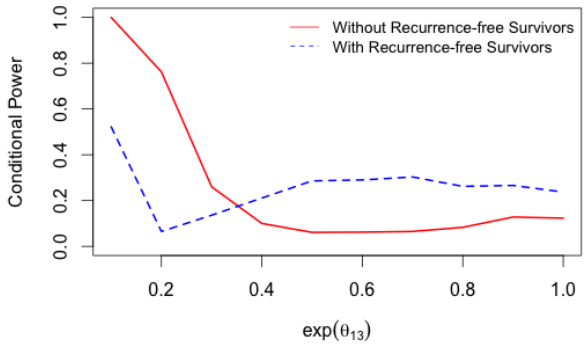


(b) Model 2

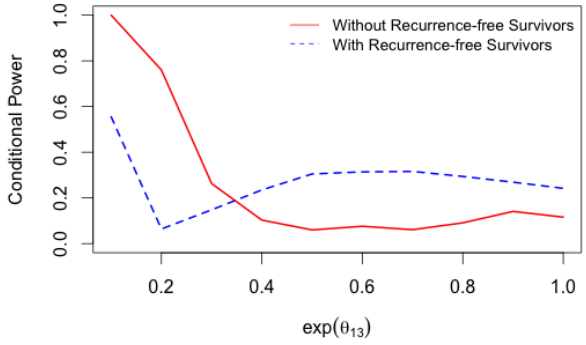


(c) Model 3

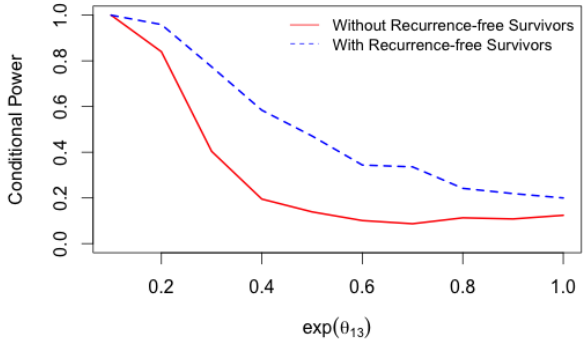
Figure B.1: The weighted sum of treatment effect is 1.2



(a) Model 1

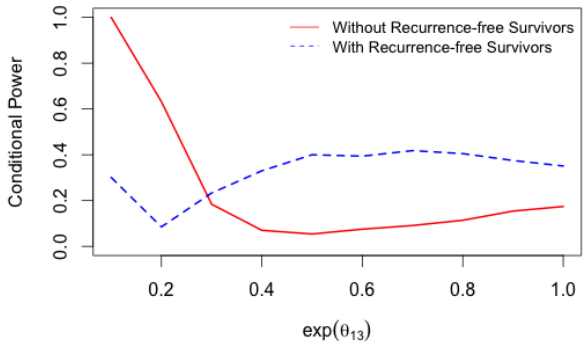


(b) Model 2

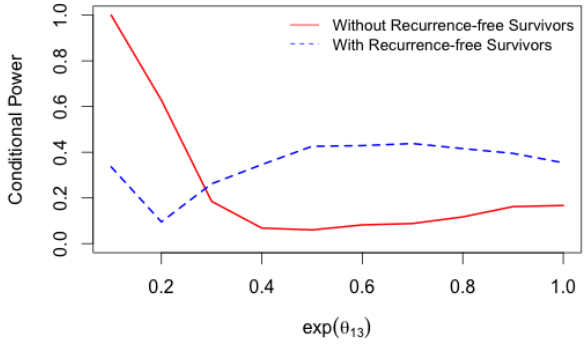


(c) Model 3

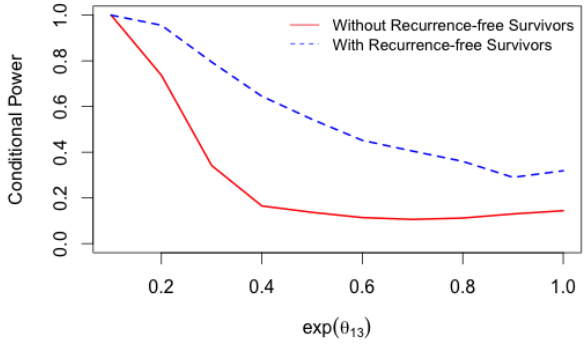
Figure B.2: The weighted sum of treatment effect is 1.3



