LUND UNIVERSITY

Department of Statistics
Master's Thesis
Fall 2010

Surviving Breast Cancer

A study of cancer recurrence and mortality after mastectomy in postmenopausal women receiving radiotherapy and Tamoxifen

Authors:
Oxana Tarassiouk
Rebecca Rylance

Supervisor:
Björn Holmquist

Abstract

This paper analyzes mortality, cancer recurrence (loco-regional, other types of cancer and distant) and the impact of progesterone status on the effect of Tamoxifen using survival analysis methods. The data consists of 258 postmenopausal women from the Southern Sweden health care region with estrogen positive receptor status receiving radiotherapy, Tamoxifen and a combination of those after mastectomy. The period of the study is from 1978 to 2003. Three additional explanatory variables are included in the regression models: patient age, number of positive lymph nodes and tumor size.

The combined treatment is found to be significantly better than radiotherapy from the 6th year and on. The death rate is decreased by 36.5 %. The difference in cumulative incidence of loco-regional breast cancer recurrences is significant between the Tamoxifen vs. the combined group and the radiotherapy vs. the combined group. The lowest and highest cumulative incidence of loco-regional recurrences is found in the combined group and Tamoxifen group, respectively. For other types of cancer recurrence (non-breast related), the hazard is 3.27 times higher in the combined treatment group compared to the radiotherapy group. The effect of the treatment variable on distant cancer recurrences interacts with the age variable, i.e. age at the beginning of the study. One extra year of age results in a 10.4 % decrease of the hazard in the radiotherapy group, while the rate of distant recurrence for every additional year of age increases by 1.1 % in the combined group. The progesterone status for the survival rate of patients receiving Tamoxifen is significant. Progesterone status 400 days after the randomization date results in 49.7 % lower hazard in the progesterone receptor positive group compared to the progesterone receptor negative group.

Keywords: breast cancer, competing risks, Cox PH, radiotherapy, subdistribution hazard, survival analysis, Tamoxifen

Acknowledgments

We would like to thank Harald Anderson from the Department of Oncology, Lund University Hospital, for his assistance in data collection as well as taking time to explain the most important aspects of the original study.

Contents

1 Introduction

The first chapter provides a brief introduction to breast cancer. The purpose and review of previous studies follows.

1.1 Background

Breast cancer is the second most prominent cancer in the world after lung cancer and the most common cancer for females (IARC, 2008). It continues to affect women around the world even though the survival rate for breast cancer has increased over the past twenty years due to early detection.

Table 1.1 shows the top five estimated age-standardized[1] incidence and mortality rates as well as the absolute number of cases for cancer. Breast cancer is almost three times as common as its follower, cervical cancer, and represents 23 % of all new female cancers.

Cancer type	**Age-standardized rate**	**Number of cases**
INCIDENCE		
Breast	39.0	1 383 523
Cervix uteri	15.2	529 409
Colorectum	14.6	570 099
Lung	13.5	513 637
Stomach	9.1	349 042
All cancers	*164.9*	*6 038 545*
MORTALITY		
Breast	12.5	458 367
Lung	11.0	427 392
Cervix uteri	7.8	274 883
Colorectum	7.0	288 049
Stomach	6.9	273 634
All cancers	*87.6*	*3 345 834*

Table 1.1: Yearly age-standardized rate per 100 000 women and the absolute number of occurrences for cancer incidence and mortality worldwide among women (IARC, 2008).

Finding an appropriate treatment to fight breast cancer is of great importance. Different treatments are required, depending on how far the breast cancer has spread when discovered. There are four main characteristics of breast cancer that are used in order to define the stage and appropriate treatment. These are: invasive or non-invasive breast cancer[2], tumor size, number of metastatic lymph nodes and how far the cancer has spread.

[1] To adjust for different age distributions among countries, all the statistics are age-standardized, i.e. the numbers are counted as if all the countries had the same age structure. This is done in order to make different parts of the world compatible.

[2] The non-invasive breast cancer is confined within normal breast tissue, while the invasive breast cancer has spread beyond this barrier and can invade other parts of the body (Breast Cancer Organization, 2010).

Based on the characteristics described on the previous page, there are five types of treatments available (Breast Cancer Organization, 2010). These are:

1. Removal of the tumor with surgery, removal of the whole breast, or removal of the whole breast and the affected lymph nodes.
2. Radiotherapy: the usage of radioactive energy on the affected parts of the body in order to damage cancer cells.
3. Chemotherapy: a whole body medical treatment with the purpose of destroying cancer cells or at least making them incapable of growing and spreading.
4. Hormonal therapy: a medical treatment, e.g. Tamoxifen given to women with hormone-sensitive tumors.
5. Targeted therapy: a treatment that affects some of the cancer cells, disabling cancer growth and activity.

In the majority of cases, more than one of these treatments is considered.

1.2 Definitions

This section aims to give a brief description of definitions for cancer-related terms used in this paper.

Cancer: Cancer is a medical condition where a cell replicates uncontrollably and invades other tissues (American Cancer Society, 2010). Normal cells replicate as well, however in a controlled manner and do not invade other cells like cancer cells do.

Tumor: An abnormal cell growth concentrated in one place (National Cancer Insitute, 2010). Tumors can be cancerous or non-cancerous.

Breast cancer: In most kinds of breast cancer, the lobules, or ducts where breast milk is made is the primary location for the cancer (Breast Cancer Organization, 2010). Breast cancer can easily spread to the lymph system because the lymph vessels of the breast are connected to the lymph nodes under the arm, which are called the axillary lymph nodes.

Breast cancer stage II: There are five breast cancer stages (0, I, II, III and IV). Patients examined in the paper are diagnosed with stage II breast cancer, which is defined as one of the earlier stages and describes invasive cancer with no tumor, but cancer cells in the axillary lymph nodes or with a tumor up to 5 cm in size with possible cancer cells in the axillary nodes or a tumor larger than 5 cm, but no cancer cells in the axillary lymph nodes (Breast Cancer Organization, 2010).

Mastectomy: Breast removal surgery (Breast Cancer Organization, 2010).

Metastasis: Metastasis conveys the spread of disease or cancer from one organ to another non-adjacent organ or part (National Cancer Insitute, 2010).

Positive lymph nodes: A lymph node containing some cancer cells (Breast Cancer Organization, 2010). Also known as metastatic lymph nodes. Opposed to a negative lymph node, i.e. a node free from cancer.

Local cancer recurrence: Cancer reappears in the same site as the original tumor (MFMER, 2010).

Regional cancer recurrence: Regional recurrence occurs in the lymph nodes or in the same general area as the original cancer (MFMER, 2010).

Distant cancer recurrence: Cancer that has spread (metastasized) far away from where the original cancer appeared (MFMER, 2010).

Estrogen and progesterone: Primary female sex hormones (Breast Cancer Organization, 2010).

Hormone receptor: A protein that binds a hormone and can be found within or outside a cell (National Cancer Insitute, 2010). Estrogen and progesterone proteins can be found in female reproductive organs, some other tissues and some cancer cells.

Hormone receptor positive: Receptor positive cells have a protein to which the hormone can bind (National Cancer Insitute, 2010). If cancer cells are estrogen/progesterone positive (sometimes shortened as "estrogen positive" or "progesterone positive" in the paper), they may need estrogen/progesterone to grow. Hormone treatment can stop or reduce the growth.

Hormone receptor negative: Receptor negative cells' growth does not depend on hormones (National Cancer Insitute, 2010). Therefore, if the cancer cells are receptor negative, the hormone therapy is not effective.

Tamoxifen: A hormone therapy that controls estrogen's action and can block or at least reduce breast cancer cell growth (Breast Cancer Organization, 2010).

Radiotherapy: A treatment method that utilizes radiation in order to destroy cancer cells and reduce the tumor size (National Cancer Insitute, 2010).

1.3 Purpose

Three treatment arms from a randomized study of women in the Southern Sweden health care region with stage II breast cancer are analyzed from 1978 to 2003. All patients are postmenopausal (5 or more years of amenorrhea), younger than 72 years at the beginning of the study and have estrogen positive receptor status. These women have all undergone mastectomies and had various numbers of lymph nodes removed. The post-surgery treatments are radiotherapy, Tamoxifen and a combined treatment of radiotherapy and Tamoxifen.

The goal of this paper is to answer the three following questions:

1. **Which treatment is best in terms of survival: radiotherapy, Tamoxifen or a combination of them?**
   The overall mortality is death by any cause.
2. **Which of the treatments is best in terms of cancer recurrences?**
   The recurrences examined are local and regional breast cancer, other types of cancer (non-breast related cancer recurrence) and distant cancer.
3. **Does progesterone receptor status matter for survival of patients receiving Tamoxifen?**
   Tamoxifen is usually prescribed to patients with estrogen positive receptor status. The aim is to analyze whether progesterone status is a decisive factor for those

already receiving Tamoxifen.

1.4 Previous Studies

The goal of a study completed in the Southern Sweden health care region analyzed survival and recurrence in stage II breast cancer (Killander et al., 2007)[3]. The study had three treatment arms: radiotherapy, Tamoxifen and a combination of radiotherapy and Tamoxifen. According to the results of their study, survival was best in the Tamoxifen arm at twenty years whereas radiotherapy was best in the case of loco-regional recurrence. Systemic disease, which is defined as distant recurrence or death, was ten percent better in the combined group compared to radiotherapy alone.

A 2×2 factorial design study of breast-conserving surgery (BCS), BCS + radiotherapy, BCS + Tamoxifen and BCS + radiotherapy + Tamoxifen was done between 1991 and 1998 with 347 patients (Winzer et al., 2010). The results of this study are displayed for the treatment groups individually. A conclusion was that those who had only undergone BCS were three times as likely to suffer from a recurrence or die compared to the other groups. BCS alone seriously increased the rate of local recurrences, even for those patients who had a favorable prognosis. No differences could be found between the three groups with supplementary treatment for event free survival or distant disease free survival, defined as time to distant recurrence, another cancer recurrence or death.

The results of a Danish study found tumor size, number of positive lymph nodes and radiotherapy in addition to Tamoxifen to be prognostic covariates (Overgaard et al., 1999). The study was based on radiotherapy being beneficial in preventing loco-regional occurrences, but also wanted to evaluate the effects of adding radiotherapy to Tamoxifen for distant recurrence and overall survival. The incidence of loco-regional recurrences was significantly higher in the Tamoxifen group compared to the group treated with both Tamoxifen and radiotherapy. Distant recurrence was more prevalent in the combined group. After all recurrences were examined, the combined group had significantly higher disease free survival. Overall survival after ten years was 45 % in the radiotherapy + Tamoxifen group and 36 % for Tamoxifen alone.

