Calculating confidence intervals for the cumulative incidence function while accounting for competing risks: comparing the Kalbfleisch-Prentice Method and the Counting Process Method

CALCULATING CONFIDENCE INTERVALS FOR THE CUMULATIVE INCIDENCE FUNCTION WHILE ACCOUNTING FOR COMPETING RISKS: COMPARING THE KALBFLEISCH-PRENTICE METHOD AND THE COUNTING PROCESS METHOD

BY

TZVIA R. ILJON, B.Sc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF MATHEMATICS & STATISTICS

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

© Copyright by Tzvia R. Iljon, August 2013

All Rights Reserved

Master of Science (2013) McMaster University

(Mathematics & Statistics) Hamilton, Ontario, Canada

TITLE:	Calculating confidence intervals for the cumulative incidence function while accounting for competing risks: comparing the Kalbfleisch-Prentice Method and the Counting Process Method
AUTHOR:	Tzvia R. Iljon B.Sc., (Statistics) University of Ottawa, Ottawa, Canada
SUPERVISOR:	Dr. Gregory R. Pond
NUMBER OF PAGES:	xi, 64

Abstract

Subjects enrolled in a clinical trial may experience a competing risk event which alters the risk of the primary event of interest. This differs from when subject information is censored, which is non-informative. In order to calculate the cumulative incidence function (CIF) for the event of interest, competing risks and censoring must be treated appropriately; otherwise estimates will be biased. There are two commonly used methods of calculating a confidence interval (CI) for the CIF for the event of interest which account for censoring and competing risk: the Kalbfleisch-Prentice (KP) method and the Counting Process (CP) method. The goal of this paper is to understand the variances associated with the two methods to improve our understanding of the CI. This will allow for appropriate estimation of the CIF CI for a single-arm cohort study that is currently being conducted. Previous work has failed to address this question because researchers typically focus on comparing two treatment arms using statistical tests that compare cause-specific hazard functions and do not require a CI for the CIF. The two methods were compared by calculating CIs for the CIF using data from a previous related study, using bootstrapping, and a simulation study with varying event rates and competing risk rates. The KP method usually estimated a larger CIF and variance than the CP method. When event rates were low ($< 1\%$), the KP method estimated disproportionately large CIs and the CP method estimated

CIs that exceeded the interval $[0, 1]$. In cases with low event rates ($< 5\%$), bootstrapping estimation of the CIF is recommended. In cases with moderate or high event rates ($> 5\%$), the CP method is recommended as it yields more consistent results than the KP method. The CP method is recommended for the proposed study since event rates are expected to be moderate (5-10%).

Acknowledgements

I would like to thank my parents for their endless love and their support in all of my academic pursuits. I would also like to thank my partner, Nick, for the kindness and support he has shown. I would like to thank Dr. Pond for his guidance and encouragement with this paper.

Contents

List of Tables

List of Figures

Notation and abbreviations

KM	Kaplan-Meier
CIF	Cumulative incidence function
CI	Confidence interval
BCa	Bias-corrected and accelerated
KP	Kalbfleisch-Prentice
p	Probability of local recurrence
q	Probability of competing risk event
T	Random variable denoting failure time
C	Random variable denoting censoring time
K	Number of competing risks
i	Index for subject i
t_i	Observed failure time for subject i
c_i	Observed censoring time for subject i
x_i	Minimum of observed failure time and censoring time for subject i
δ_i	Indicator specifying causes of failure
t	Time

k	Cause of failure at time t
$h_k(t)$	Cause-specific hazard function at time, t, for cause of failure, k
$H_k(t)$	Cause-specific cumulative hazard function
$S(t)$	Overall survival function
$F_k(t)$	Cumulative incidence function
d_{kj}	Number of failures from cause k up to time t_j
n_j	Number of subjects at risk of failure at time t_j
$\hat{H}_k(t)$	Estimator for cause-specific cumulative hazard function
$\hat{S}_k(t)$	Estimator for overall survival function
$\hat{F}_k(t)$	Estimator for cumulative incidence function
z_α	$100(1-\alpha)$ percentile of the standard normal distribution
η	State of a counting process
$P_\eta(t)$	Probability that a process is in state η at time t
ν	Absorbing state of a counting process
$\alpha_\eta(t)$	Force of transition for state η
N	A subset of $\{1, ..., v\}$
μ	A state that is not in N
$\gamma(t)$	Total forces of transition to the set of states $\{1, ..., v\}$
$\gamma_N(t)$	Total forces of transition to the subset N
$\beta_\eta(t)$	Cumulative forces of transition to the set of states $\{1, ..., v\}$
$\beta_N(t)$	Cumulative forces of transition to the subset of states N
$\omega_N(t)$	The probability in the partial model of not leaving state 0 in the time interval $[0, t]$

$\rho_N(t)$	The probability in the partial model of leaving state 0 in the time interval $[0, t]$
$P_\eta(t, N)$	The probability of transition from state 0 to state η in the time interval $[0, t]$ in the partial chain for the subset N
$\Phi_\eta(t)$	The number of processes in state η at time t, for $\eta > 0$
$\Phi_N(t)$	The number of processes in the subset of states N at time t
$\tau(t)$	The number of processes in state 0 at time t
$\Psi(t)$	A second formulation for the number of processes in state 0 at time t
$\psi(t)$	An estimate of a single measurement of $\Psi(t)$
$\hat{\omega}_N(t)$	An estimate of $\omega_N(t)$
X_η	A Gaussian process on the time interval $[0, 1]$
$Y_{\eta,\phi}(\cdot, N)$	A Gaussian process

Chapter 1

Introduction

Survival analysis is a field of biostatistics that focuses on time-to-event data. Such data can be obtained from clinical trials, which are studies in medical research and drug development that generate safety and efficacy data for health interventions. Time-to-event data is generated by measuring the time from when a patient first enters the study until they experience an event of interest. Examples of events of interest include time from diagnosis of a fatal illness to death, or time of progression from onset of Alzheimer's to full-blown dementia. The time at which a participant experiences the event of interest is referred to as the event time. Objectives of survival analysis include estimating mean time-to-event for a population, the proportion of a population that will survive past a given time, the rate that those survivors will experience an event, or the difference between treatment options.

A fundamental aspect of time-to-event data is that not all patients generally experience the event of interest before completion of the study. The data for these patients is referred to as censored data, which is a form of missing data when the value of a measurement or observation is only partially known. There are many forms of

censoring, such as right censoring, left censoring, and interval censoring. This paper will focus on right censored data, which refers to data where event times throughout the duration of the study are missing; however, the start time is known. This includes data from patients that are lost to follow-up, patients who do not experience the event of interest during the study, and patients who experience a competing risk event that prevents the event of interest from being observed. Analysis techniques which do not account for the fact that some data are censored are not suitable for time-to-event data (Altman and Bland, 1998). Comparing the average time-to-event between two groups using a t-test or linear regression is not appropriate because it does not account for the information from censored data. This results in bias and poor estimation of expected time-to-event. Comparing the proportion of events among the two groups using the odds ratio or logistic regression ignores time. This may be a suitable method for comparing success rates of multiple treatment options, but it is inadequate for estimation of time-to-event.