(a) Model 1

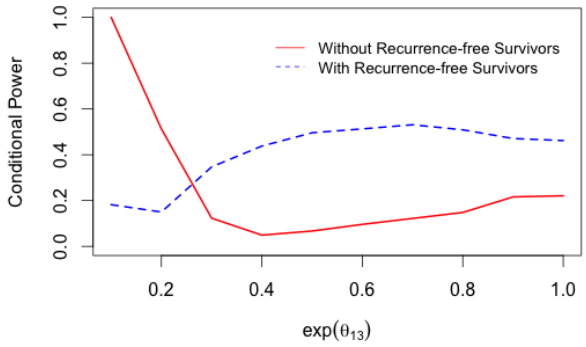


(b) Model 2

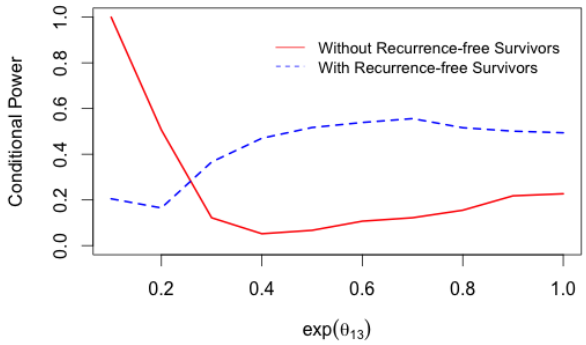


(c) Model 3

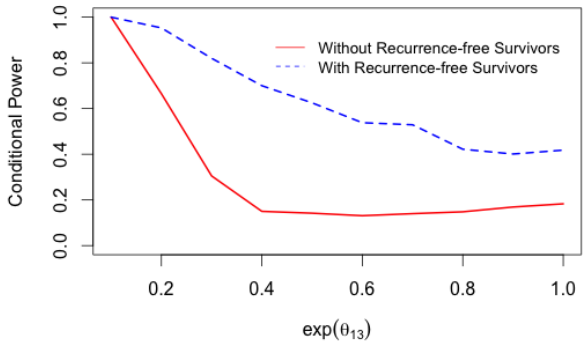
Figure B.3: The weighted sum of treatment effect is 1.4