In a population based study, progesterone was found to be independently statistically significant in both the whole cohort and estrogen receptor positive patients (Liu et al., 2010). Progesterone positive patients who used Tamoxifen had 24 % higher breast cancer survival than progesterone negative patients.

Another study focused on the interplay between progesterone and estrogen for postmenopausal women receiving Tamoxifen therapy (Lamy et al., 2002). The goal of the study was to find the level of expression in the hormone factors and evaluate the relationship between the levels of expression and mortality; the study defined 5 categories of estrogen and 3 categories of progesterone. Multivariate analysis found tumor size, number of lymph node status (positive or negative) and progesterone levels to be significant. The study shows higher levels of progesterone in estrogen positive women are associated with better survival.

[3]The data for the analysis in this paper is based on this study.

2 Survival Analysis

Chapter 2 discusses the theoretical background of the analysis techniques used in this paper. The presentation of survival analysis basics continues with a description of semi-parametric and parametric models along with their assumptions. Finally, alternative analysis strategies in the presence of competing risks are discussed.

2.1 Introduction and Basics

Survival analysis can be described as the time until an event occurs. Time can be measured in months, weeks, years etc. from the time the individual or interest of study starts to be under observation until the event transpires (Kleinbaum and Klein, 2005). The event is a changing of state like recovery, relapse from remission, conception or quitting smoking. The event in survival analysis can also refer to failure of some kind; two examples could be the recurrence of cancer or death.

2.1.1 Survivor and Hazard Functions

If T is the time until an event occurs, then survival time can be thought of in terms of t being a specific value of T. The distribution of T can be described by the probability density function, $f(t)$, the cumulative distribution function, $F(t)$, the survivor function, $S(t)$ and the hazard function, $h(t)$ (Cameron and Trivedi, 2005). The cumulative distribution function (cdf) describes the probability that a random variable T takes on a value less or equal to t. In other words, cdf represents the probability that an individual survives up to time t or exactly to time t. If T is a continuous random variable, then

$$F(t) = \Pr[T \leq t] = \int_0^t f(s)ds \tag{2.1}$$

The survivor function $S(t)$ is the probability that an individual survives beyond time t and is related to cdf $F(t)$ by

$$F(t) + S(t) = \Pr[T \leq t] + \Pr[T > t] = 1 \tag{2.2}$$

or more commonly

$$S(t) = 1 - F(t) \tag{2.3}$$

Therefore the survivor function and cumulative distribution functions are complementary.

The probability density function is simply the derivative of $F(t)$

$$f(t) = \frac{d}{dt}F(t) \tag{2.4}$$

The hazard function "gives the instantaneous potential per unit time for the event to occur given that the individual has survived up to time t" (Kleinbaum and Klein, 2005). In other words, the hazard function is the risk that a person has the event in the

next instant of time given that the person has not yet experienced the event (Collett, 1994). The hazard function is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr[t \leq T < t + \Delta t \ | T \geq t]}{\Delta t} = \frac{f(t)}{S(t)} \tag{2.5}$$

The relationship between the hazard function, $h(t)$ and the survivor function, $S(t)$ can also be expressed as

$$h(t) = -\frac{d}{dt} \ln(S(t)) \tag{2.6}$$

or,

$$S(t) = e^{-\int_0^t h(u)du} \tag{2.7}$$

which implies that knowing one of them is enough to obtain the other one (Cameron and Trivedi, 2005).

The survivor and hazard functions are the springboards of further analysis. Numerical or graphical summaries of the survival times of individuals are usually a first step in survival analysis. A special feature of this type of data is that it is not symmetrical in nature; it is often positively skewed, with a longer tail on the right (Collett, 1994). This makes traditional statistical tests difficult because many of them require the data be normally distributed.

2.1.2 Censoring

Censoring is a common feature of survival analysis and can be thought of as a timeline with the beginning of the timeline being when the individual entered the study. If the end-point of interest, the event, is not observed, then the event is said to be right-censored because it happens on the right side of the timeline (Collett, 1994). Right censoring can be either fixed or random (Cameron and Trivedi, 2005). Random censoring occurs when it is out of the control of the researcher. An example would be a patient moving away or when a patient stops coming to follow-up appointments. The end of study is an example of fixed right censoring.

Left censoring is far less common. An example is when the patient's official examination occurs three months after treatment and cancer is detected, but the actual time that the cancer appeared is not known. Therefore it is censored on the left side of the timeline (Collett, 1994).

Further discussion of the methodology is limited to right censoring.

2.2 Kaplan-Meier Estimator

The Kaplan-Meier (KM) approach is commonly used when estimating the survivor function. The survival probability at time t is a conditional probability of surviving beyond time t, given survival to at least time t (Kim, 2007). The KM estimate at time t is a product of all the estimated conditional probabilities for failure time t and earlier (Kleinbaum and Klein, 2005). The KM survival probability is calculated at each failure time in the sample and plotted in a graph where the x-axis is the timeline and the y-axis

is the estimated survival probability. Therefore the Kaplan-Meier is a decreasing step function where each failure or each event is a step down (Cameron and Trivedi, 2005).

If the event times are ordered such that $t_1 < t_2 < ... < t_j < ... < t_k$ in a sample of size N and $N \geq k$, the KM estimated conditional survival probability at time j is defined as

$$\hat{S}(t_j) = \prod_{i=1}^{j} \hat{\Pr}[T > t_i | T \geq t_i] = \prod_{i=1}^{j} \frac{n_i - m_i}{n_i} \tag{2.8}$$

where m_i is the number of events at time t_i for $i = 1, .., k$; n_i is the number of subjects "at risk" just prior to time t_i, i.e. the number of subjects that have not been censored or failed up until the moment right before time t_i for $i = 1, .., k$ (Cameron and Trivedi, 2005).

To evaluate if two or more KM curves are statistically equivalent, the log-rank test is used (Kleinbaum and Klein, 2005). The test uses observed versus expected cell counts and under the null hypothesis, the log-rank statistic is chi-squared distributed with [the number of groups being compared minus one] degrees of freedom.

Other tests like Wilcoxon and Tarone-Ware are used if the strongest effect of the treatment is expected in the beginning of the timeline and weighs early failures more heavily (Kleinbaum and Klein, 2005). The Flemington-Harrington test lets the tester choose how to distribute the weights in the test.

2.3 Cox Proportional Hazards Model

The Cox proportional hazards (PH) model is a commonly used survival regression method for estimating the relationship between survival and one or several explanatory variables (Kleinbaum and Klein, 2005). The general formula for the Cox PH model is given by the hazard function

$$h(t, \mathbf{X}) = h_0(t) e^{\sum_{i=1}^{p} \beta_i X_i} \tag{2.9}$$

where $\mathbf{X}$ indicates a vector of predictor variables $(X_1, X_2, ..., X_p)$ and t indicates time. The baseline hazard function, $h_0(t)$, is a function of time and is unspecified, which makes the model semi-parametric because the survival times do not have any predescribed distribution. The exponential expression of $\beta_i X_i$ represents the sum over the p explanatory variables where the X_i's are time-independent, which signifies that the values do not change over time. Some examples of such variables are sex or smoking status. If all the X_i's would be zero, the formula becomes the baseline hazard function. If there are no explanatory variables in the model, then the model also reduces to the baseline hazard function, which would be a function of time (Kleinbaum and Klein, 2005).

2.3.1 Parameter Estimation and Hazard Ratio

The parameters of the Cox PH model, the β_i's, are estimated by using the method of maximum likelihood (Collett, 1994). The maximum likelihood estimates are derived by maximizing the natural logarithm of the likelihood function, L, with respect to each β in the model, and solving a system of equations using iteration (Kleinbaum and Klein,

2005). The likelihood function for the Cox PH model (see Equation 2.10) is a partial likelihood function because it includes only those who fail, although those subjects who were censored are included in the risk set up until they are censored. The function uses ordered survival times and does not take into account the actual event times.

$$L = \prod_{j=1}^{k} \frac{e^{\beta' \mathbf{X}_j}}{\sum_{l \in R(t_j)} e^{\beta' \mathbf{X}_l}} \tag{2.10}$$

In the formula above the k ordered event times are denoted by $t_1 < t_2 < ... < t_j < ... < t_k$; $\mathbf{X}_j$ denotes the vector of covariates for the individual who experienced the event at t_j; $R(t_j)$ indicates subjects at risk at t_j (Collett, 1994).

If two or more subjects experience an event at the same time, Equation 2.10 should be modified. The description of the modification methods is beyond the scope of this chapter. By default, SAS uses Breslow's method for ties, but in order to get more precise results an exact computation method should be used (Kleinbaum and Klein, 2005). Although, it is preferred to Breslow's method if a large number of events occur simultaneously. Otherwise, the difference between the methods is insignificant.

In combination with the parameter estimates, the hazard ratios are used to make statistical conclusions. The hazard ratio (HR) is the hazard function of one individual divided by the hazard function of another individual, usually with a unit difference in the explanatory variable (Singer and Willett, 2003).

$$\widehat{HR} = \frac{h_0(t)e^{\widehat{\beta} X^*}}{h_0(t)e^{\widehat{\beta} X}} = e^{\widehat{\beta}(X^* - X)} = e^{\widehat{\beta}(1-0)} = e^{\widehat{\beta}} \tag{2.11}$$

The hazard ratio does not involve time because the baseline hazards cancel out. As shown in Equation 2.11, the final expression of the estimate of HR is the exponential of the estimated regression coefficient and therefore cannot be negative (Collett, 1994). In Equation 2.11, the value of the explanatory variable for the first individual is denoted as X^* and the corresponding value for the second individual is denoted as X. Note that the values of 0 and 1 are just examples and can be replaced by any values appropriate for the covariate of interest.

In the comparison of two groups (i.e. the variable is dummy coded) and if $\widehat{HR}$ is less than one, then the group coded as 1 has a lower hazard than the group coded as 0 (Kleinbaum and Klein, 2005). If $\widehat{HR}$ is greater than one, the hazard of experiencing the event is higher for the group coded as 1. The $\widehat{HR}$ of one implies no difference between the groups. For continuous variables, the interpretation of the $\widehat{HR}$ refers to the change in one unit.