A common method for analyzing time to event data is the Kaplan-Meier (KM) method. The KM method estimates the survival function for an event of interest, which describes the probability that a patient will survive beyond a given time (Kaplan and Meier, 1958). The KM method takes into account right censored data. The cumulative incidence function (CIF) is calculated as one minus the survival function and describes the cumulative probability of an event from a specific cause over time. When comparing treatment arms, the most effective treatment can be determined by comparing various treatments using the log rank test (Lin *et al.*, 1997) or Cox proportional hazards regression methods (Lau *et al.*, 2009).

When conducting a time-to-event analysis, one potential complicating factor is a

competing risk event. A competing risk is defined as any type of event that either hinders the observation of the event of interest or alters its probability of occurrence (Gooley *et al.*, 1999). For example, the event of interest could be recurrence of breast cancer in patients who have had a lumpectomy, a surgical procedure to remove the cancer. Competing risks include other cancers, whose treatment may affect recurrence of the localized breast cancer, or death from other causes.

The KM method is not appropriate in situations where there are competing risks. It assumes that censoring is independent of the event of interest and, as a consequence, ignores competing risks. This assumption is violated when a competing risk exists, and such a violation can lead to biased estimates. When an individual experiences a competing risk or censoring event, they are removed from the group that is at risk, leading to overestimation of the cumulative incidence of the event of interest (Gooley *et al.*, 1999; Satagopan *et al.*, 2004). This inflation of the cumulative incidence results in undesirable and avoidable bias (Gooley *et al.*, 1999). The assumption of independence compromises the interpretation of the CIF by supposing that the probability of experiencing an event of interest is the same in the presence and absence of competing risks (Gooley *et al.*, 1999). Often, competing risks and the event of interest are not independent. For example, in a study investigating lifetime risk of development of coronary heart disease, a competing risk is kidney failure, which is also related to high blood pressure (MacMahon *et al.*, 1997). Treatment for kidney failure may affect later development of heart disease. As such, the KM method is not well substantiated in this medical context and may not be clinically meaningful (Tai *et al.*, 2001).

Alternative methods have been proposed for calculating the CIF which account for

competing risks. Kalbfleisch and Prentice derived a two-step process for calculating cumulative incidence while accounting for the informative censoring resulting from competing risk events (Kalbfleisch and Prentice, 2002). Another estimator for the cumulative incidence can be derived using the counting process-martingale formulation (Aalen, 1978). These methods have been compared extensively, both through simulation (Gooley *et al.*, 1999) and application to real data (Gooley *et al.*, 1999; Tai *et al.*, 2001; Kim, 2007). However, in the presence of competing risks, researchers are typically interested in comparing time to event data for multiple treatment options for a particular ailment. The tests that have been derived for comparing treatment arms, such as the Gray test and the log-rank test, do not compare the CIF directly (Kim, 2007; Zhang *et al.*, 2008), so a confidence interval (CI) for the CIF is not required. As such, few studies have been conducted which investigate the confidence bands associated with the CIF estimates (Lin *et al.*, 1997; Fine and Gray, 1999). The performance of these variance estimators needs to be assessed for use in a clinical trial which is underway, for which the CIF estimate and the confidence bounds are of interest. This trial, for which the goal is to estimate the CIF and associated CI, is a single-arm cohort study and therefore no comparison of the CIF between treatment arms can be conducted. Hence, an analysis to compare different variance estimators of the CIF was conducted, leading to improved understanding of CIs and accuracy of estimation of the CIF. This study will allow improved inference of the results from the ongoing clinical trial.

Chapter 2

Setting

Breast cancer patients typically undergo a series of treatments to remove the cancer with the goal of ensuring that it does not reoccur. For early stage breast cancer patients, one option is to have a lumpectomy, which is a type of surgery, followed by radiation and chemotherapy. However, up to 30% of patients in North America who are candidates for this treatment regimen do not undergo radiation because of inconvenience and cost (Whelan *et al.*, 2010).

A study is currently being conducted to investigate the possibility of safely treating some breast cancer patients with surgery alone. It is hypothesized that some subset of breast cancer patients have such a good prognosis that the risks associated with breast radiation outweigh the potential benefit. This single-arm cohort study is collecting time-to-event data for breast cancer patients who have not undergone chemotherapy or radiation after having a lumpectomy (NCT01791829, www.clinicaltrials.gov). Although a randomized, two-arm, non-inferiority clinical trial would be preferred, the sample size for such a trial is not feasible and hence, the cohort design was adopted. The primary outcome of interest is the upper bound of the 95% CI for the CIF. If

this upper bound is below a proportion specified by the investigators, then it may be considered safe for women with a specific clinical profile to go without radiation. Hence, the primary goal of this study is estimation of the CIF, not statistical testing. However, as mentioned previously, there are two methods of calculating the CI for the CIF. Proper estimation of the CIF requires that these methods be evaluated in order to determine which method is more appropriate for use in this context. Data from a prior study comparing two radiation regimens for breast cancer treatment with similar patient characteristics will be used as a historical control to define the upper bound of the 95% CI that would be considered safe.

The historical control data comes from a study published in 2010 investigating the long-term results of hypofractionated radiation therapy for breast cancer (Whelan *et al.*, 2010). The goal of the study was to determine whether a 3-week schedule of whole-breast radiation is as effective as a 5-week schedule in reducing the risk of recurrence of breast cancer after the lumpectomy procedure. There were 1234 women enrolled in this study and they were followed for up to 14 years. Whelan *et al.* (2010) determined rates of local recurrence and overall survival using the KM method. A 97.5% CI was used to compare rates of local recurrence in the two groups, and the null hypothesis that the hypofractionated treatment would be worse than the standard treatment was rejected in favour of the non-inferiority hypothesis ($p < 0.0001$). Secondarily, a log-rank test was used to compare the probability of survival over time and it was found that survival was not statistically different between the two groups ($p = 0.79$). Cox proportional-hazards models were used to evaluate the consistency of treatment effects across various subgroups of interest. Since the study was designed when CIF methods were not well developed, the authors used

KM methods. As summarized earlier, the use of KM is not optimal for the analysis of local recurrence because the assumption that competing events are independent is violated, since some patients will die without having had a local recurrence, and informative censoring is not taken into account. This data was re-analyzed later using statistical tests which do account for competing risks. These analyses demonstrated the same result: the 3-week schedule was deemed to be equally as effective as the 5-week schedule (Parpia *et al.*, 2013).

Chapter 3

Methods

Study Design

The primary objective of this study was to evaluate methods of estimating confidence intervals for the CIF. This will allow for appropriate estimation of the CIF for the cohort study that is currently being conducted. Two confidence intervals that take into account informative censoring were investigated. These estimation methods do not assume independence of competing risks; thus, they are not susceptible to the bias associated with violating this assumption. The first was derived by Kalbfleisch and Prentice (2002) and will be referred to as the KP method. The second was derived by Aalen (1978) and will be referred to as the Counting Process (CP) method. The formulas for these CIs are given in the subsections that follow.