(a) Model 1



(b) Model 2



(c) Model 3

Figure B.4: The weighted sum of treatment effect is 1.5

Appendix C

R CODE

```
library(survival)
library(zoo)
library(lmtest)

SimV <- function(n1,n2,t1,t2,beta1,beta2,beta3,beta4,cen1){
  T <- runif(n1,0,t1)
  g1 <- as.numeric(T<=t1/10)
  g2 <- as.numeric(T>t1/10 & T<=2*t1/10)
  g3 <- as.numeric(T>2*t1/10 & T<=3*t1/10)
  g4 <- as.numeric(T>3*t1/10 & T<=4*t1/10)
  g5 <- as.numeric(T>4*t1/10 & T<=5*t1/10)
  g6 <- as.numeric(T>5*t1/10 & T<=6*t1/10)
  g7 <- as.numeric(T>6*t1/10 & T<=7*t1/10)
  g8 <- as.numeric(T>7*t1/10 & T<=8*t1/10)
  g9 <- as.numeric(T>8*t1/10 & T<=9*t1/10)
  g10 <- as.numeric(T>9*t1/10 & T<=10*t1/10)
  mat <- data.frame(T,g1,g2,g3,g4,g5,g6,g7,g8,g9,g10)
  ## Prev VS Inci ##
  for(i in 1:n1){
    mat$g11[i] <- ifelse(mat$g1[i]==1,rbinom(1,1,0),0)
    mat$g12[i] <- ifelse(mat$g2[i]==1,rbinom(1,1,1/12),0)
    mat$g13[i] <- ifelse(mat$g3[i]==1,rbinom(1,1,2/12),0)
    mat$g14[i] <- ifelse(mat$g4[i]==1,rbinom(1,1,3/12),0)
```

```
  mat$g15[i] <- ifelse(mat$g5[i]==1,rbinom(1,1,4/12),0)
  mat$g16[i] <- ifelse(mat$g6[i]==1,rbinom(1,1,5/12),0)
  mat$g17[i] <- ifelse(mat$g7[i]==1,rbinom(1,1,6/12),0)
  mat$g18[i] <- ifelse(mat$g8[i]==1,rbinom(1,1,7/12),0)
  mat$g19[i] <- ifelse(mat$g9[i]==1,rbinom(1,1,8/12),0)
  mat$g20[i] <- ifelse(mat$g10[i]==1,rbinom(1,1,9/12),0)
}
## Treatment vs Control ##
for(i in 1:n1){
  mat$g21[i] <- if(mat$g1[i]==1 & mat$g11[i]==1)mat$g21[i]=rbinom(1,1,0.5)
  else if(mat$g1[i]==1 & mat$g11[i]==0)mat$g21[i]=rbinom(1,1,0.5) else mat$g21[i]=0
  mat$g22[i] <- if(mat$g2[i]==1 & mat$g12[i]==1)mat$g22[i]=rbinom(1,1,0.5)
  else if(mat$g2[i]==1 & mat$g12[i]==0)mat$g22[i]=rbinom(1,1,0.5) else mat$g22[i]=0
  mat$g23[i] <- if(mat$g3[i]==1 & mat$g13[i]==1)mat$g23[i]=rbinom(1,1,0.5)
  else if(mat$g3[i]==1 & mat$g13[i]==0)mat$g23[i]=rbinom(1,1,0.5) else mat$g23[i]=0
  mat$g24[i] <- if(mat$g4[i]==1 & mat$g14[i]==1)mat$g24[i]=rbinom(1,1,0.5)
  else if(mat$g4[i]==1 & mat$g14[i]==0)mat$g24[i]=rbinom(1,1,0.5) else mat$g24[i]=0
  mat$g25[i] <- if(mat$g5[i]==1 & mat$g15[i]==1)mat$g25[i]=rbinom(1,1,0.5)
  else if(mat$g5[i]==1 & mat$g15[i]==0)mat$g25[i]=rbinom(1,1,0.5) else mat$g25[i]=0
  mat$g26[i] <- if(mat$g6[i]==1 & mat$g16[i]==1)mat$g26[i]=rbinom(1,1,0.5)
  else if(mat$g6[i]==1 & mat$g16[i]==0)mat$g26[i]=rbinom(1,1,0.5) else mat$g26[i]=0
  mat$g27[i] <- if(mat$g7[i]==1 & mat$g17[i]==1)mat$g27[i]=rbinom(1,1,0.5)
  else if(mat$g7[i]==1 & mat$g17[i]==0)mat$g27[i]=rbinom(1,1,0.5) else mat$g27[i]=0
  mat$g28[i] <- if(mat$g8[i]==1 & mat$g18[i]==1)mat$g28[i]=rbinom(1,1,0.5)
  else if(mat$g8[i]==1 & mat$g18[i]==0)mat$g28[i]=rbinom(1,1,0.5) else mat$g28[i]=0
  mat$g29[i] <- if(mat$g9[i]==1 & mat$g19[i]==1)mat$g29[i]=rbinom(1,1,0.5)
  else if(mat$g9[i]==1 & mat$g19[i]==0)mat$g29[i]=rbinom(1,1,0.5) else mat$g29[i]=0
  mat$g30[i] <- if(mat$g10[i]==1 & mat$g20[i]==1)mat$g30[i]=rbinom(1,1,0.5)
  else if(mat$g10[i]==1 & mat$g20[i]==0)mat$g30[i]=rbinom(1,1,0.5) else mat$g30[i]=0
```

```
}
## hazard ##
for(i in 1:n1){
  mat$g31[i] <-
  if(mat$g1[i]==1 & mat$g11[i]==0 &mat$g21[i]==1)mat$g31[i]=rexp(1,rate=exp(beta1))
  else if(mat$g1[i]==1 & mat$g11[i]==0 &mat$g21[i]==0)mat$g31[i]=rexp(1,rate=exp(beta2))
  else if(mat$g1[i]==1 & mat$g11[i]==1 &mat$g21[i]==1)mat$g31[i]=rexp(1,rate=exp(beta3))
  else if(mat$g1[i]==1 & mat$g11[i]==1 &mat$g21[i]==0)mat$g31[i]=rexp(1,rate=exp(beta4))
  else mat$g31[i]=0
  mat$g32[i] <-
  if(mat$g2[i]==1 & mat$g12[i]==0 &mat$g22[i]==1)mat$g32[i]=rexp(1,rate=exp(beta1))