A $(100 - \alpha)$ % confidence interval (CI) is obtained by exponentiating the confidence limits of $\widehat{\beta}$ (Collett, 1994).

$$\text{CI for } \widehat{HR} = e^{\widehat{\beta} \pm z_{\alpha/2} SE_{\widehat{\beta}}} \tag{2.12}$$

2.3.2 Evaluating Proportional Hazards Assumption

The Cox model is called the proportional hazards model because there is an assumption of proportional hazards, which means that the hazard ratio should be time-independent (Kleinbaum and Klein, 2005). The assumption check can be performed both graphically and with statistical tests. It is best to assess whether the PH assumption is met by applying at least two techniques.

One graphical method is the log-log plot, uses the KM survival probability by taking the double natural logarithm of the estimates, $\ln(-\ln(\widehat{S}))$ and plotting it against the natural logarithm of time (Kleinbaum and Klein, 2005). Since a probability lies between 0 and 1 it will always produce a negative value when taking the natural logarithm. It is therefore necessary to put a minus sign in front of the first logarithm. If the PH assumption is met, the lines will be approximately parallel.

The log-log plots can be applied to variables with only a few levels, for example the variable gender, which has two categories. If the variables are continuous or contain many levels, other methods are recommended (Kleinbaum and Klein, 2005).

One of the methods for evaluating the PH assumption that works for all types of variables is the Schoenfeld residuals plot (Kleinbaum and Klein, 2005). The Schoenfeld-type residuals, also known as partial residuals, are defined separately for each explanatory variable in the model and every individual who has experienced the event. The residuals for each covariate should be plotted against time with a smooth line (Singer and Willett, 2003). A horizontal line, i.e. no tendency to increase or decrease across time, indicates that the assumption is met.

An alternative method for assessing the PH assumption uses the Cox PH model by including interactions of the explanatory variables with time in the model. The assumption is violated if one or more interactions are found to be statistically significant (Scrucca et al., 2010).

If the PH assumption is not met and the variable that violates the assumption is categorical or can be categorized, it is possible to stratify that variable. This is called the stratified Cox procedure (Kleinbaum and Klein, 2005). In an example where there are three predictor variables: treatment, smoking and marital status and treatment and marital status are assumed to have proportional hazards, but the smoking variable does not, treatment and marital status are included in the model and smoking is the stratified variable. As a result of this, there will be two baseline functions, one for each smoking status, smoker and nonsmoker. The estimated coefficients, standard errors and hazard ratios are available for both the treatment and the marital status, but not for the smoking variable because it is not included in the model. As a consequence, no conclusions can be made concerning the hazards of smoker and nonsmoker groups when compared to one another.

If there is more than one variable not fulfilling the PH assumption or the variable violating the assumption is of particular interest, an extended Cox model should be considered (Singer and Willett, 2003). Such a model allows time-dependent variables as predictors. Possible extensions are testing various functions of time in combination with

the explanatory variable or splitting the timeline into intervals if the variable's effect differs between periods.

2.4 Parametric Models

Parametric models differ from semi-parametric or non-parametric models in that the outcome (i.e. time in the survival models) follows a specific distribution (Kleinbaum and Klein, 2005). Parametric models can be either proportional hazard (PH) models or accelerated failure time (AFT) models. The most widely used parametric models in survival analysis are the Weibull, exponential, Gompertz, log-logistic, gamma and lognormal.

The parametric PH models are similar to the Cox PH model (Bradburn et al., 2003). Both models produce similar results and the meaning of the of the hazard ratios is the same. The main difference is that in parametric PH models the survival times follow a certain distribution whereas in Cox PH model the distribution is unspecified. Parametric models are also more accurate than the Cox PH model because they yield smaller standard errors, but ascertaining the distribution of the survival times may be difficult.

While the effect of the covariates in PH models is related to hazard ratios, the estimates of AFT models are thought of in terms of survival times (Kleinbaum and Klein, 2005). The underlying assumption of AFT models can be illustrated as

$$S_2(t) = S_1(\gamma t) \tag{2.13}$$

for $t \geq 0$; $S_1(t)$ is the survivor function for one group and $S_2(t)$ is the survivor function for another group. The accelerator factor, γ, is a ratio of survival times in relation to any fixed value of $S(t)$. If the AFT assumption holds, γ is constant over time.

The accelerator factor can be parameterized as a function of predictor variables (Cameron and Trivedi, 2005).

$$\gamma = e^{\sum_{i=1}^{p} \beta_i X_i} \tag{2.14}$$

for $(X_1, X_2, ..., X_p)$ explanatory variables.

The Akaike's information criterion is used as a goodness-of-fit statistic for choosing the best parametric model (Kleinbaum and Klein, 2005).

2.5 Competing Risks

Competing risks occur when there are more than one type of event an individual can experience, but it is only possible to fail from one of these events (Kleinbaum and Klein, 2005). An example might be dying from cancer, a heart attack, or another reason.

The competing risks methodology is also used for events that can happen more than once, e.g. disease relapse, or when one event is not excluding the other one, e.g. having a stroke and having a heart attack. In this case, the focus is put on the first failure of the event that occurs first (Kleinbaum and Klein, 2005).

2.5.1 Cumulative Incidence Estimation

Cumulative incidence (CI) of an event can be thought as an alternative way to describe the survival probability. Cumulative incidence is defined as the probability of experiencing the event (Kleinbaum and Klein, 2005).

If there is only one type of event, CI can be simply calculated by subtracting the Kaplan-Meier (KM) estimated survival probability from one. In the case when there are more than one event types, the KM method may not be the most informative because it assumes independent censoring (Putter et al., 2007). A practical example could be a disease recurrence and death as its competing risk. The KM approach would treat both those who did not experience the event of interest (i.e. recurrence) and those who died as censored observations. It is known that those who died can not develop a recurrence and therefore the risk of it is equal to 0, while those who are "truly" censored might still experience a recurrence and their risk is greater than 0. Hence, the result would be biased. Another notable point is that even if the event of interest is recurrence, death as a competing event is still of great importance for researchers and patients (Kim, 2007).

An alternative to the KM approach uses marginal probabilities instead of conditional probabilities (Kleinbaum and Klein, 2005). If the event times are ordered as $t_1 < t_2 < ... < t_j < ... < t_k$, the estimated cumulative incidence for event of type c is defined as

$$\widehat{CI}_c(t_j) = \sum_{i=1}^{j} \widehat{S}(t_{(j-1)}) \frac{m_{cj}}{n_j} \tag{2.15}$$

where m_{ci} is the number of events of type c at time t_i for $i = 1, .., k$, n_i is the number of subjects at risk just prior to time t_i and $S(t_i)$ is the overall survival probability at time i estimated using the KM method[1].

By calculating the cumulative incidence for each cause of failure and plotting it against time, different failure types can be compared to one another visually (Scrucca et al., 2010).

Gray's test is used to evaluate whether the CI curves of a particular event type for two or more groups are significantly different (Kleinbaum and Klein, 2005). The test compares weighted averages of the cumulative incidence hazards. The test statistic is chi-squared distributed with [the number of groups minus one] degrees of freedom under the null hypothesis (Gray, 1988).

2.5.2 Regression Models for Competing Risks Data

There are different analysis methods to be explored when competing risks are an issue. One method is a cause-specific Cox PH, which produces a separate model for each various event type (Kleinbaum and Klein, 2005). This method is not recommended if the competing risks are dependent because only the event of interest is counted while individuals who experience the competing risks are censored along with individuals who dropped out of the study or did not experience the event until after the end of the study.

[1] i.e. the probability of not experiencing any of the competing events beyond time i, given survival to time i.

As a consequence, the results might be misleading. Moreover, the estimated effect of covariates using cause-specific Cox PH can be very different from the estimated effects when the competing risks are included in the regression (Kim, 2007).

An alternative approach developed by Lunn and McNeil (1995) produces only one Cox PH model where [number of competing events minus one] dummy variables are created in order to separate different event types. Although even this approach requires competing risks to be independent, which is rarely the case.

Another helpful method to find prognostic factors for each type of failure, proposed by Fine and Gray (1999), is a Cox PH model for the cumulative incidence of the competing events discussed in Section 2.5.1. The regression estimates the effect of covariates on the subdistribution hazard[2] of each event type. The subdistribution hazard for failure of type c, denoted as h_c^*, is defined as the hazard of experiencing a particular event in the presence of competing risks given that the individual is event-free or already experienced one of the competing events, other than c.

$$\begin{aligned} h_c^*(t) &= \lim_{\Delta t \to 0} \frac{\Pr[t \leq T < t + \Delta t, C = c \,|T \geq t \cup (T \leq t \cap C \neq c)]}{\Delta t} \\ &= -\frac{d}{dt} \ln(1 - F_c(t)) \end{aligned} \tag{2.16}$$

where

$$F_c(t) = \Pr[T \leq t, C = c] \tag{2.17}$$

for $c = 1, ..., r$ competing events (Fine and Gray, 1999).

Therefore, the main difference of this approach compared to the cause-specific Cox PH is that the individuals who fail from other events than the event of interest are not removed from the risk set (Putter et al., 2007).

The general formula for the subdistribution PH model for failure type c is given by the subdistribution hazard function

$$h_c^*(t, \mathbf{X}) = h_{c0}^*(t) e^{\sum_{i=1}^{p} \beta_i X_i} \tag{2.18}$$

where $\mathbf{X}$ is a vector of p covariates $(X_1, X_2, ..., X_p)$.

The parameter estimation and PH assumption tests for the subdistribution PH model are identical to the standard Cox PH, previously described in Sections 2.3.1 and 2.3.2 (Scrucca et al., 2010).

[2] In the original article, cumulative incidence for a specific event among other competing events is referred to as subdistribution hazard.

3 Data and Method

Chapter 3 discusses the data in great detail and provides some descriptive statistics. The final part of the chapter relays the applied analysis methods.

3.1 Data Sources and Modifications

The data in this study comes from a study completed in the Southern Swedish health care region. The original data set includes 713 patients randomized to the following post-surgery treatments: radiotherapy, Tamoxifen and the combination of Tamoxifen and radiotherapy (more information on the treatments details, surgery and other decisions made about the medical design of the study can be found in the original article by Killander et al. (2007)).

Tamoxifen was wrongly prescribed to estrogen negative patients during the time when the study was initiated. The effect of radiotherapy does not depend on the hormone status and therefore, in order to avoid underestimating Tamoxifen's effect, 383 observations with negative or unknown estrogen status are excluded.

One of the entry criteria for the study is that the women have been amenorrheic for five years. In the original study by Killander et al. (2007), the researchers include women who do not fulfill the entry criteria of cessation of menstruation for five years. In this analysis, these individuals are excluded as well as those whose menopausal state is questionable. Hence, an additional 53 observation are removed from the data set.