To investigate the two CIF methods, CIs for the CIF for the historical data were calculated using each method. The distribution of possible values for the lower and upper bounds of the 95% CI for the CIF were further investigated by performing

a bootstrap analysis using the historical data. Finally, a simulation study was performed to evaluate if the results held over a variety of censoring and event rates. The width of a CI is affected by the sample size. As such, each simulation had 600 subjects, which reflects the anticipated sample size for the cohort study.

Analysis of Historical Data

Confidence intervals for the CIF were calculated for each of the two treatment groups from the historical control. All CIF estimates were made at 10 years, following the time frame of interest of the original trial.

95% reference intervals for the estimates of the 95% CI upper and lower bounds were calculated based on the bootstrapped data. The intervals that were generated are referred to as reference intervals, rather than confidence intervals, because CI are not random variables. Hence, the range of estimates for the lower and upper bounds of the CI are fixed and cannot be described by a CI. For example, the calculation of the bootstrap reference interval for the lower bound for the KP method is as follows: 10,000 bootstrap samples were taken from the historical data set. For each sample, the lower bound of the CI for the CIF was calculated using the KP method. A 95% reference interval was then determined for the lower bounds that were calculated. A 95% BCa reference intervals was also calculated for the upper bound of the 95% CI, as calculated by each of the two methods. Additionally, a 95% bootstrap CI was calculated for the CIF, as calculated by each method.

Simulation Study

Simulations were conducted to further evaluate the differences between the CP method and the KP method. The simulation was designed to replicate a clinical trial with 600 patients who are followed for 15 years, which is based on the structure of the historical study and the cohort study that is currently being conducted. Patients are at risk of either a local recurrence, a competing risk, or neither. For each scenario, 1000 simulated clinical trials were generated, with 600 patients per sample. Each scenario had a specified probability of local recurrence, p, and probability of competing risk, q. A vector of outcomes for local recurrence was generated from the Bernoulli distribution with 600 trials and probability of success, p. A vector of event times for local recurrence events was then generated from the Exponential distribution with mean 1/3. Similarly, for competing risks, a vector of indicators was generated from the Bernoulli distribution with 600 trials and probability of success, q. A vector of event times for competing risk events was then generated from the Exponential distribution, also with mean 1/3. The times of local recurrence and competing risk events were then compared to determine the time and status of the earliest event, if any, resulting in a vector of event times and statuses. The status 0 indicated no event and the corresponding time indicates that the patient has dropped out of the experiment and is censored. No event will be referred to as a "censored patient" and the corresponding time will be referred to as the "censoring time." The status 1 indicated a local recurrence, and the status 2 indicated a competing risk event. Using the KP method, and CP method, the CIF, variance, and 95% CI were calculated for each of the 1000 samples. The mean values were then calculated. The mean difference in CIF between the two methods were also calculated. 16 scenarios were investigated, where

p and q had very high, high, moderate, and low probabilities of occurring. These probabilities were 0.90, 0.50, 0.10, and 0.01, respectively.

A second group of simulations investigated data with a lower probability of censoring towards the beginning of the experiment. Event times and statuses were generated for local recurrence and competing risk events as in the first group of simulations. However, when a 0 was generated in either status vector, the corresponding time was replaced with a time from a different distribution. The new distribution was created based on the distribution of censoring times in the historical data, ignoring treatment groups, as pictured in Figure 3.1. This distribution of event times is more similar to real life. Some patients are censored earlier in the trial for various reasons, including moving or withdrawal of consent. However, most patients remain in the study until its completion, 10-15 years after their enrollment. The probability of censoring was calculated for each year from 0 to 15 years. A piece-wise function was then created to represent the distribution based on these probabilities. Analysis was conducted as in the first group of simulations. The code used to generate results in the simulation study can be found in Appendix A.

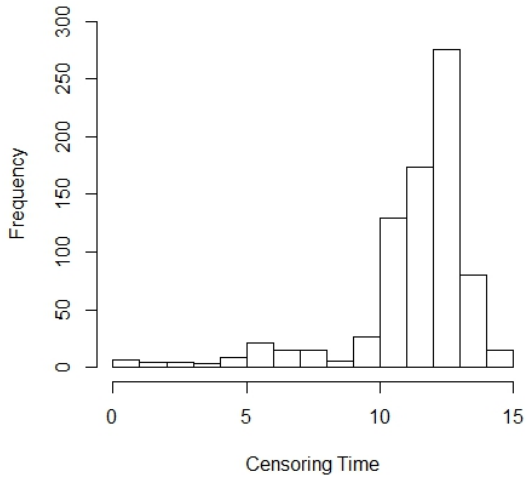
Simulations, as described above, were also conducted with 1000 subjects per sample to explore differences due to sample size.

Kalbfleisch-Prentice Method

The point estimate and variance for the CIF were derived by Kalbfleisch and Prentice in the book *The Statistical Analysis of Failure Time Data* (Kalbfleisch and Prentice, 2002). Experiencing the event of interest or a competing risk is referred to as failure.

Let T and C be two continuous random variables that denote failure and censoring

Figure 3.1: Histogram of censoring times for historical data



times, respectively. For data with K competing risks, the pairs (x_i, δ_i) are observed, where $x_i = \min(t_i, c_i)$ and $\delta_i = 0, ..., K$ is an indicator with value 0 if the individual was censored or value $1, ..., K$, specifying the causes of failure, including the event of interest and the competing risks. In this framework, any of the K events can be considered the event of interest, and all other risks will be considered competing risks. The cause-specific hazard function at time t is the instantaneous rate of failure due to cause k, conditional on survival to time t. The simplest case is when there is a single competing risk event, and therefore $K = 1$. The cause-specific hazard function for each competing risk event can be summarized as

$$h_k(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, \delta = k \mid T \geq t)}{\Delta T}, \quad k = 1, ..., K.$$

From this, the cause-specific cumulative hazard function, $H_k(t)$, and the overall survival function, $S(t)$, which is the probability of survival beyond time t, can be calculated as:

$$\begin{aligned} H_k(t) &= \int_0^t h_k(u)du \\ S(t) &= P(T > t) = \exp\left(-\sum_{k=1}^{K} H_k(t)\right). \end{aligned}$$

Since the cumulative incidence function describes the probability of failure due to cause k prior to time t, it is therefore calculated as:

$$F_k(t) = P(T \leq t, \delta = k) = \int_0^t S(u)h_k(u) = \int_0^t S(u)dH_k(u), \quad k = 1, ..., K. \tag{3.1}$$

Let $t_1 < t_2 < ... < t_j < ... < t_n$ be distinct failure times from any cause k. Let d_{kj} denote the number of failures from cause k up to time t_j, and let n_j denote the number of subjects at risk of failure at time t_j. Then the cause-specific cumulative hazard function can be estimated with the Nelson-Aalen estimator,

$$H_k(t) = \int_0^t h_k(u)du \approx \hat{H}_k(t) = \sum_{t_j \leq t} \frac{d_{kj}}{n_j}. \tag{3.2}$$