  else if(mat$g2[i]==1 & mat$g12[i]==0 &mat$g22[i]==0)mat$g32[i]=rexp(1,rate=exp(beta2))
  else if(mat$g2[i]==1 & mat$g12[i]==1 &mat$g22[i]==1)mat$g32[i]=rexp(1,rate=exp(beta3))
  else if(mat$g2[i]==1 & mat$g12[i]==1 &mat$g22[i]==0)mat$g32[i]=rexp(1,rate=exp(beta4))
  else mat$g32[i]=0
  mat$g33[i] <-
  if(mat$g3[i]==1 & mat$g13[i]==0 &mat$g23[i]==1)mat$g33[i]=rexp(1,rate=exp(beta1))
  else if(mat$g3[i]==1 & mat$g13[i]==0 &mat$g23[i]==0)mat$g33[i]=rexp(1,rate=exp(beta2))
  else if(mat$g3[i]==1 & mat$g13[i]==1 &mat$g23[i]==1)mat$g33[i]=rexp(1,rate=exp(beta3))
  else if(mat$g3[i]==1 & mat$g13[i]==1 &mat$g23[i]==0)mat$g33[i]=rexp(1,rate=exp(beta4))
  else mat$g33[i]=0
  mat$g34[i] <-
  if(mat$g4[i]==1 & mat$g14[i]==0 &mat$g24[i]==1)mat$g34[i]=rexp(1,rate=exp(beta1))
  else if(mat$g4[i]==1 & mat$g14[i]==0 &mat$g24[i]==0)mat$g34[i]=rexp(1,rate=exp(beta2))
  else if(mat$g4[i]==1 & mat$g14[i]==1 &mat$g24[i]==1)mat$g34[i]=rexp(1,rate=exp(beta3))
  else if(mat$g4[i]==1 & mat$g14[i]==1 &mat$g24[i]==0)mat$g34[i]=rexp(1,rate=exp(beta4))
   else mat$g34[i]=0
  mat$g35[i] <-
  if(mat$g5[i]==1 & mat$g15[i]==0 &mat$g25[i]==1)mat$g35[i]=rexp(1,rate=exp(beta1))
```

```
else if(mat$g5[i]==1 & mat$g15[i]==0 &mat$g25[i]==0)mat$g35[i]=rexp(1,rate=exp(beta2))
else if(mat$g5[i]==1 & mat$g15[i]==1 &mat$g25[i]==1)mat$g35[i]=rexp(1,rate=exp(beta3))
else if(mat$g5[i]==1 & mat$g15[i]==1 &mat$g25[i]==0)mat$g35[i]=rexp(1,rate=exp(beta4))
else mat$g35[i]=0
mat$g36[i] <-
if(mat$g6[i]==1 & mat$g16[i]==0 &mat$g26[i]==1)mat$g36[i]=rexp(1,rate=exp(beta1))
else if(mat$g6[i]==1 & mat$g16[i]==0 &mat$g26[i]==0)mat$g36[i]=rexp(1,rate=exp(beta2))
else if(mat$g6[i]==1 & mat$g16[i]==1 &mat$g26[i]==1)mat$g36[i]=rexp(1,rate=exp(beta3))
else if(mat$g6[i]==1 & mat$g16[i]==1 &mat$g26[i]==0)mat$g36[i]=rexp(1,rate=exp(beta4))
else mat$g36[i]=0
mat$g37[i] <-
 if(mat$g7[i]==1 & mat$g17[i]==0 &mat$g27[i]==1)mat$g37[i]=rexp(1,rate=exp(beta1))
else if(mat$g7[i]==1 & mat$g17[i]==0 &mat$g27[i]==0)mat$g37[i]=rexp(1,rate=exp(beta2))
else if(mat$g7[i]==1 & mat$g17[i]==1 &mat$g27[i]==1)mat$g37[i]=rexp(1,rate=exp(beta3))
else if(mat$g7[i]==1 & mat$g17[i]==1 &mat$g27[i]==0)mat$g37[i]=rexp(1,rate=exp(beta4))
else mat$g37[i]=0
mat$g38[i] <-
if(mat$g8[i]==1 & mat$g18[i]==0 &mat$g28[i]==1)mat$g38[i]=rexp(1,rate=exp(beta1))
else if(mat$g8[i]==1 & mat$g18[i]==0 &mat$g28[i]==0)mat$g38[i]=rexp(1,rate=exp(beta2))
else if(mat$g8[i]==1 & mat$g18[i]==1 &mat$g28[i]==1)mat$g38[i]=rexp(1,rate=exp(beta3))
else if(mat$g8[i]==1 & mat$g18[i]==1 &mat$g28[i]==0)mat$g38[i]=rexp(1,rate=exp(beta4))
else mat$g38[i]=0
mat$g39[i] <-
if(mat$g9[i]==1 & mat$g19[i]==0 &mat$g29[i]==1)mat$g39[i]=rexp(1,rate=exp(beta1))
else if(mat$g9[i]==1 & mat$g19[i]==0 &mat$g29[i]==0)mat$g39[i]=rexp(1,rate=exp(beta2))
else if(mat$g9[i]==1 & mat$g19[i]==1 &mat$g29[i]==1)mat$g39[i]=rexp(1,rate=exp(beta3))
else if(mat$g9[i]==1 & mat$g19[i]==1 &mat$g29[i]==0)mat$g39[i]=rexp(1,rate=exp(beta4))
else mat$g39[i]=0
mat$g40[i] <-
```

```
  if(mat$g10[i]==1 & mat$g20[i]==0 &mat$g30[i]==1)mat$g40[i]=rexp(1,rate=exp(beta1))
  else if(mat$g10[i]==1 & mat$g20[i]==0 &mat$g30[i]==0)mat$g40[i]=rexp(1,rate=exp(beta2))
  else if(mat$g10[i]==1 & mat$g20[i]==1 &mat$g30[i]==1)mat$g40[i]=rexp(1,rate=exp(beta3))
  else if(mat$g10[i]==1 & mat$g20[i]==1 &mat$g30[i]==0)mat$g40[i]=rexp(1,rate=exp(beta4))
  else mat$g40[i]=0
}
mat$Inci <- as.numeric(rowSums(mat[,(12:21)])>0)
mat$Trt <- as.numeric(rowSums(mat[,(22:31)])>0)
mat$Failure <- rowSums(mat[,(32:41)])