The subsequent 7 exclusions are from patients with breast cancer stages one, three and four. Twelve observations where the actual treatment dates were not registered are removed because of the uncertainty that the patients received the treatment they were prescribed. In total, 258 observations are included in the analysis.

In five cases, the received treatments are different from the planned ones. The discrepancies are also adjusted. Two radiotherapy patients received the combined treatment, one Tamoxifen patient received the radiotherapy instead and one of the two patients in the combined randomization group received radiotherapy while the second one received just Tamoxifen.

3.2 Data Characteristics

The start time of the trial is the treatment randomization date, which is commonly used in clinical studies (Anderson, 2010). It is from this point that the patients are comparable, i.e. possible differences are random. The women were studied a maximum of 25 years; some moved away, some were censored and some women died.

Besides the variable of interest, i.e. treatment, there are three explanatory variables: age, tumor size and number of positive lymph nodes.

In the beginning of the study, the median age for all the patients is 63 years, the median tumor size is 25 mm and the median number of positive lymph nodes is 1. The data is quite balanced for the three treatment groups (see Table 3.1). The age is known for all the patients. The number of positive lymph nodes is not known for two patients

Variable	RT ($n = 82$)	Tam ($n = 78$)	RT+Tam ($n = 98$)
Age (years)			
1st quartile	59	58	59
Median	63	62	63
3rd quartile	66	64	67
Range: 49-71 years			
Tumor size (mm)			
1st quartile	20	20	20
Median	25	25	22
3rd quartile	30	34	28
Range: 4-50 mm			
Number of positive lymph nodes			
1st quartile	0	0	0
Median	1	1	1
3rd quartile	2	3	3
Range: 0-20 lymph nodes			

Table 3.1: Data characteristics for 258 estrogen positive postmenopausal patients receiving radiotherapy (RT), Tamoxifen (Tam) and a combination of those (RT+Tam).

in the radiotherapy group and 1 patient in the Tamoxifen group. The tumor size is missing in 11 cases: 4 for radiotherapy, 4 for Tamoxifen and 3 for the combined group.

The progesterone status is determined for 64 patients within the Tamoxifen group. Descriptive statistics for each progesterone status group (see Table 3.2) are relatively proportional with somewhat larger tumor sizes in the receptor negative group. The number of positive lymph nodes is missing for 1 receptor negative patient. The tumor size is not registered for 1 progesterone negative and 2 progesterone positive patients.

Variable	PR+ ($n = 37$)	PR- ($n = 27$)
Age (years)		
1st quartile	59	59
Median	63	62
3rd quartile	65	64
Range: 53-71 years		
Tumor size (mm)		
1st quartile	20	20
Median	25	30
3rd quartile	30	37
Range: 10-50 mm		
Number of positive lymph nodes		
1st quartile	0	0
Median	1	1
3rd quartile	4	3
Range: 0-12 lymph nodes		

Table 3.2: Data characteristics for 64 patients in the Tamoxifen group with positive (PR+) or negative (PR-) progesterone status.

3.3 Analysis Methodology

As mentioned earlier, women were assigned to three treatment groups: radiotherapy, Tamoxifen and the combined therapy. The effects of Tamoxifen can be evaluated when the combined therapy is compared with radiotherapy. The effects of radiotherapy can also be compared in a similar manner when considering Tamoxifen against the combined group. Hence, the two estimated models are radiotherapy vs. combined, where radiotherapy is coded as 0 and the combined treatment is coded as 1 and Tamoxifen vs. combined, where Tamoxifen is coded as 0 and combined is coded as 1.

3.3.1 Overall Mortality Analysis

The overall mortality is analyzed for 258 patients in the three treatments groups. It is also analyzed in the Tamoxifen group separately for 64 patients comparing the positive and the negative progesterone status.

The analysis is performed in SAS using PROC LIFETEST in order to obtain the Kaplan-Meier (KM) plot and perform the log-rank test. The KM plots are calculated for only the treatment arms, irrelevant of other factors. Hence, further analysis is needed to be able to gain better knowledge of the data. The log-rank test is used to compare factor variables, i. e. treatment in the current analysis, to see if there are differences between groups. A typical significance level for rejecting the null hypothesis of no differences is 20-25 %.

PROC PHREG is used to estimate the Cox PH model[1]. At first, all of the explanatory variables are included in the model. Those variables that are not significant at the 5 % level of significance are removed one at a time beginning with the one that has the highest p-value. Once all the remaining explanatory variables are significant, all possible interactions between the variables are checked and the ones that are significant, if any, are included in the model.

If there is more than one model that seems to be appropriate, the goodness-of-fit statistics are used in choosing the best one. The Bayesian information criterion (BIC) is used because it takes into account both the number of parameters and sample size (Cameron and Trivedi, 2005).

$$BIC = -2 \ln L + d \ln n$$

where d is the number of estimated parameters and n is the number of observations.

The first test of PH assumption is the log-log plot. It is only possible for variables with a interpretable number of levels and therefore is only performed for treatment. The log-log curves are created using PROC LIFETEST in SAS and should be proportional if the assumption is met. The second check of the PH assumption tests all the variables in the chosen model for time-dependence. The procedure is performed in PROC PHREG by estimating the Cox PH model where each of the variables is included by itself and

[1]Some standard PH parametric models are estimated with PROC LIFEREG, but since the Cox PH model produces similar results to the PH parametric models without having to determine the distribution of survival times, the analysis is limited to the Cox PH regression.

with an interaction with time. If no significant time-dependent variables are detected, the model fulfills the PH assumption. An interpretation of the hazard ratios along with their confidence intervals follows.

The graphs included in the paper are recreated in R software working with the function `survfit` in the `survival` library.

3.3.2 Cancer Recurrence Analysis

In the analysis of breast cancer recurrences, there are three combinations of events that are considered competing events (see Table 3.3). Local and regional recurrences are combined into one group because of the low incidence number (16 local recurrences as a first event and 10 regional recurrences as a first event).

Event of interest	**Competing event**
Loco-regional breast recurrence	Distant recurrence or death
Other type of cancer recurrence	Distant recurrence or death
Distant recurrence	Death

Table 3.3: Type of the recurrence studied and its competing risk.

The choice of the competing event can be explained by the severity of the recurrence, where distant cancer is obviously worse than loco-regional cancer or other types of cancer recurrences since distant cancer has the potential to affect the whole body. The worst case, in other words death, is included as the competing risk in all of the combinations. Whether the other types of cancer recurrence (non-breast related) or loco-regional breast cancer recurrence is more serious is hard to determine and therefore these two types are studied separately.

The analysis of cancer recurrence is performed in R software in the `cmprsk` library using `cuminc` and `crr` functions. The analysis begins with the estimation of the cumulative incidence of the recurrence within each treatment group along with its competing risk. Gray's test compares the cumulative incidence of the treatment groups and rejects the null hypothesis of no difference between the groups at the 20-25 % significance level. The multivariate subdistribution PH model for the event of interest including all the explanatory variables continues with excluding the variables that are not significant at the 5 % level one at a time beginning with the one that has highest p-value. The interaction check and the choice of the best model are made in a similar manner as in the overall mortality analysis (see Section 3.3.1). The PH assumption is checked with the help of the Schoenfeld residuals plot for each covariate in the model. No tendency to increase or decrease with time indicates the adequacy of the model. An additional assumption check tests the time-dependence of the variables. None of the interactions with time should be significant. If the PH assumption is fulfilled, different models are produced for each specific competing event to evaluate the effect of the explanatory variables on the event of interest and compare it with the effect of the competing event.

4 Results

This chapter discusses the results in the same order as in the Purpose Section (Section 1.3), i. e. overall mortality, which means death independent of its cause, analysis of recurrences and finally the evaluation of progesterone's impact on mortality in the Tamoxifen treatment group. The most important computation results are included in the chapter. For complete results see the Appendices.

4.1 Overall Mortality

Starting with a Kaplan-Meier (KM) plot, patients within each treatment group have approximately the same survival probability until the 6th year (see Figure 4.1), after which the radiotherapy group has a slightly worse survival probability than the other groups. The probability of survival is approximately 50 % and Tamoxifen becomes the best treatment while radiotherapy becomes the worst at 14 years.

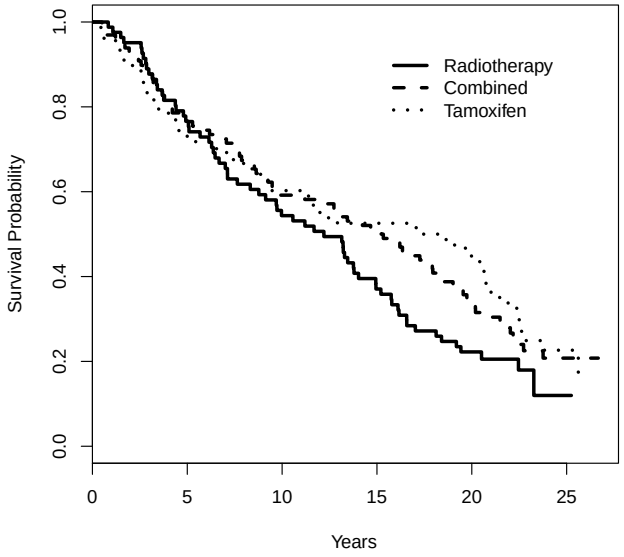

Figure 4.1: Kaplan-Meier plot for radiotherapy, Tamoxifen and combined treatment groups.

The results of the log-rank test indicate no significant difference between Tamoxifen and the combined treatment groups ($p = 0.6286$). The null hypothesis of no differences between radiotherapy and the combined treatment is rejected at the 15 % significance level ($p = 0.1279$).

Further analysis is performed first for Tamoxifen vs. the combined therapy and then for radiotherapy vs. combined. The results for the Cox PH model selection, interaction check and the assumption check for overall mortality models can be found in Appendix A.

Tamoxifen vs. combined treatment

The multivariate check of the Cox PH model including all four variables, i.e. treatment,

age, number of positive lymph nodes and tumor size indicates that treatment is highly insignificant ($p = 0.98$), confirming the results of the log-rank test. There is no statistical evidence that the effect of Tamoxifen on the overall survival varies from the combined treatment. Therefore further investigation is discontinued.

Radiotherapy vs. combined treatment

In the multivariate model, which includes all four variables, age is not significant ($p = 0.19$), so it is dropped from the model. The next step is to re-estimate the model without age. All of the variables are significant at the 5 % level. None of the possible interaction terms (i.e. interaction of treatment and positive lymph nodes, interaction of treatment and tumor size and interaction of positive lymph nodes and tumor size) are found to be statistically significant. The model which only includes significant coefficients is reffered to as model A in Table 4.1.