The original Kaplan-Meier estimator of failure from cause k, ignoring competing risks, is

$$\hat{S}_k(t) = \prod_{t_j \leq t} \left(1 - \frac{d_{kj}}{n_j}\right)$$

and the overall Kaplan-Meier estimator of failure from any cause is

$$\hat{S}(t) = \prod_{t_j \leq t} \left(1 - \frac{d_j}{n_j}\right). \tag{3.3}$$

Equations 3.2 and 3.3 can be plugged into Equation 3.1 to arrive at an estimate for the CIF,

$$\hat{F}_k(t) = \int_0^t S(u) dH_k(u) \approx \hat{F}_k(t) = \sum_{t_j \leq t} \hat{S}(t_{j-1}) \frac{d_{kj}}{n_j}.$$

From this, using the Delta method, the variance of $\hat{F}_k(t)$ can be derived as:

$$\begin{aligned} Var(\hat{F}_k(t)) = \sum_{t_j \leq t} \Bigg\{ & [\hat{F}_k(t) - \hat{F}_k(t_j)]^2 \frac{d_j}{n_j(n_j - d_j)} + [\hat{S}(t_{j-1})]^2 \frac{n_j - d_{kj}}{n_j^3} \\ & -2[\hat{F}_k(t) - \hat{F}_k(t_j)][\hat{S}(t_{j-1})] \frac{d_{kj}}{n_j^2} \Bigg\}. \end{aligned} \tag{3.4}$$

Further details are provided in Appendix B.

Equation 3.4 may lead to limits outside of $[0, 1]$. This can be resolved by using a $log(-log)$ transformation on the estimated CIF. The standard error of $\log(-\log(\hat{F}_k(t)))$ can be derived using the Delta method,

$$SE(\log(-\log(\hat{F}_k(t)))) = \frac{SE(\hat{F}_k(t))}{\hat{F}_k(t)[\log(\hat{F}_k(t))]}.$$

A pointwise confidence interval for $F_k(t)$, the cumulative incidence function for cause k, is (e^{-e^L}, e^{-e^U}), where

$$\begin{aligned} L &= \log(-\log(\hat{F}_k(t))) - z_{\alpha/2}(SE(\log(-\log(\hat{F}_k(t))))) \\ U &= \log(-\log(\hat{F}_k(t))) + z_{\alpha/2}(SE(\log(-\log(\hat{F}_k(t))))) \end{aligned}$$

where z_α is the $100(1-\alpha)$ percentile of the standard normal distribution.

Counting-Process Martingale Method

The point estimate and variance using the Counting-Process Martingale Method was derived by Aalen (1978) in the publication *Non-Parametric Estimation of Partial Transition Probabilities in Multiple Decrement Models*. Let $P_\eta(t)$ be the probability that a process is in state η at time t, given that it started in state 0 at time 0. We can describe the survival model with a time-continuous Markov chain with one transient state, labeled 0, and ν absorbing states. The transient state represents the state where no event has occurred and the absorbing states represent the competing risks. The infinitesimal transition probability from state 0 to state η at time t is

$$\alpha_\eta(t) = \frac{P'_\eta(t)}{P_0(t)}, \quad \eta = 1, ..., \nu.$$

$\alpha_\eta(t)$ will be referred to as forces of transition. Existence of $\alpha_\eta(t)$ depends on existence of the derivative $P'_\eta(t)$. As such, it is assumed that $\alpha_\eta(t)$ exists and is continuous everywhere for $\eta = 1, ..., \nu$. Let N be a subset of $\{1, ..., \nu\}$. A partial chain is a new model defined such that $\alpha_\mu(0) \equiv 0$ for all μ not in N. The new model allows conclusions to be drawn about a specific subset, assuming that the forces of transition of the events not included in the set N are zero. The total forces of transition to the set of states $\{1, ..., \nu\}$, and the subset N, are $\gamma(t)$ and $\gamma_N(t)$, respectively. They are

$$\gamma(t) = \sum_{\eta=1}^{\nu} \alpha_\eta(t), \qquad \gamma_N(t) = \sum_{\eta \in N} \alpha_\eta(t).$$

The cumulative forces of transition are

$$\beta_\eta(t) = \int_0^t \alpha_\eta(s)ds, \qquad \beta_N(t) = \int_0^t \gamma_N(s)ds.$$

Let $\omega_N(t)$ be the probability in the partial model of not leaving state 0 in the time interval $[0, t]$.

$$\omega_N(t) = \exp(-\beta_N(t))$$

Let $\rho_N(t)$ be the probability in the partial model of leaving state 0 in the time interval $[0, t]$.

$$\rho_N(t) = 1 - \omega_N(t)$$

The probability of transition from state 0 to state η in the time interval $[0, t]$ in the partial chain for the subset N is

$$P_\eta(t, N) = \int_0^t \alpha_\eta(s)\omega_N(t)ds = \int_0^t \omega_N(s)d\beta_\eta(s) \tag{3.5}$$

$P_\eta(t, N)$ will be estimated by substituting estimators of $\omega_N(s)$ and $\beta_\eta(s)$.
Suppose that over the time interval $[0, 1]$, ϕ independent processes with the same set of forces of transition are observed continuously. All processes start in state 0. Let $\Phi_\eta(t)$ denote the number of processes in state η at time t, a right-continuous process for $\eta > 0$. Let

$$\Phi_N(t) = \sum_{\eta \in N} \Phi_\eta(t)$$

denote the number of processes in the subset of states N at time t. Let $\tau(t) = \Phi_0(t)$, a left-continuous process. Let $\Psi(t)$ describe the number of processes in state 0 at

time t, given by

$$\Psi(t) = \begin{cases} \tau(t)^{-1} & \tau(t) > 0 \\ 0 & \tau(t) = 0 \end{cases}$$

and $\psi(t)$ be an estimate of a single measurement of $\Psi(t)$. If the subset N is thought of as a single state, the Kaplan-Meier estimator can be used for estimation of $\omega_N(t)$.

$$\hat{\omega}_N(t) = \exp\left\{\int_0^t \log(1 - \Psi(s))d\Phi_N(s)\right\} \tag{3.6}$$

The Nelson-Aalen estimator can be used for estimation of the cumulative force of transition, $\beta_\eta(t)$, given by:

$$\hat{\beta}_\eta(t) = \int_0^t \Psi(s)d\Phi_\eta(s). \tag{3.7}$$

Substituting Equations 3.6 and 3.7 into Equation 3.5 yields

$$\hat{P}_\eta(t, N) = \int_0^t \hat{\omega}_N(s-0)d\hat{\beta}_\eta(s) = \int_0^t \hat{\omega}_N(s-0)\Psi(s)d\Phi_\eta(s). \tag{3.8}$$

An estimate for the variance of $\hat{P}_\eta(t, N)$ can be derived by studying its convergence in distribution. Let $X_1, ..., X_n$ be independent Gaussian processes on the interval of time $[0, 1]$ with independent increments. Each X_η is such that $E(X_\eta) = 0$ and $Var(X_\eta) = \int_0^t \alpha_\eta(s)\psi(s)ds$. Aalen (1978) proves that the vector consisting of all processes of the form

$$Y_{\eta,\phi}(\cdot, N) = \phi^{1/2}(\hat{P}_{\eta,\phi}(\cdot, N) - P_\eta(\cdot, N)), \quad \eta \in N, N \subset \{1, ..., \nu\}$$

converges weakly to the vector consisting of the Gaussian process $Y_\eta(\cdot, N)$, defined by

$$Y_\eta(t,N) = -\int_0^t \left[\int_0^t \omega_N(u)\alpha_\eta(u)du - \omega_N(s)\right] dX_\eta(s) + \sum_{\mu\in N-\{\eta\}} \int_0^t \int_s^t \omega_N(u)\alpha_\eta(u)dudX_\mu(s).$$