mat$CenT <- rexp(n1,rate=exp(cen1))
mat$Death <- ifelse(mat$Failure<mat$CenT,1,0)
mat$Failure1 <- ifelse(mat$Failure<mat$CenT,mat$Failure,mat$CenT)
mat$StudyT <- mat$Failure1+mat$T
data <- data.frame(mat$T,mat$Inci,mat$Trt,mat$Failure1,mat$Death,mat$StudyT)
colnames(data) <- c("StartT","Status","Trt","Failure","Death","StudyT")
data$Death4 <- ifelse(data$StudyT<t1 & data$Death==1,1,0)
data$StudyT4 <- ifelse(data$StudyT<t1,data$StudyT,t1)
data$Confirm <- rep(0, n1)
## Now confirmatory analysis ##
T2 <- runif(n2,0,t2)
c1 <- as.numeric(T2<=t2/3)
c2 <- as.numeric(T2>t2/3 & T2<=2*t2/3)
c3 <- as.numeric(T2>2*t2/3 & T2<=t2)
Cmat <- data.frame(T2,c1,c2,c3)
## Prev VS Inci of confirmatory analysis ##
for(i in 1:n2){
  Cmat$c11[i] <- ifelse(Cmat$c1[i]==1,rbinom(1,1,10/12),0)
  Cmat$c12[i] <- ifelse(Cmat$c2[i]==1,rbinom(1,1,11/12),0)
  Cmat$c13[i] <- ifelse(Cmat$c3[i]==1,rbinom(1,1,1),0)
```

```
}
## Treatment vs Control of confirmatory analysis ##
for(i in 1:n2){
  Cmat$c21[i] <- if(Cmat$c1[i]==1 & Cmat$c11[i]==1)Cmat$c21[i]=rbinom(1,1,0.5)
  else if(Cmat$c1[i]==1 & Cmat$c11[i]==0)Cmat$c21[i]=rbinom(1,1,0.5) else Cmat$c21[i]=0
  Cmat$c22[i] <- if(Cmat$c2[i]==1 & Cmat$c12[i]==1)Cmat$c22[i]=rbinom(1,1,0.5)
  else if(Cmat$c2[i]==1 & Cmat$c12[i]==0)Cmat$c22[i]=rbinom(1,1,0.5) else Cmat$c22[i]=0
  Cmat$c23[i] <- if(Cmat$c3[i]==1 & Cmat$c13[i]==1)Cmat$c23[i]=rbinom(1,1,0.5)
  else if(Cmat$c3[i]==1 & Cmat$c13[i]==0)Cmat$c23[i]=rbinom(1,1,0.5) else Cmat$c23[i]=0
}
## Hazard of confirmatory analysis ##
for(i in 1:n2){
  Cmat$c31[i] <-
  if(Cmat$c1[i]==1 & Cmat$c11[i]==0 &Cmat$c21[i]==1)Cmat$c31[i]=rexp(1,rate=exp(beta1))
  else if(Cmat$c1[i]==1 & Cmat$c11[i]==0 &Cmat$c21[i]==0)
  Cmat$c31[i]=rexp(1,rate=exp(beta2))
  else if(Cmat$c1[i]==1 & Cmat$c11[i]==1 &Cmat$c21[i]==1)
  Cmat$c31[i]=rexp(1,rate=exp(beta3))
  else if(Cmat$c1[i]==1 & Cmat$c11[i]==1 &Cmat$c21[i]==0)
  Cmat$c31[i]=rexp(1,rate=exp(beta4))
  else Cmat$c31[i]=0
  Cmat$c32[i] <-
   if(Cmat$c2[i]==1 & Cmat$c12[i]==0 &Cmat$c22[i]==1)Cmat$c32[i]=rexp(1,rate=exp(beta1))
  else if(Cmat$c2[i]==1 & Cmat$c12[i]==0 &Cmat$c22[i]==0)
  Cmat$c32[i]=rexp(1,rate=exp(beta2))
  else if(Cmat$c2[i]==1 & Cmat$c12[i]==1 &Cmat$c22[i]==1)
  Cmat$c32[i]=rexp(1,rate=exp(beta3))
  else if(Cmat$c2[i]==1 & Cmat$c12[i]==1 &Cmat$c22[i]==0)
  Cmat$c32[i]=rexp(1,rate=exp(beta4))
```

```
  else Cmat$c32[i]=0
  Cmat$c33[i] <-
   if(Cmat$c3[i]==1 & Cmat$c13[i]==0 &Cmat$c23[i]==1)Cmat$c33[i]=rexp(1,rate=exp(beta1))
  else if(Cmat$c3[i]==1 & Cmat$c13[i]==0 &Cmat$c23[i]==0)
  Cmat$c33[i]=rexp(1,rate=exp(beta2))
  else if(Cmat$c3[i]==1 & Cmat$c13[i]==1 &Cmat$c23[i]==1)
  Cmat$c33[i]=rexp(1,rate=exp(beta3))
  else if(Cmat$c3[i]==1 & Cmat$c13[i]==1 &Cmat$c23[i]==0)
  Cmat$c33[i]=rexp(1,rate=exp(beta4))
  else Cmat$c33[i]=0
}
Cmat$Inci <- as.numeric(rowSums(Cmat[,(5:7)])>0)
Cmat$Trt <- as.numeric(rowSums(Cmat[,(8:10)])>0)
Cmat$Failure <- rowSums(Cmat[,(11:13)])
Cmat$CenT <- rexp(n2,rate=exp(cen1))
Cmat$Death <- ifelse(Cmat$Failure<Cmat$CenT,1,0)
Cmat$Failure1 <- ifelse(Cmat$Failure<Cmat$CenT,Cmat$Failure,Cmat$CenT)
Cmat$StudyT <- Cmat$Failure1+Cmat$T2
Newdata <- data.frame(Cmat$T2,Cmat$Inci,Cmat$Trt,Cmat$Failure1,Cmat$Death,Cmat$StudyT)
colnames(Newdata) <- c("StartT","Status","Trt","Failure","Death","StudyT")
Newdata$Death4 <- ifelse(Newdata$StudyT<t1 & Newdata$Death==1,1,0)
Newdata$StudyT4 <- ifelse(Newdata$StudyT<t1,Newdata$StudyT,t1)
Newdata$Confirm <- rep(1,n2)