In Figure 4.1, radiotherapy and the combined treatment curves cross one another a few times until the 6th year. Thus, a model where treatment is divided into two periods is examined. Two new variables are created. The first one is denoted as "treatment, $t <$ y5" and equals the original treatment variable for all the patients censored or failed up to the 6th year, otherwise it is equal to zero. The second variable is denoted as "treatment, $t \geq$ y5" and is constructed in the same way. It equals zero for all the patients censored or failed up to the 6th year while for the rest of the observations it is equal to the initial treatment variable. Treatment, $t <$ y5 is not significant ($p = 0.69$). In the next step of the model selection age is also dropped ($p = 0.21$). The rest of the variables are significant and this model is called model B.

Table 4.1 includes the two final model candidates. Please note that model A does not include the treatment variable divided into two periods while model B does not include the original treatment variable.

	MODEL A		**MODEL B**	
Variable	**Hazard ratio (95 % CI)**	***p*-value**	**Hazard ratio (95 % CI)**	***p*-value**
Treatment	0.704 (0.499,0.993)	0.046		
# lymph nodes	1.098 (1.047,1.151)	0.000	1.092 (1.042,1.144)	0.000
Tumor size	1.023 (1.000,1.046)	0.046	1.023 (1.001,1.046)	0.043
Treatment, $t \geq$ y5			0.635 (0.420,0.962)	0.032
BIC	*1208.549*		*1207.967*	

Table 4.1: Cox PH model for overall mortality: radiotherapy vs. combined treatment groups

The significance of treatment in the second interval (6th year and on) in the time extended model has a lower p-value than in model A including treatment for the whole period ($p = 0.032$ vs. $p = 0.046$). Moreover, model B has the lowest Bayesian Information Criterion (BIC) value and is considered to be the best choice.

The log-log plot, which serves as a graphical check of the proportional hazards (PH) assumption, is only applicable for factor variables; therefore only treatment is tested. The curves are somewhat parallel after year 6, which suggests that treatment ($t \geq$ y5) in model B satisfies the PH assumption. The test of the time-dependence of the variables in model B does not reject the null hypothesis of proportionality and supports

the result of the adequacy of the model.

The interpretation of the obtained estimates is as follows: the hazard ratio for treatment, which is 0.635 (see Table 4.1), indicates that the hazard for the radiotherapy group is 1/0.635 = 1.57 times the combined treatment group after the sixth year. In other words, if treatment changes from radiotherapy to combined, the rate of death at the 6th year and on decreases by (100 % - 63.5 %) = 36.5 %. If all other variables are held constant, then for each additional positive lymph node, the rate of death increases by 9.2 % and if all other variables are held constant for every additional millimeter of tumor, the rate of death increases by 2.3 %.

The 95 % confidence interval for the hazard ratio is (0.420,0.962) for treatment variable in Model B, which surrounds the point estimate of 0.635. This is quite a wide interval, which reveals that the point estimate for treatment is not precise. In light of the 95 % confidence interval, the combined therapy would result in a decreased death rate between 3.8 to 58 % compared to radiotherapy after the 6th year.

4.2 Recurrences

Various types of cancer recurrence are evaluated in the following sections.

4.2.1 Loco-regional Breast Cancer

In the analysis of loco-regional breast cancer, the competing risk is distant recurrence or death. The highest cumulative incidence of loco-regional breast cancer recurrence occurs in the Tamoxifen group (see Figure 4.2). The lowest cumulative incidence of the competing risk also occurs in the Tamoxifen group. This suggests that Tamoxifen is better at preventing distant cancer and death, but is more responsible for loco-regional recurrences.

Gray's test results in Table 4.2 indicate highly significant differences between Tamoxifen and the combined group cumulative incidence curves for loco-regional recurrence and its competing risk. But yet again, the high rate of loco-regional recurrences can be a consequence of a low incidence of distant cancer or death.

Fail cause	**RT vs. RT+Tam**	**Tam vs. RT+Tam**
LR recurrence	0.1577	0.0000
Competing event	0.2435	0.0121

Table 4.2: Results of Gray's test (p-values).

The difference in cumulative incidence for the radiotherapy and the combined treatment is significant at the 20 % level for loco-regional recurrences and at the 25 % for its competing event.

The number of patients with loco-regional recurrences as a first event (5 for radiotherapy, 2 for combined and 19 for Tamoxifen) is too small for any further analysis.

4.2.2 Other Types of Cancer

The cumulative incidence of other types of cancer recurrences is the lowest in the radiotherapy group (see Figure 4.3), but at the same time this group has the highest

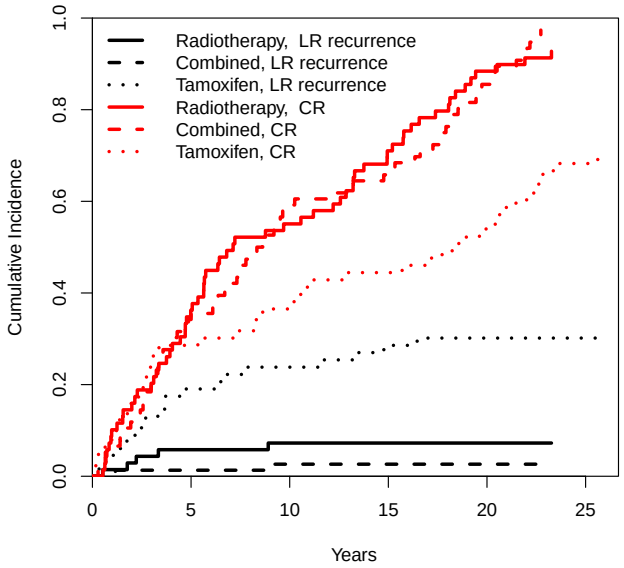

Figure 4.2: Loco-regional (LR) recurrence vs. competing risk (CR), i.e. distant recurrence or death

cumulative incidence of the competing event, which is distant recurrence or death. For the other treatment groups, the cumulative incidence curves are very close to one another and cross a few times.

The results of Gray's test display significant differences between the radiotherapy treatment and the combined treatment in both other cancer recurrence and its competing event (see Table 4.3).

Fail cause	**RT vs. RT+Tam**	**Tam vs. RT+Tam**
Other types of cancer	0.0012	0.1388
Competing event	0.0009	0.7824

Table 4.3: Results of Gray's test (p-values).

There are no statistically significant differences between Tamoxifen and the combined treatment concerning the competing risk, but there are differences at the 15 % level of significance in other cancer recurrences.

The number of patients with the other types of cancer recurrence as a first event is 9 in the radiotherapy group, 31 in the combined group and 17 in the Tamoxifen group. The subdistribution PH model estimation results follows. The model selection, interaction and the assumption test can be found in Appendix B.

Tamoxifen vs. combined treatment

The multivariate model including all the explanatory variables indicates that age is not significant ($p = 0.88$). In the model including treatment, positive lymph nodes and tumor size, the treatment variable is not significant ($p = 0.38$). Therefore, the difference between Tamoxifen and the combined treatment group in terms of other cancer recur-

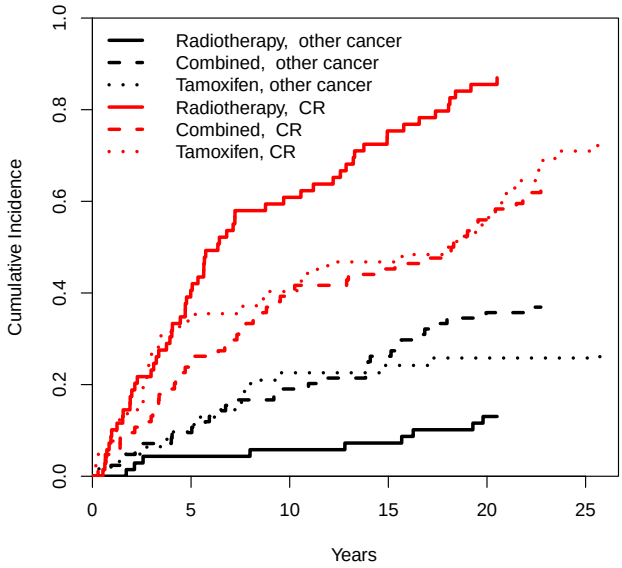

Figure 4.3: Other cancer recurrence vs. competing risk (CR), i.e. distant recurrence or death.

rence with competing risk of distant recurrence or death is not statistically significant and the analysis is discontinued.

Radiotherapy vs. combined treatment

In the model selection, tumor size ($p = 0.62$) is excluded first. In the subsequent step, age is dropped ($p = 0.58$). In the model containing only treatment and number of positive lymph nodes, the p-value of the positive lymph nodes variable is equal to 0.067. Since it is not far from 5 %, both this model and a model with only treatment are compared to one another.

The interaction term between treatment and positive lymph nodes is not statistically significant, therefore model A in Table 4.4 contains only the original variables. Model B includes only treatment and according to the Bayesian Information Criterion (BIC), this model is the optimal choice.

	MODEL A		**MODEL B**	
Variable	**Hazard ratio**	***p*-value**	**Hazard ratio**	***p*-value**
Treatment	3.291	0.001	3.27	0.002
# lymph nodes	0.883	0.067		
BIC	*396.45*		*395.41*	

Table 4.4: Subdistribution PH models for other types of cancer recurrence: radiotherapy vs. combined treatment groups.

Both the proportional assumption checks indicate model adequacy and the interpretation is as follows. The estimated hazard ratio of 3.27 reveals that the other cancer type recurrence rate increases by 227 % if the combined treatment is administered. Expressed in another way, the hazard for the combined group is 3.27 times the radiotherapy group, which means that adding Tamoxifen dramatically increases the risk of having a cancer

recurrence. This result should be taken with caution due to very wide 95 % confidence interval, from 1.56 to 6.82 (see Table 4.5).

	OTHER CANCER REC. MODEL			COMPETING RISK MODEL		
Variable	**Coeff.**	**Haz. ratio (95 % CI)**	***p*-value**	**Coeff.**	**Haz. ratio (95 % CI)**	***p*-value**
Treatment	1.18	3.27 (1.56,6.82)	0.002	-0.616	0.54 (0.37,0.78)	0.001

Table 4.5: Subdistribution PH models for other types of cancer recurrence and distant recurrence or death as the competing risk: radiotherapy vs. combined treatment groups.