This representation of the Y-processes as stochastic integrals over the X-processes makes it possible to compute moments. Since the X_η are independent processes and have independent increments, the variance is given by

$$\begin{aligned} VarY_\eta(t,N) &= \int_0^t \left[\int_s^t \omega_N(u)\alpha_\eta(u)du - \omega_N(s)\right]^2 d(VarX_\eta(s)) \\ &\qquad + \sum_{\mu\in N-\{\eta\}} \int_0^t \left[\int_s^t \omega_N(u)\alpha_\eta(u)du\right]^2 d(VarX_\mu(s)) \\ &= \int_0^t \left[\int_s^t \omega_N(u)\alpha_\eta(u)du - \omega_N(s)\right]^2 \alpha_\eta(s)\psi(s)ds \\ &\qquad + \sum_{\mu\in N-\{\eta\}} \int_0^t \left[\int_0^t \omega_N(u)\alpha_\eta(u)du\right]^2 \alpha_\mu(s)\psi(s)ds. \end{aligned}$$

which can be rewritten as

$$\begin{aligned} VarY_\eta(t,N) &= \int_0^t \left[P_\eta(t,N) - P_\eta(s,N) - \omega_N(s)\right]^2 r(s)d\beta_\eta(s) \\ &\qquad + \sum_{\mu\in N-\{\eta\}} \int_0^t \left[P_\eta(t,N) - P_\eta(s,N)\right]^2 \psi(s)d\beta_\mu(s). \end{aligned} \tag{3.9}$$

In this expression, since $\tau(t)$ has a Binomial distribution with ϕ trials and probability of success, $\psi(t)^{-1}$, $\psi(t)$ may be estimated with $\phi \cdot \Psi(t)$. Substituting this estimate and the estimates in Equations 3.6, 3.7, and 3.8, into Equation 3.9 yields an estimate

of the variance of $Y_\eta(t,N)$. This estimate can be expressed as the following:

$$\begin{aligned}\hat{Var}Y_\eta(t,N) &= \phi\int_0^t \left[\hat{P}_\eta(t,N)-\hat{P}_\eta(s,N)-\hat{\omega}_N(s)\right]^2\Psi(s)d\hat{\beta}_\eta(s)\\ &\qquad+\sum_{\mu\in N-\{i\}}\phi\int_0^t\left[\hat{P}_\eta(t,N)-\hat{P}_\eta(s,N)\right]^2\Psi(s)d\hat{\beta}_\mu(s)\\ &= \phi\int_0^t \left[\hat{P}_\eta(t,N)-\hat{P}_\eta(s,N)-\hat{\omega}_N(s)\right]^2\Psi^2(s)d\Phi_\eta(s)\\ &\qquad+\sum_{j\in A-\{i\}} n\int_0^t\left[\hat{P}_i(t,A)-\hat{P}_i(s,A)\right]^2R^2(s)dN_j(s).\end{aligned}$$

Chapter 4

Results

Analysis of Historical Data

The results of the historical data analysis are given in Table 4.1. For both treatment groups, the KP method estimated a slightly higher CIF at 10 years than the CP method. For treatment group 1, the KP and CP methods estimated 10-year cumulative incidences of 0.064, and 0.054, respectively. For treatment group 2, the KP and CP methods estimated 10-year cumulative incidences of 0.067, and 0.059, respectively. The variances for the KP method were larger than the variances for the CP method, resulting in much larger CIs, as depicted in Figures 4.1 and 4.2. The CIs for the CP method were completely contained in the CIs for the KP method. For treatment group 1, the KP method CI, (0.020, 0.145), contained the CP method CI, (0.036, 0.072). For treatment group 2, the KP method CI, (0.024, 0.143), also contained the CP method CI, (0.040, 0.078). Towards the end of the study, where the number of participants had decreased, the estimates for the bounds became very large for both methods.

Table 4.2 shows the bootstrap reference intervals for the lower and upper bounds of the 95% CI for the CIF and a 95% bootstrap CI for the CIF for each method. For this analysis, 10,000 bootstrap replications of the CIF and 95% CI for the CIF were calculated using each method. Then a 95% BCa reference interval was calculated for the upper and lower bounds of the 95% CI for the CIF. A 95% BCa CI was also determined for the CIF. For treatment 1, the KP method reference interval for the lower bound, $(0.0075, 0.0378)$, was completely below the CI for the CIF, $(0.0442, 0.0875)$, which was completely below the reference interval for the upper bound, $(0.1320, 0.1635)$. This may be due to the closeness of the CIF to 0 and the known poor performance of asymptotic estimators at extreme values. For the CP method, the reference interval for the lower bound, $(0.0219, 0.0528)$, overlapped with the CI for the CIF, $(0.0371, 0.0739)$, which also overlapped with the reference interval for the upper bound, $(0.0523, 0.0949)$. The CIs for treatment 2 behaved similarly. The bootstrap CIs for the CIF were similar in width for both methods and both treatment groups; however the KP method yielded CIs which were higher by a slight amount. The bootstrap CIs for the CIF can also be compared to the CIs calculated for each treatment group in table 4.1. The calculated KP method CIs were much wider than the bootstrap CIs for both treatment groups, while, in contrast, the calculated CP method CIs were very similar to the bootstrap CIs.

Simulation Study

Table 4.3 contains the CIFs calculated by the KP and CP methods and the differences between the estimates in each of the simulation scenarios, where times were simulated

from an exponential distribution or from a distribution which mimicked the historical data. A positive value for the mean difference indicates that the KP method calculated a higher CIF than the CP method. The KP method estimated a higher CIF in almost all cases. In scenario 1, for both censoring distributions, the KP and CP methods calculated the same value for the CIF. In the scenarios where the CP method estimated a higher CIF–scenario 4 for exponential censoring and scenarios 3 and 4 for historical censoring–the differences were small. The differences in these scenarios ranged from -0.003 to -0.034. In scenarios where the KP method yielded a higher CIF estimate, the differences ranged from 0.002 to 0.411. The differences between the estimates tended to be higher in scenarios 11, 12, 15, and 16, with high event rates and high competing risk rates, regardless of the censoring distribution. The differences between the estimates were larger in the simulations with exponential censoring than the simulations with censoring that mimics the historical data. The CP method CIF estimates appeared to be more sensitive to extreme changes in the competing risk rate than the KP method. In scenarios 9-11, where $p = 0.50$ and $q = 0.01, 0.10$ and 0.50, with historical censoring, the KP method and CP method estimated similar CIF values, ranging from 0.485 to 0.490. In scenario 12, where $q = 0.90$, the KP method estimated the CIF as 0.485, similar to the other estimates. Alternately, the CP method estimated the CIF as 0.274, which is substantially lower.