data1<- data.frame(Newdata$StartT,Newdata$Status,Newdata$Trt,Newdata$Failure,
Newdata$Death,Newdata$StudyT,Newdata$Death4,Newdata$StudyT4,Newdata$Confirm)
colnames(data1)<- c("StartT","Status","Trt","Failure","Death","StudyT",
"Death4","StudyT4","Confirm")
data <- rbind(data,data1)
I.data <- data[data$Confirm==0,]
```

```
C.data <- data[data$Confirm==1,]
D.data <- rbind(data[data$Confirm==1,],data[data$Confirm==0 & data$StudyT>t1,])
D.data$StartTNew <- ifelse(D.data$Confirm==1,D.data$StartT, t1-D.data$StartT)
D.data$StudyTNew <- ifelse(D.data$Confirm==1,D.data$StudyT4,D.data$Failure)
D.data$DeathNew <- ifelse(D.data$Confirm==1,D.data$Death4,D.data$Death)
D.data[D.data$Confirm==0,]$DeathNew <- ifelse(D.data[D.data$Confirm==0,]$StudyT>2*t1,
0,D.data[D.data$Confirm==0,]$Death)
D.data[D.data$Confirm==0,]$StudyTNew <- ifelse(D.data[D.data$Confirm==0,]$StudyT>2*t1,
D.data[D.data$Confirm==0,]$StartTNew+4,D.data[D.data$Confirm==0,]$Failure)
##Delayed entry for only 1 year
D1.data <- rbind(data[data$Confirm==1,],data[data$Confirm==0 & data$StudyT>t1,])
D1.data$StartTNew <- ifelse(D1.data$Confirm==1,D1.data$StartT, t1-D1.data$StartT)
D1.data$StudyTNew <- 0
D1.data[D1.data$Confirm==1,]$StudyTNew <-
ifelse(D1.data[D1.data$Confirm==1,]$StudyT<t2,
D1.data[D1.data$Confirm==1,]$StudyT, t2)
D1.data[D1.data$Confirm==0,]$StudyTNew <-
ifelse(D1.data[D1.data$Confirm==0,]$StudyT<t1+t2,
D1.data[D1.data$Confirm==0,]$Failure,D1.data[D1.data$Confirm==0,]$StartTNew+t2)
D1.data$DeathNew <- 0
D1.data[D1.data$Confirm==1,]$DeathNew <-
ifelse(D1.data[D1.data$Confirm==1,]$StudyT<t2,
D1.data[D1.data$Confirm==1,]$Death,0)
D1.data[D1.data$Confirm==0,]$DeathNew <-
ifelse(D1.data[D1.data$Confirm==0,]$StudyT<t1+t2,
D1.data[D1.data$Confirm==0,]$Death,0)

#Model fitting
```

```
ph.I1 <- coxph(Surv(I.data$StartT,I.data$StudyT4,I.data$Death4)~I.data$Trt)
ph.I2 <- coxph(Surv(I.data$StartT,I.data$StudyT4,I.data$Death4)~I.data$Trt+I.data$Status)
ph.I3 <- coxph(Surv(I.data$StartT,I.data$StudyT4,I.data$Death4)~I.data$Trt*I.data$Status)
ph.I4 <- coxph(Surv(I.data$StartT,I.data$StudyT4,I.data$Death4)~I.data$Status)
ph.C1 <- coxph(Surv(C.data$StartT,C.data$StudyT4,C.data$Death4)~C.data$Trt)
ph.C2 <- coxph(Surv(C.data$StartT,C.data$StudyT4,C.data$Death4)~C.data$Trt+C.data$Status)
ph.C3 <- coxph(Surv(C.data$StartT,C.data$StudyT4,C.data$Death4)~C.data$Trt*C.data$Status)
ph.C4 <- coxph(Surv(C.data$StartT,C.data$StudyT4,C.data$Death4)~C.data$Status)
ph.BD1 <- coxph(Surv(D1.data$StartTNew,D1.data$StudyTNew,D1.data$DeathNew)~D1.data$Trt)
ph.BD2 <- coxph(Surv(D1.data$StartTNew,D1.data$StudyTNew,D1.data$DeathNew)~D1.data$Trt+
D1.data$Status)
ph.BD3 <- coxph(Surv(D1.data$StartTNew,D1.data$StudyTNew,D1.data$DeathNew)~
D1.data$Trt*D1.data$Status)
ph.BD4 <- coxph(Surv(D1.data$StartTNew,D1.data$StudyTNew,D1.data$DeathNew)~D1.data$Status)
ph.D1 <- coxph(Surv(D.data$StartTNew,D.data$StudyTNew,D.data$DeathNew)~D.data$Trt)
ph.D2 <- coxph(Surv(D.data$StartTNew,D.data$StudyTNew,D.data$DeathNew)~D.data$Trt
+D.data$Status)
ph.D3 <- coxph(Surv(D.data$StartTNew,D.data$StudyTNew,D.data$DeathNew)~
D.data$Trt*D.data$Status)
ph.D4 <- coxph(Surv(D.data$StartTNew,D.data$StudyTNew,D.data$DeathNew)~D.data$Status)
return(c("Pval.I1"=summary(ph.I1)$logtest[3],
         "Pval.I2"=anova(ph.I2,ph.I4)[2,4],
         "Pval.I3"=anova(ph.I3,ph.I4)[2,4],
         "Pval.IT"=summary(ph.I3)$coef[3,5],
         "Pval.C1"=summary(ph.C1)$logtest[3],
         "Pval.C2"=anova(ph.C2,ph.C4)[2,4],
         "Pval.C3"=anova(ph.C3,ph.C4)[2,4],