As with all regressions where competing risks are involved, an excellent result in recurrences can mean a poor result in more serious types of recurrence or even death. The estimated hazard ratio of the competing risk equals 0.54 and this means the radiotherapy group has $1/0.54 = 1.85$ times the combined group's hazard to suffer from a distant recurrence or die.

4.2.3 Distant Cancer

In the analysis of distant cancer the competing event is death.

The plot in Figure 4.4 indicates that the combined treatment differs from the other treatments in the cumulative incidence of distant cancer recurrences, as well as in its competing risk.

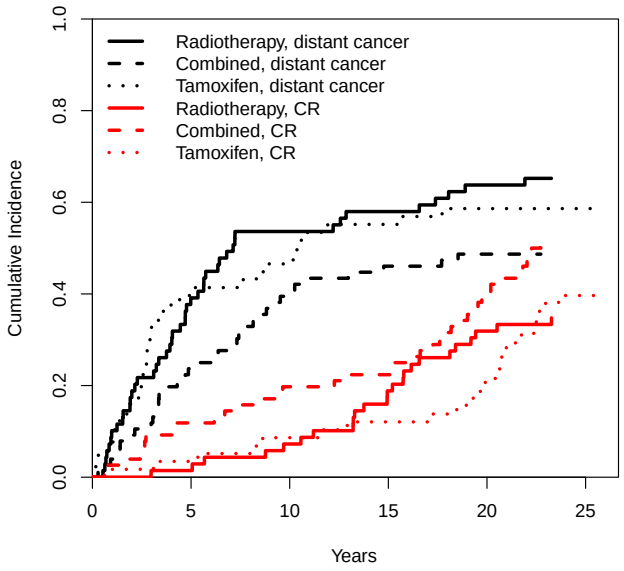

Figure 4.4: Distant cancer vs. competing risk (CR), i.e. death.

According to the plot, the combined group has the lowest cumulative incidence for distant cancer, but the highest cumulative incidence when the competing risk is death.

The difference between the cumulative incidence of distant recurrence for the combined treatment and radiotherapy is significant (see Table 4.6), while no significant difference is found between Tamoxifen and the combined treatment.

Fail cause	RT vs. RT+Tam	Tam vs. RT+Tam
Distant cancer	0.0117	0.3116
Competing event	0.2735	0.1472

Table 4.6: Results of Gray's test (p-values).

The number of patients with distant recurrence as a first event is 45 in the radiotherapy group, 37 in the combined group and 34 in the Tamoxifen group. Tamoxifen vs. the combined treatment and radiotherapy vs. combined treatment model selection, interactions check and evaluation of the PH assumption can be found in Appendix C.

Tamoxifen vs. combined treatment

In the model with all the variables, age is least significant ($p = 0.91$). In the model without age, the treatment variable is the next candidate for exclusion ($p = 0.59$). Hence, no statistical evidence of differences between Tamoxifen and the combined therapy in the appearance of distant recurrences with the competing risk of death can be detected.

Radiotherapy vs. combined treatment

All the explanatory variables are significant in the multivariate model. The interactions check determines a significant interaction between treatment and age. Schoenfeld residual plots for each of the variables indicate some possible problems, but various time transformations do not improve the results. Since no time dependence of the variables is found in the second check of the PH assumption, the analysis continues.

Table 4.7 includes detailed estimation results for the defined model along with the corresponding model for the competing risk, i.e death.

	DISTANT RECURRENCE MODEL			COMPETING RISK MODEL		
Variable	**Coeff.**	**Haz. ratio (95 % CI)**	***p*-value**	**Coeff.**	**Haz. ratio (95 % CI)**	***p*-value**
Treatment	-8.1457	0.000 (0.000,0.122)	0.008	8.5204	$5015.9(0.3,7\times10^{8})$	0.082
Patient age	-0.1099	0.896 (0.843,0.952)	0.000	0.2002	1.222(1.077,1.390)	0.002
# lymph nodes	0.1509	1.163 (1.110,1.214)	0.000	-0.1724	0.842(0.739,0.958)	0.009
Tumor size	0.0401	1.041 (1.010,1.070)	0.005	-0.0299	0.971(0.935,1.010)	0.120
Treatment × Age	0.1205	1.128 (1.020,1.243)	0.015	-0.1279	0.880(0.759,1.020)	0.089

Table 4.7: Subdistribution PH models for distant recurrence and death as the competing risk: radiotherapy vs. combined treatment groups.

The number of positive lymph nodes is significant in the both models. One node increase, assuming that everything else is held constant, results in (1.163 - 1) = 16.3 % increase in the rate of distant recurrence while the rate of death as a competing event decreases by (1 - 0.842) = 15.8 %. The effect of tumor size is significant in the recurrence model only and if all other variables are held unchanged, every additional millimeter increases the rate of distant recurrences by (1.041 - 1) = 4.1 %.

In the presence of an interaction term, the interacting variables should be interpreted with consideration to one another. If the treatment changes from radiotherapy to combined and all other variables are held constant and age is equal to zero, the rate of distant recurrence is decreased by (1 - 0.000) = 100 % (more precise 99.97 %). Such

an extreme result depends on the assumption that age = 0 which is not applicable on the studied data.

The hazard ratio of age is 0.896, which means if age increases by one year and all other variables are held constant and the patients received radiotherapy, the rate of recurrence decreases by (1 - 0.896) = 10.4 %. If age increases by one year, assuming all other variables are held constant and the treatment is combined, the hazard ratio is $(\exp(-0.1099 + 0.1205)) = 1.0107$ and therefore the rate of recurrence increases by (1.011 - 1) = 1.1 %.

The following example provides an illustration of the obtained results. Consider 4 patients in Table 4.8.

Patient	**Treatment**	**Age**	**# lymph nodes**	**Tumor size**
Patient 1	Radiotherapy	60 years	1 node	24 mm
Patient 2	Combined	60 years	1 node	24 mm
Patient 3	Radiotherapy	70 years	1 node	24 mm
Patient 4	Combined	70 years	1 node	24 mm

Table 4.8: Characteristics of four patients included in the illustration example.

Two patients who are both 60 years old receive different treatments: one receives radiotherapy and another receives the combined treatment. The other two patients are both 70 years old and they receive the analogous treatments as the 60 year old patients. The number of positive lymph nodes and the tumor size are assumed the same for all the patients.

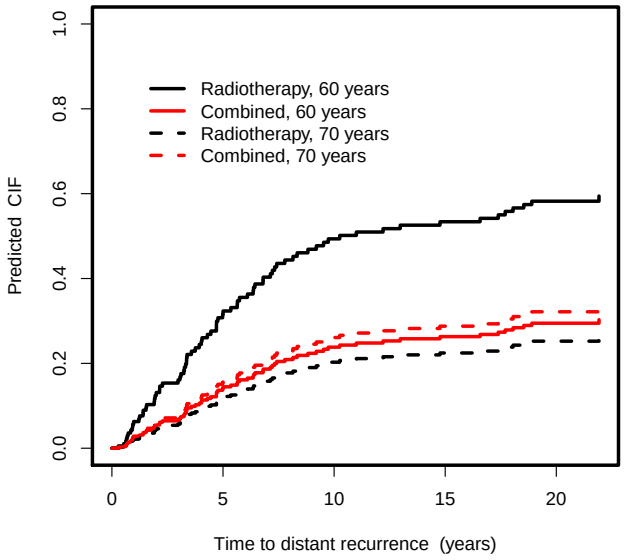

Figure 4.5: Cumulative incidence functions for 4 patients described in Table 4.8.

The predicted cumulative incidence functions (CIF) for the 4 patients in the example above are illustrated in Figure 4.5. The hazard of distant recurrence for the older woman is slightly higher than for the younger woman in the combined group, but the difference is very small. The hazard of recurrence for the 60 year old woman is much higher than for

70 year old woman in the radiotherapy group. The hazard for the 60 year old woman in the radiotherapy group is higher compared to all other women. The 70 year old woman receiving radiotherapy has the lowest risk of experiencing a distant recurrence. Therefore the patient age reflects which treatment is best.

In the model for the competing risk, the interaction term and the treatment variable are not significant at the 5 % level, but both of them are significant at the 10 % level (see Table 4.7). If all the variables are held constant and the treatment is radiotherapy, then for every additional year of age, the rate of death increases by (1.222 - 1) = 22.2 %. If the treatment is combined, a one year increase in age, holding the rest of the variables constant, leads to $(\exp(0.2002 - 0.1279) - 1) = 7.5$ % increase in the death rate.

As for distant recurrence, the interpretation of the treatment variable itself is not useful because age is assumed to be zero.

4.3 Progesterone

In the last part of the analysis, only the Tamoxifen group is examined. Progesterone status is known for 64 patients, 37 of which are progesterone positive (coded as 1 in Cox PH regression) while 27 are progesterone negative (coded as 0).

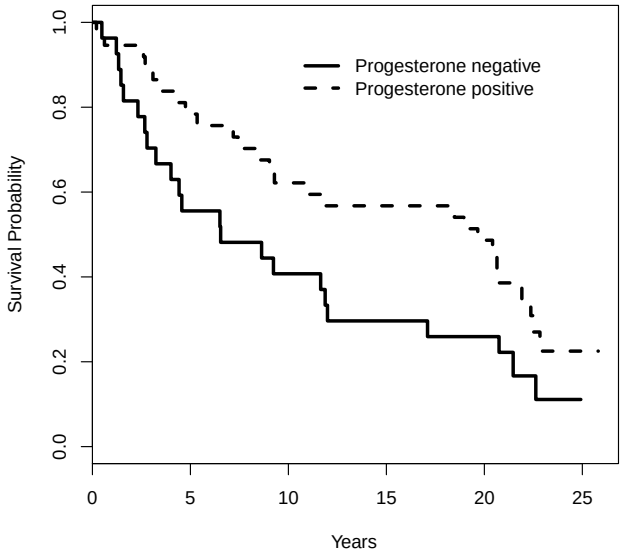

Figure 4.6: Kaplan-Meier survival plot for the Tamoxifen treatment group stratified on progesterone status.

The Kaplan-Meier plot indicates that progesterone positive patients have a higher survival probability than patients whose progesterone status is negative (see Figure 4.6). The probability of survival is around 50 % at year 7 in the progesterone negative group, while in the progesterone positive group the same probability level is reached at approximately 20 years. Gray's test rejects the null hypothesis of no difference with p-value = 0.0545.

The multivariate Cox PH model selection reveals that tumor size is not significant

($p = 0.179$). The age variable is excluded in the subsequent model ($p = 0.096$). The progesterone status and the number of positive lymph nodes in the multivariate model are both statistically significant at the 5 % level. The interaction term is not found to be significant. The first final model candidate (Model A) is shown in Table 4.9.