Table 4.4 contains the variances calculated by the KP and CP methods in each of the simulation scenarios for both censoring distributions. The KP method variances were larger than the CP method variances in every simulation. For both methods, the variances for the simulations with exponential censoring tended to be larger than the variances for simulations with historical censoring. For the KP method, this was true

in all scenarios, however, for the CP method, in scenarios 4, 13, 14, 15, and 16, the variance for the simulations with exponential censoring were larger than or equal to the variances for the simulation with historical censoring. In each of these scenarios, the probability of local recurrence or competing risk was high; $p = 0.90$ or $q = 0.90$. For both censoring distributions, as the competing risk rate, q, increased, the KP method variances increased. Alternatively, the CP method variances decreased as q increased, except in the four cases where the local recurrence rate was very high ($p = 0.90$). In these cases, the CP method variances increased as q increased and the CIF estimates decreased as q increased.

In the simulations with exponential censoring, the KP method variances decreased as the local recurrence rate, p, increased. No other trends related to the local recurrence rate were observed. Overall, the CP method yielded consistently small variances, ranging from 0.00001 to 0.00221, while the KP method yielded larger variances, ranging from 0.00030 to 0.04503.

Tables 4.5 and 4.6 contain 95% CIs calculated by each method for simulations with exponential censoring times and historical censoring times, respectively. In Table 4.5, the KP method always estimated a higher upper bound for the CI compared with the CP method and the CIs for the KP method were very wide. Small CIF estimates, paired with large variances, resulted in particularly large intervals. Alternately, the CP method, which calculated smaller variances, yielded much smaller CIs. For example, in scenario 3 with exponential censoring, the KP method CIF estimate was 0.022 and the 95% CI was $(0.000, 0.952)$. The CP method CIF estimate was 0.021 and the 95% CI was (0.009, 0.033). In Table 4.6, the KP method also estimated a higher upper bound for the CI than the CP method; however, the differences tended to be

slightly less extreme than in the cases with exponential censoring times. The CIs for the KP method remained very wide in comparison to the CP method. In scenario 1 with exponential censoring times, the CP method estimated a negative lower bound for the CI. The CIF estimate was 0.033 and the CI was $(-0.001, 0.067)$. The KP method CI never exceeded 0 or 1 due to the linear transformation which was applied.

The results from the simulations with 1,000 subjects can be found in Appendix C. All of trends above were observed, however, the variances calculated by both methods were slightly smaller in the simulations with the large sample size.

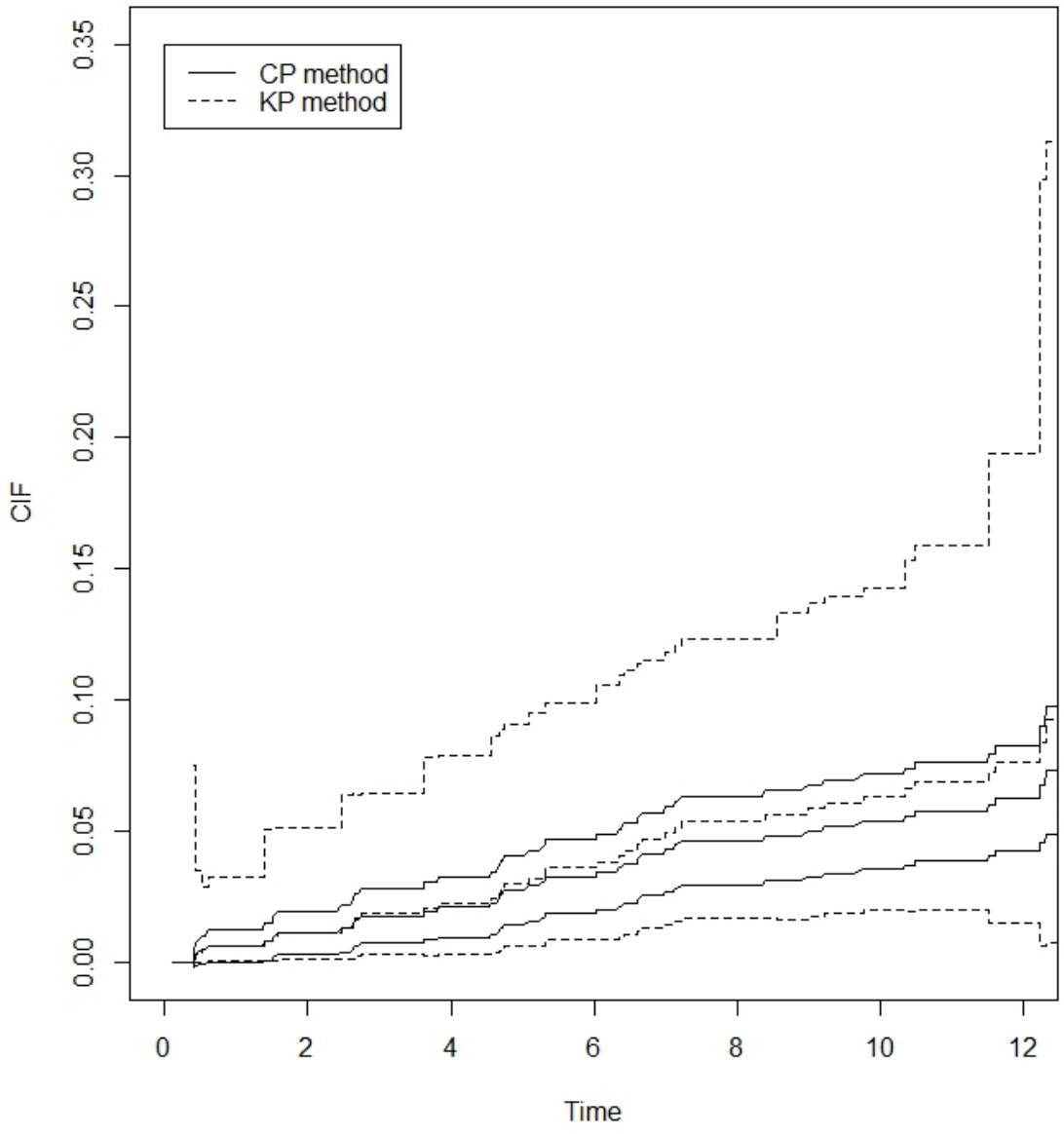


Figure 4.1: The CIF, lower, and upper bounds of the 95% CI for the CIF, calculated by the KP method and CP method for treatment group 1 of the historical data