         "Pval.CT"=summary(ph.C3)$coef[3,5],
         "Pval.BD1"=summary(ph.BD1)$logtest[3],
```

```
            "Pval.BD2"=anova(ph.BD2,ph.BD4)[2,4],
            "Pval.BD3"=anova(ph.BD3,ph.BD4)[2,4],
            "Pval.BDT"=summary(ph.BD3)$coef[3,5],
            "Pval.D1"=summary(ph.D1)$logtest[3],
            "Pval.D2"=anova(ph.D2,ph.D4)[2,4],
            "Pval.D3"=anova(ph.D3,ph.D4)[2,4],
            "Pval.DT"=summary(ph.D3)$coef[3,5],
            "Recurrence"=dim(data[data$Confirm==0 & data$StudyT>t1,])[1]
  ))
}

Proportion <- function(rep,n1,n2,t1,t2,beta1,beta2,beta3,beta4,cen1){
  ## First, replicate and then transpose it
  Total <- t(replicate(rep,SimV(n1,n2,t1,t2,beta1,beta2,beta3,beta4,cen1)))
  Total <- as.data.frame(Total)

  ## Dividing the data into 2 groups: ph holds, ph does not hold
  NonPH <- Total

  ## Stratifying Bigdat by three different models (nonph)
  Model1 <- NonPH[,c("Pval.I1.pvalue","Pval.C1.pvalue","Pval.BD1.pvalue","Pval.D1.pvalue")]
  Model2 <- NonPH[,c("Pval.I2","Pval.C2","Pval.BD2","Pval.D2")]
  Model3 <- NonPH[,c("Pval.I3","Pval.C3","Pval.BD3","Pval.D3")]

  ## Each case in Model1
  M1SI <- dim(Model1[Model1$Pval.I1.pvalue<0.05,])[1]
  M1SISC <- dim(Model1[Model1$Pval.I1.pvalue<0.05 & Model1$Pval.C1.pvalue<0.05,])[1]
  M1SISBD <- dim(Model1[Model1$Pval.I1.pvalue<0.05 & Model1$Pval.BD1.pvalue<0.05,])[1]
  M1SISD <- dim(Model1[Model1$Pval.I1.pvalue<0.05 & Model1$Pval.D1.pvalue<0.05,])[1]
```

```
M1powerC <- round(M1SISC/M1SI,3)
M1powerBD <- round(M1SISBD/M1SI,3)
M1powerD <- round(M1SISD/M1SI,3)

## Each case in Model2
M2SI <- dim(Model2[Model2$Pval.I2<0.05,])[1]
M2SISC <- dim(Model2[Model2$Pval.I2<0.05 & Model2$Pval.C2<0.05,])[1]
M2SISBD <- dim(Model2[Model2$Pval.I2<0.05 & Model2$Pval.BD2<0.05,])[1]
M2SISD <- dim(Model2[Model2$Pval.I2<0.05 & Model2$Pval.D2<0.05,])[1]

M2powerC <- round(M2SISC/M2SI,3)
M2powerBD <- round(M2SISBD/M2SI,3)
M2powerD <- round(M2SISD/M2SI,3)

## Each case in Model3
M3SI <- dim(Model3[Model3$Pval.I3<0.05,])[1]
M3SISC <- dim(Model3[Model3$Pval.I3<0.05 & Model3$Pval.C3<0.05,])[1]
M3SISBD <- dim(Model3[Model3$Pval.I3<0.05 & Model3$Pval.BD3<0.05,])[1]
M3SISD <- dim(Model3[Model3$Pval.I3<0.05 & Model3$Pval.D3<0.05,])[1]

M3powerC <- round(M3SISC/M3SI,3)
M3powerBD <- round(M3SISBD/M3SI,3)
M3powerD <- round(M3SISD/M3SI,3)

mat1 <- data.frame(c(M1powerC,M1powerBD,M1powerD),
                   c(M2powerC,M2powerBD,M2powerD),
                   c(M3powerC,M3powerBD,M3powerD))
```

```
  Recurrence <- c(ceiling(quantile(NonPH$Recurrence,0.05)),
  ceiling(mean(NonPH$Recurrence)),ceiling(quantile(NonPH$Recurrence,0.95)))

  mat1 <- rbind(mat1,Recurrence)

  colnames(mat1) <- c("Model1","Model2","Model3")
  rownames(mat1) <- c("PowerC","PowerBD","PowerD","Recurrence")
  return(mat1)

}

##For HR(AVE)=1.2

set.seed(4)
AVE12-11 <- Proportion(1000,1200,300,3,1,-1.7,-1.882,-2.7,-2.882,-2)
AVE12-12 <- Proportion(1000,1200,300,3,1,-1.7,-1.993,-2.7,-2.828,-2)
AVE12-13 <- Proportion(1000,1200,300,3,1,-1.7,-1.987,-2.7,-2.764,-2)
AVE12-14 <- Proportion(1000,1200,300,3,1,-1.7,-2.045,-2.7,-2.688,-2)
AVE12-15 <- Proportion(1000,1200,300,3,1,-1.7,-2.105,-2.7,-2.595,-2)
AVE12-16 <- Proportion(1000,1200,300,3,1,-1.7,-2.170,-2.7,-2.477,-2)
AVE12-17 <- Proportion(1000,1200,300,3,1,-1.7,-2.239,-2.7,-2.323,-2)
AVE12-18 <- Proportion(1000,1200,300,3,1,-1.7,-2.314,-2.7,-2.109,-2)
AVE12-19 <- Proportion(1000,1200,300,3,1,-1.7,-2.394,-2.7,-1.784,-2)
AVE12-20 <- Proportion(1000,1200,300,3,1,-1.7,-2.481,-2.7,-1.177,-2)
```