The hormone therapy (Tamoxifen) began within four weeks after the surgery and lasted one year (Killander et al., 2007). Hence, the treatment effect and progesterone's impact might not be fully visible during approximately 400 days. Also, in Figure 4.6, the curves cross one another in the beginning of the time line. Therefore, the progesterone status is divided into two periods. Progesterone under the first period is denoted as "progesterone, $t <$ d400" and equals the original progesterone status variable for the patients who are censored or failed during first 399 days since the randomization date. Otherwise it is equal to zero. The second period variable, "progesterone, $t \geq$ d400" is equal to zero from day 0 to day 399. For the patients who were censored or failed day 400 and on, "progesterone, $t \geq$ d400" is equal to the initial progesterone receptor status variable.

Excluding one insignificant variable at a time, beginning with the least significant one, the model selection reveals that $t \geq$ d400 and the number of positive lymph nodes should be included in the final model, denoted as model B in Table 4.9. Please note that model A does not include the progesterone variable divided into two period while model B does not include the original progesterone variable.

	MODEL A		**MODEL B**	
Variable	**Hazard ratio (95 % CI)**	***p*-value**	**Hazard ratio (95 % CI)**	***p*-value**
Progesterone	0.534 (0.299,0.955)	0.035		
# lymph nodes	1.129 (1.034,1.234)	0.007	1.131 (1.035,1.235)	0.007
Progest. > 400 days			0.503 (0.276,0.916)	0.025
BIC	*337.413*		*336.846*	

Table 4.9: Cox PH model for overall mortality: progesterone negative vs. progesterone positive group.

According to the Bayesian Information Criterion (BIC), model B is the best choice.

The graphical proportional hazards assumption check, i.e. the log-log plot, of the progesterone status variable indicates no violations after approximately 400 days, where the curves cross. The null hypothesis of no time-dependence for all the variables in model B is not rejected and consequently, model B meets the PH assumption.

The coefficient of the number of positive lymph nodes variable indicates that if the other variable is held constant, then for each additional positive lymph node, the rate of death increases by (1.131 - 1) = 13.1 %. If the number of positive lymph nodes is held constant and the progesterone status is changed from negative to positive, the rate of death after 400 days decreases by (1 - 0.503) = 49.7 %. In other words, the hazard ratio for the progesterone negative group is approximately 2 times the progesterone positive group. The 95 % confidence interval is quite wide (0.276,0.916). Hence, with the probability of 95 % the positive progesterone status results in 8.4 to 72.4 % decrease in the mortality rate and the result is valid from 400 days after the randomization.

5 Summarizing discussion

From the descriptive statistics, Tamoxifen is visibly the best at twenty years for overall survival, which was also seen in the original study by Killander et al. (2007). The results of the Cox PH regression reveal that adding Tamoxifen to radiotherapy significantly improves survival from the 6th year and on. The death rate is decreased by 36.5 %. The study by Killander et al. (2007) along with the study by Overgaard et al. (1999), did not find any significant differences in these groups, but instead in the Tamoxifen vs. combined group. The first study associated higher survival rates with the Tamoxifen treatment only, while the second study found survival was best in the combined group. In the current analysis no differences could be observed between Tamoxifen and the combined group in overall mortality.

Even though the number of individuals who experienced a loco-regional recurrence from the original 258 are few, the differences were significant with Gray's test.

The lowest cumulative incidence of other types of cancer is found in the radiotherapy group while the highest cumulative incidence for distant cancer or death is also found in the same group. The combined and Tamoxifen groups cross a few times in the instance of other cancer as well as in distant cancer or death. The result of no significant differences between the Tamoxifen and the combined groups is similar to the article by Winzer et al. (2010). In the competing risks regression for other types of cancer recurrence the hazard is 3.27 times higher in the combined treatment group compared to the radiotherapy group. This result should be taken with caution because of the wide confidence interval. Also, in the regression for the competing event, the radiotherapy group has 1.85 times the hazard of the combined group to experience one of the competing risks, which are distant recurrence or death.

The difference in the cumulative incidence for distant recurrence for radiotherapy and the combined therapy is significant. Radiotherapy has the worst result for distant cancer, while the combined therapy has the best result for distant cancer and the worst for the competing event of death. The result is consistent with the article by Winzer et al. (2010), where radiotherapy alone had a higher number of distant recurrences than the group receiving the combined therapy. In the regression, the effect of the treatment variable on distant cancer recurrences interacts with the age variable, i.e. age at the beginning of the study. One extra year of age results in 10.4 % decrease of the hazard in the radiotherapy group, while the rate of the distant recurrence for every additional year of age increases by 1.1 % in the combined group.

The progesterone status for the survival rate of patients receiving Tamoxifen is significant. Progesterone status 400 days after the randomization date results in a 49.7 % lower hazard in the progesterone receptor positive group compared to the progesterone receptor negative group. Progesterone was found to significantly contribute as a prognostic factor in the studies by Liu et al. (2010) and Lamy et al. (2002), either independently or in a combination of them. These are the same findings as in the current study.

After the analysis for this study was complete they are a few issues, which can be

discussed further. The final sample size is a bit small with only 258 observations, but it was important to remove those who were estrogen receptor negative to be able to evaluate the true effect of Tamoxifen. Other variables could have been tested showing other significant covariates. The lack of thorough medical knowledge may also have caused some wrong choices in data inclusion and statistical decisions.

Bibliography

American Cancer Society. Information and Resources for Cancer: Breast, Colon, Prostate, Lung and Other Forms, 2010. URL `http://www.cancer.org/`.

H. Anderson. The Department of Oncology in Lund, 2010. e-mail correspondence.

M. J. Bradburn, T. J. Clark, S. B. Love, and D. G. Altman. Survival Analysis Part II: Multivariate data analysis - an introduction to concepts and methods. *British Journal of Cancer*, 89:431–436, 2003.

Breast Cancer Organization. Breast Cancer Treatment Information and Pictures, 2010. URL `http://www.breastcancer.org/`.

A. C. Cameron and P. K. Trivedi. *Microeconometrics: Methods and Applications*. Cambridge University Press, New York, 2005.

D. Collett. *Modelling survival data in medical research*. Chapman & Hall, London, 1994.

J. P. Fine and R. J. Gray. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*, 94(446):496–509, June 1999.

R. J. Gray. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, 16(2):1141–1154, 1988.

IARC. Globocan. International Agency for Research on Cancer, 2008. URL `http://www-dep.iarc.fr/`.

F. Killander, H. Anderson, S. Ryden, T. Möller, K. Aspegren, J. Ceberg, C. Danewid, and P. Malmström. Radiotherapy and tamoxifen after mastectomy in postmenopausal women 20 year follow-up of the South Sweden Breast Cancer group randomised trial SSBCG II:I. *European Journal of Cancer*, 43:2100–2108, 2007.

H. T. Kim. Cumulative Incidence in Competing Risks Data and Competing Risks Regression Analysis. *Clinical Cancer Research*, 13(2):559–565, January 2007.

D. G. Kleinbaum and M. Klein. *Survival Analysis, a Self-Learning Text*. Springer Science+Business Media, Inc., New York, second edition, 2005.

P.-J. Lamy, P. Pujol, S. Thezenasl, A. Kramar, P. Rouanet, F. Guilleux, and J. Grenier. Progesterone receptor quantication as a strong prognostic determinant in postmenopausal breast cancer women under tamoxifen therapy. *Breast Cancer Research and Treatment*, 76:65–71, 2002.

S. Liu, S. K. Chia, E. Mehl, S. Leung, A. Rajput, M. C. U. Cheang, and T. O. Nielsen. Progesterone receptor is a significant factor associated with clinical outcomes and effect of adjuvant tamoxifen therapy in breast cancer patients. *Breast Cancer Research and Treatment*, 119(1):53–61, 2010.

M. Lunn and D. McNeil. Applying Cox Regression to Competing Risks. *Biometrics*, 51(2):524–532, 1995.

MFMER. Mayo Foundation for Medical Education and Research: Mayo Clinic medical information and tools for healthly living, 2010. URL `http://www.mayoclinic.com/`.

National Cancer Insitute. Comprehensive Cancer Infromation, 2010. URL `http://www.cancer.gov/`.

M. Overgaard, M.-B. Jenseni, J. Overgaard, P. S. Hansen, C. Rose, M. Anderssoni, C. Kamby, M. Kjaer, C. C. Gadeberg, B. Bruun Rasmussen, M. Blichert-Toft, and H. T. Mouridsen. Postoperative radiotherapy in high-risk postmenopausal breast cancer patients given adjuvant tamoxifen: Danish Breast Cancer Cooperative Group DBCG 82c randomised trial. *The Lancet*, 353:1641–1648, May 1999.

H. Putter, M. Fiocco, and R. B. Geskus. Tutorials in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*, 26:2389–2430, 2007.

L. Scrucca, A. Santucci, and F. Aversa. Regression modelling of competing risk using R: an in depth guide for clinicians. *Bone Marrow Transplantation*, 45:1388–1395, 2010.

J. D. Singer and J. B. Willett. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press, New York, 2003.

K.-J. Winzer, W. Sauerbrei, M. Braun, L. Liersch, J. Dunst, H. Guski, and M. Schumacher. Radiation therapy and tamoxifen after breast-conserving surgery: Updated results of a 2x2 randomised clinical trial in patients with low risk of recurrence. *European Journal of Cancer*, 46(1):95–101, 2010.