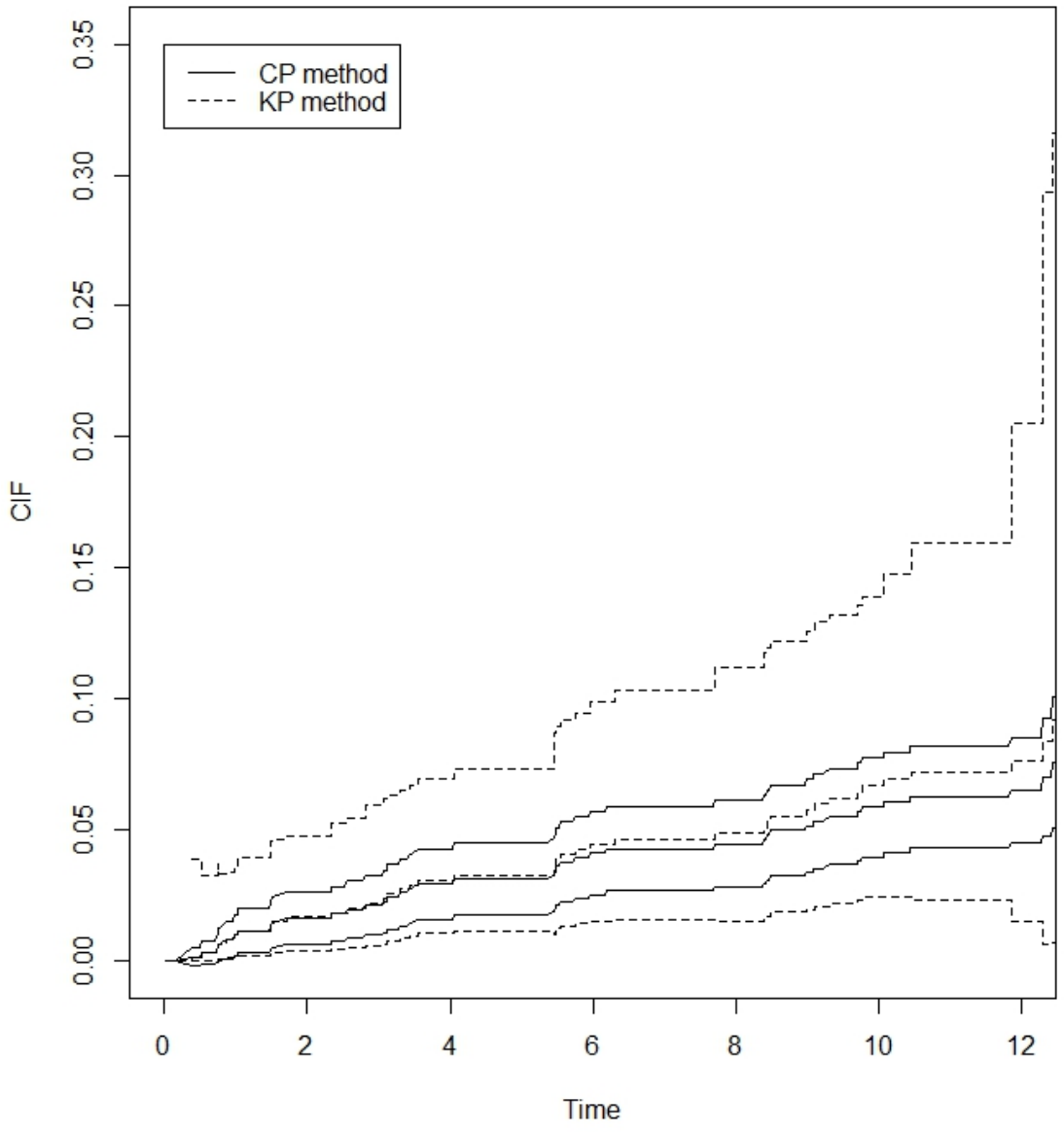


Figure 4.2: The CIF, lower, and upper bounds of the 95% CI for the CIF, calculated by the KP method and CP method for treatment group 2 of the historical data

Treatment	Method	CIF	Variance	95% CI
1	KP	0.064	0.00101	(0.020, 0.145)
1	CP	0.054	0.00009	(0.036, 0.072)
2	KP	0.067	0.00092	(0.024, 0.143)
2	CP	0.059	0.00009	(0.040, 0.078)

Table 4.1: CIF, variance, and 95% CIs calculated for each treatment group of the historical data using the KP method and CP method.

For treatment group 1, $n = 612$ and for treatment group 2, $n = 622$.

Treatment	Method	95% CI for Lower Bound of 95% CI for CIF	95% CI for Upper Bound of 95% CI for CIF	95% CI for CIF
1	KP	(0.0075, 0.0378)	(0.1320, 0.1635)	(0.0442, 0.0875)
1	CP	(0.0219, 0.0528)	(0.0523, 0.0949)	(0.0371, 0.0739)
2	KP	(0.0104, 0.0428)	(0.1283, 0.1632)	(0.0478, 0.0919)
2	CP	(0.0258, 0.0585)	(0.0580, 0.1022)	(0.0419, 0.0803)

Table 4.2: Bootstrap confidence intervals for each treatment group of the historical data.

For each CI, 10,000 bootstrap samples were taken. For each sample, the CIF and bounds of the 95% CI were calculated using the KP method and CP method. 95% CIs were then determined for each statistic.

	p	q	Exponential Censoring			Historical Censoring		
			Mean KP	Mean CP	Mean Difference	Mean KP	Mean CP	Mean Difference
1	0.01	0.01	0.033	0.033	0.000	0.010	0.010	0.000
2	0.01	0.10	0.031	0.027	0.004	0.010	0.010	0.000
3	0.01	0.50	0.022	0.021	0.002	0.010	0.013	-0.003
4	0.01	0.90	0.011	0.044	-0.033	0.010	0.044	-0.034
5	0.10	0.01	0.278	0.274	0.004	0.099	0.098	0.001
6	0.10	0.10	0.269	0.233	0.036	0.098	0.093	0.005
7	0.10	0.50	0.198	0.109	0.089	0.099	0.074	0.024
8	0.10	0.90	0.119	0.058	0.061	0.097	0.055	0.042
9	0.50	0.01	0.812	0.804	0.007	0.490	0.487	0.002
10	0.50	0.10	0.801	0.731	0.070	0.490	0.466	0.024
11	0.50	0.50	0.729	0.464	0.266	0.488	0.369	0.119
12	0.50	0.90	0.556	0.282	0.274	0.485	0.274	0.211
13	0.90	0.01	0.950	0.945	0.005	0.873	0.869	0.004
14	0.90	0.10	0.949	0.898	0.052	0.872	0.830	0.042
15	0.90	0.50	0.940	0.691	0.250	0.873	0.661	0.212
16	0.90	0.90	0.906	0.495	0.411	0.871	0.492	0.379

Table 4.3: Mean CIF estimates calculated for simulated data by the KP method and CP method.

p denotes the probability of a local recurrence. q denotes the probability of a competing risk event. Scenarios in the first column had censoring times that followed the exponential distribution with mean 1/3. Scenarios in the second column had censoring times that mimicked the censoring times for the historical data. Every combination was simulated 1000 times with $n = 600$.

	p	q	Exponential Censoring		Historical Censoring	
			KP	CP	KP	CP
1	0.01	0.01	0.04298	0.00042	0.00030	0.00002
2	0.01	0.10	0.04412	0.00026	0.00049	0.00002
3	0.01	0.50	0.04503	0.00005	0.00206	0.00002
4	0.01	0.90	0.04483	0.00001	0.01251	0.00002
5	0.10	0.01	0.02832	0.00221	0.00041	0.00015
6	0.10	0.10	0.02878	0.00151	0.00059	0.00014
7	0.10	0.50	0.03469	0.00029	0.00212	0.00012
8	0.10	0.90	0.04092	0.00010	0.01193	0.00009
9	0.50	0.01	0.00365	0.00082	0.00057	0.00043
10	0.50	0.10	0.00415	0.00080	0.00073	0.00042
11	0.50	0.50	0.00815	0.00058	0.00197	0.00039
12	0.50	0.90	0.02099	0.00035	0.00875	0.00033
13	0.90	0.01	0.00030	0.00011	0.00022	0.00019
14	0.90	0.10	0.00042	0.00018	0.00033	0.00024
15	0.90	0.50	0.00126	0.00038	0.00109	0.00038
16	0.90	0.90	0.00527	0.00042	0.00437	0.00042