Appendix D

COEFFICIENTS USED FOR THE SIMULATIONS

Table D.1: Hazard(ET) indicates the hazard for early patients in the treatment group and Hazard(LC) indicates the hazard for late patients in the control group.

HR(WS)	**Hazard(ET)**	**Hazard(EC)**	**Hazard(LT)**	**Hazard(LC)**
1.2	0.183	0.084	0.067	0.308
	0.183	0.091	0.067	0.168
	0.183	0.099	0.067	0.121
	0.183	0.107	0.067	0.098
	0.183	0.114	0.067	0.084
	0.183	0.122	0.067	0.075
	0.183	0.129	0.067	0.068
	0.183	0.137	0.067	0.063
	0.183	0.145	0.067	0.059
	0.183	0.152	0.067	0.056
1.3	0.183	0.077	0.067	0.284
	0.183	0.084	0.067	0.155
	0.183	0.091	0.067	0.112
	0.183	0.098	0.067	0.090
	0.183	0.105	0.067	0.077
	0.183	0.112	0.067	0.069
	0.183	0.119	0.067	0.063
	0.183	0.126	0.067	0.058
	0.183	0.134	0.067	0.055
	0.183	0.141	0.067	0.052
1.4	0.183	0.072	0.067	0.264
	0.183	0.078	0.067	0.144
	0.183	0.085	0.067	0.104
	0.183	0.091	0.067	0.084
	0.183	0.098	0.067	0.072
	0.183	0.104	0.067	0.064
	0.183	0.111	0.067	0.058
	0.183	0.117	0.067	0.054
	0.183	0.124	0.067	0.051
	0.183	0.130	0.067	0.048
1.5	0.183	0.067	0.067	0.246
	0.183	0.073	0.067	0.134
	0.183	0.079	0.067	0.097
	0.183	0.085	0.067	0.078
	0.183	0.091	0.067	0.067
	0.183	0.097	0.067	0.060
	0.183	0.104	0.067	0.054
	0.183	0.110	0.067	0.050
	0.183	0.116	0.067	0.047
	0.183	0.122	0.067	0.045