A Overall mortality

Variable	***p*-value**
STEP 1	
Treatment	0.9792
Age	0.0021
# pos. lymph nodes	0.0003
Tumor size	0.3811
BIC: 1134.961	

Table A.1: Cox PH model selection: Tamoxifen vs. combined

Variable	***p*-value**
STEP 1	
Treatment	0.0509
Age	0.1913
# pos. lymph nodes	0.0003
Tumor size	0.0635
BIC: 1211.708	
STEP 2	
Treatment	0.0457
# pos. lymph nodes	0.0001
Tumor size	0.0464
BIC: 1208.549	

Table A.2: Cox PH model selection: radiotherapy vs. combined

Variable	***p*-value**
STEP 1	
Treatment	0.0679
# pos. lymph nodes	0.0682
Tumor size	0.0442
Treat × pos. nodes	0.7864
STEP 2	
Treatment	0.4263
# pos. lymph nodes	0.0001
Tumor size	0.0179
Treat × tumor size	0.1533
STEP 3	
Treatment	0.0432
# pos. lymph nodes	0.0620
Tumor size	0.0466
Pos. nodes × tumor size	0.5803

Table A.3: Cox PH model, interaction check: radiotherapy vs. combined

Variable	***p*-value**
STEP 1	
Treatment t $<$ y5	0.6659
Treatment t $\geq$ y5	0.0389
Age	0.2063
# pos. lymph nodes	0.0006
Tumor size	0.0600
BIC: 1215.978	
STEP 2	
Treatment t $\geq$ y5	0.0387
Age	0.2098
# pos. lymph nodes	0.0006
Tumor size	0.0592
BIC: 1211.265	
STEP 3	
Treatment t $\geq$ y5	0.0320
# pos. lymph nodes	0.0002
Tumor size	0.0428
BIC: 1207.967	

Table A.4: Cox PH model selection: radiotherapy vs. combined; the treatment variable divided into two periods: before the 6th year and equal or after the 6th year

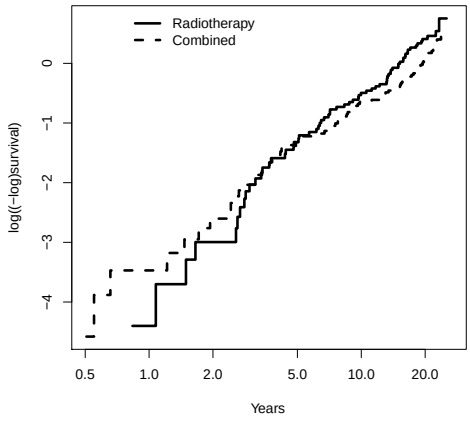

Figure A.1: The log-log plot

Variable	**Estimate**	***p*-value**
Treatment t $\geq$ y5	-1.80849	0.1675
# lymph nodes	0.10400	0.0184
Tumor size	0.01222	0.6447
Treatment t $\geq$ y5 $\times$ time	0.53905	0.2972
# lymph nodes $\times$ time	-0.00947	0.7198
Tumor size $\times$ time	0.00531	0.6740

Table A.5: Cox PH model including time-dependent variables: radiotherapy vs. combined

B Other Types of Cancer Recurrences

Variable	p**-value**
STEP 1	
Treatment	0.370
Age	0.880
# pos. lymph nodes	0.140
Tumor size	0.098
STEP 2	
Treatment	0.380
# pos. lymph nodes	0.140
Tumor size	0.093

Table B.1: Subdistribution PH model selection: Tamoxifen vs. combined

Variable	p**-value**
STEP 1	
Treatment	0.002
Age	0.480
# pos. lymph nodes	0.064
Tumor size	0.620
STEP 2	
Treatment	0.0014
Age	0.5800
# pos. lymph nodes	0.0720
STEP 3	
Treatment	0.0014
# pos. lymph nodes	0.0670
STEP 4	
Treatment	0.0016

Table B.2: Subdistribution PH model selection: radiotherapy vs. combined

Variable	p**-value**
STEP 1	
Treatment	0.016
# pos. lymph nodes	0.420
Treat $\times$ pos. nodes	0.690

Table B.3: Subdistribution PH model, interaction check: radiotherapy vs. combined

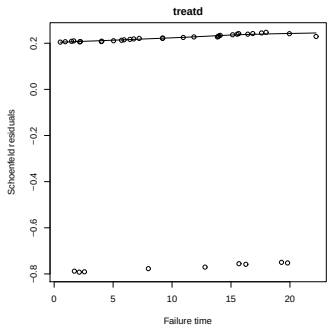

Figure B.1: Schoenfeld residuals

Variable	**Estimate**	p**-value**
Treatment	1.0654	0.11
Treatment × time	0.0002	1.00

Table B.4: Subdistribution PH model including time-dependent variables: radiotherapy vs. combined

C Distant Cancer Recurrences

Variable	***p*-value**
STEP 1	
Treatment	0.5900
Age	0.9100
# pos. lymph nodes	0.0000
Tumor size	0.0089
STEP 2	
Treatment	0.5900
# pos. lymph nodes	0.0000
Tumor size	0.0068

Table C.1: Subdistribution PH model selection: Tamoxifen vs. combined

Variable	***p*-value**
STEP 1	
Treatment	0.0048
Age	0.0160
# pos. lymph nodes	0.0000
Tumor size	0.0067

Table C.2: Subdistribution PH model selection: radiotherapy vs. combined

Model:	***p*-value**
TEST 1	
Treatment	0.008
Age	0.000
# pos. lymph nodes	0.000
Tumor size	0.005
Treat × age	0.015
TEST 2	
Treatment	0.013
Age	0.016
# pos. lymph nodes	0.005
Tumor size	0.008
Treat × pos. nodes	0.990
TEST 3	
Treatment	0.065
Age	0.015
# pos. lymph nodes	0.000
Tumor size	0.120
Treat × tumor size	0.290
TEST 4	
Treatment	0.005
Age	0.018
# pos. lymph nodes	0.920
Tumor size	0.009
Age × pos. nodes	0.840
TEST 5	
Treatment	0.005
Age	0.800
# pos. lymph nodes	0.000
Tumor size	0.550
Age × tumor size	0.690
TEST 6	
Treatment	0.005
Age	0.011
# pos. lymph nodes	0.190
Tumor size	0.051
Pos. nodes × tumor size	0.100

Table C.3: Subdistribution PH model, interaction check: radiotherapy vs. combined

Variable	**Estimate**	***p*-value**
Treatment	-7.8747	0.014
Age	-0.1473	0.065
# pos. lymph nodes	0.1040	0.810
Tumor size	0.0343	0.039
Treatment × age	0.1147	0.026
Treatment × time	0.0144	0.760
Age × time	0.0446	0.580
# pos. lymph nodes × time	0.0472	0.920
Tumor size × time	0.0478	0.520

Table C.4: Subdistribution PH model including time-dependent variables: radiotherapy vs. combined

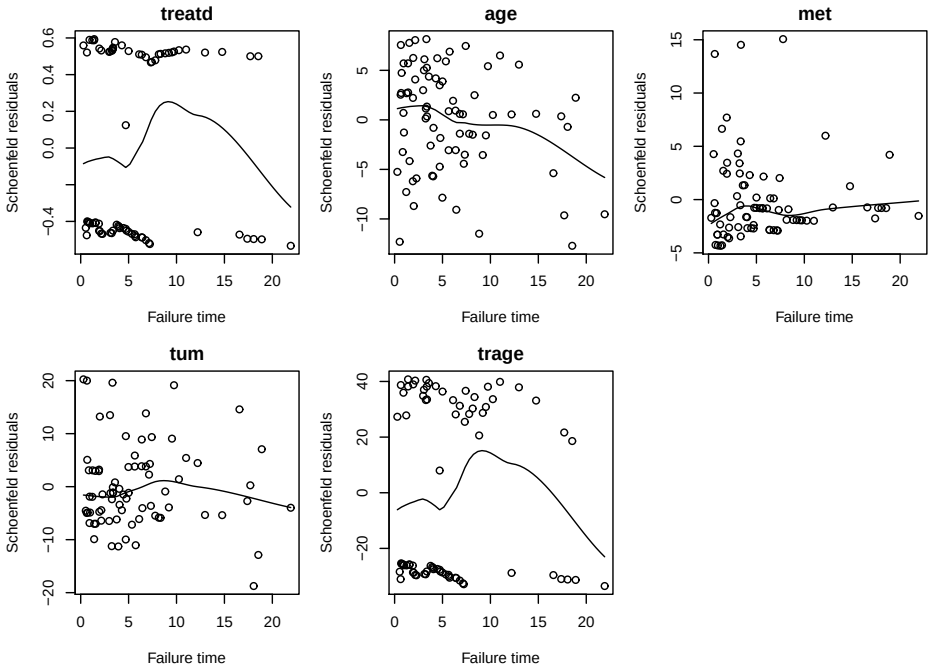

Figure C.1: Schoenfeld residuals

D Progesterone

Progesterone is shortened as PR in the tables.

Model:	p**-value**
STEP 1	
PR status	0.0611
Age	0.0704
# pos. lymph nodes	0.0110
Tumor size	0.1790
BIC: 321.892	
STEP 2	
PR status	0.0329
Age	0.0955
# pos. lymph nodes	0.0036
BIC: 338.453	
STEP 3	
PR status	0.0345
# pos. lymph nodes	0.0071
BIC: 337.413	

Table D.1: Cox PH model selection: PR+ vs. PR-

Model:	p**-value**
STEP 1	
PR status	0.1189
# pos. lymph nodes	0.0375
PR × pos. nodes	0.7625

Table D.2: Cox PH model, interaction check: PR+ vs. PR-

Model:	***p*-value**
STEP 1	
PR status t $\leq$ 400d	0.8617
PR status t $>$ 400d	0.0466
Age	0.0715
# pos. lymph nodes	0.0110
Tumor size	0.1766
BIC: 325.252	
STEP 2	
PR status t $>$ 400d	0.0464
Age	0.0706
# pos. lymph nodes	0.0107
Tumor size	0.1771
BIC: 321.454	
STEP 3	
PR status t $>$ 400d	0.0238
Age	0.0972
# pos. lymph nodes	0.0034
BIC: 337.915	
STEP 4	
PR status t $>$ 400d	0.0246
# pos. lymph nodes	0.0067
BIC: 336.846	

Table D.3: Cox PH model selection: PR+ vs. PR-, the progesterone variable divided into two periods: before day 400 and day 400 and on

Variable	**Estimate**	***p*-value**
PR status t $>$ 400d	-4.77735	0.1089
# lymph nodes	0.14137	0.6383
PR status t $>$ 400d $\times$ time	0.51841	0.1676
# lymph nodes $\times$ time	-0.00326	0.9356

Table D.4: Cox PH model including time-dependent variables: PR+ vs. PR-

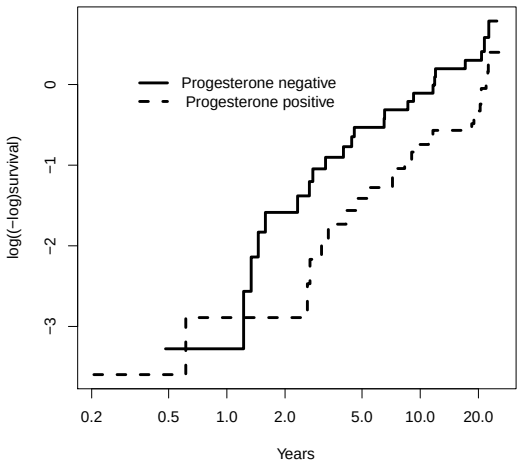

Figure D.1: The log-log plot