Table 4.4: Mean variances calculated for simulated data by the KP method and CP method.

p denotes the probability of a local recurrence. q denotes the probability of a competing risk event. Scenarios in the first column had censoring times that followed the exponential distribution with mean 1/3. Scenarios in the second column had censoring times that mimicked the censoring times for the historical data. Every combination was simulated 1000 times with $n = 600$.

	p	q	KP Method			CP method		
			Mean CIF	Mean Variance	95% CI	Mean CIF	Mean Variance	95% CI
1	0.01	0.01	0.033	0.04298	(0.000, 0.902)	0.033	0.00042	(-0.001, 0.067)
2	0.01	0.10	0.031	0.04412	(0.000, 0.914)	0.027	0.00026	(0.000, 0.054)
3	0.01	0.50	0.022	0.04503	(0.000, 0.952)	0.021	0.00005	(0.009, 0.033)
4	0.01	0.90	0.011	0.04483	(0.000, 0.989)	0.044	0.00001	(0.037, 0.050)
5	0.10	0.01	0.278	0.02832	(0.044, 0.603)	0.274	0.00221	(0.183, 0.364)
6	0.10	0.10	0.269	0.02878	(0.038, 0.600)	0.233	0.00151	(0.158, 0.308)
7	0.10	0.50	0.198	0.03469	(0.010, 0.598)	0.109	0.00029	(0.076, 0.142)
8	0.10	0.90	0.119	0.04092	(0.000, 0.646)	0.058	0.00010	(0.038, 0.077)
9	0.50	0.01	0.812	0.00365	(0.656, 0.901)	0.804	0.00082	(0.748, 0.860)
10	0.50	0.10	0.801	0.00415	(0.635, 0.896)	0.731	0.00080	(0.675, 0.786)
11	0.50	0.50	0.729	0.00815	(0.506, 0.863)	0.464	0.00058	(0.417, 0.511)
12	0.50	0.90	0.556	0.02099	(0.249, 0.780)	0.282	0.00035	(0.245, 0.319)
13	0.90	0.01	0.950	0.00030	(0.901, 0.975)	0.945	0.00011	(0.925, 0.966)
14	0.90	0.10	0.949	0.00042	(0.888, 0.977)	0.898	0.00018	(0.872, 0.924)
15	0.90	0.50	0.940	0.00126	(0.807, 0.981)	0.691	0.00038	(0.653, 0.729)
16	0.90	0.90	0.906	0.00527	(0.577, 0.978)	0.495	0.00042	(0.455, 0.536)

Table 4.5: Mean lower and upper bounds for the 95% CI for the CIF for simulations with exponential censoring times

p denotes the probability of a local recurrence. q denotes the probability of a competing risk event. Censoring times followed the exponential distribution with mean 1/3. Each scenario was simulated 1000 times with $n = 600$.

	p	q	KP Method			CP method		
			Mean CIF	Mean Variance	95% CI	Mean CIF	Mean Variance	95% CI
1	0.01	0.01	0.010	0.00030	(0.000, 0.141)	0.010	0.00002	(0.002, 0.018)
2	0.01	0.10	0.010	0.00049	(0.000, 0.213)	0.010	0.00002	(0.001, 0.017)
3	0.01	0.50	0.010	0.00206	(0.000, 0.579)	0.013	0.00002	(0.006, 0.020)
4	0.01	0.90	0.010	0.01251	(0.000, 0.934)	0.044	0.00002	(0.037, 0.051)
5	0.10	0.01	0.099	0.00041	(0.064, 0.143)	0.098	0.00015	(0.074, 0.122)
6	0.10	0.10	0.098	0.00059	(0.057, 0.152)	0.093	0.00014	(0.070, 0.117)
7	0.10	0.50	0.099	0.00212	(0.033, 0.211)	0.075	0.00012	(0.053, 0.096)
8	0.10	0.90	0.097	0.01193	(0.004, 0.409)	0.054	0.00009	(0.037, 0.073)
9	0.50	0.01	0.490	0.00057	(0.442, 0.536)	0.487	0.00043	(0.447, 0.528)
10	0.50	0.10	0.490	0.00073	(0.436, 0.542)	0.466	0.00042	(0.425, 0.506)
11	0.50	0.50	0.488	0.00197	(0.399, 0.572)	0.369	0.00039	(0.330, 0.408)
12	0.50	0.90	0.485	0.00875	(0.296, 0.651)	0.274	0.00033	(0.238, 0.310)
13	0.90	0.01	0.873	0.00022	(0.841, 0.900)	0.869	0.00019	(0.842, 0.896)
14	0.90	0.10	0.872	0.00033	(0.832, 0.904)	0.830	0.00024	(0.800, 0.860)
15	0.90	0.50	0.873	0.00109	(0.791, 0.924)	0.661	0.00038	(0.623, 0.699)
16	0.90	0.90	0.871	0.00437	(0.660, 0.952)	0.492	0.00042	(0.452, 0.532)

Table 4.6: Mean lower and upper bounds for the 95% CI for the CIF for simulations with historical censoring times.

p denotes the probability of a local recurrence. q denotes the probability of a competing risk event. Censoring times mimicked the censoring times of the historical distribution. Each scenario was simulated 1000 times with $n = 600$.

Chapter 5

Discussion

In a clinical trial that produces time-to-event data, a statistic of interest is the CIF, which describes the proportion of individuals that have experienced the event of interest at a given time. To estimate the CIF without bias, competing risks and censoring must be taken into account. A competing risk is an event whose occurrence will alter the probability of occurrence of the event of interest. As a result, competing risk events are informative, as opposed to censoring events which are not informative beyond the time of censoring. Previous research has compared and evaluated valid methods of estimating the CIF extensively; however, little research has been conducted investigating calculation of the variance associated with these estimates using different methods. Most clinical trials are concerned with comparing the CIFs of two treatment arms using statistical tests, so research regarding the CIF has been focused in this area. There are two commonly used methods that have been proposed for calculating the CIF and the associated variance: the Kalbfleisch-Prentice method and the Counting Process method by Aalen.

The primary objective of this study was to improve our understanding of the CIF

and its variance calculated using these two methods. The results will help to guide estimation of the CIF for a single-arm cohort study that is currently underway. The cohort study is concerned with the upper bound of the 95% CI for the CIF. Here, the KP and CP methods were compared using historical data from a related study, which was used as the historical control rate in planning the cohort study. Additionally, bootstrapping was performed along with a simulation study using estimates designed to replicate plausible outcomes for the cohort study.

For each treatment group of the historical data, estimates for the CIF, variance, and 95% CI for the CIF were calculated using each method. The KP method estimated a slightly higher CIF than the CP method for both treatment groups; although, this difference may be considered negligible in certain contexts. The variances estimated by the KP method were larger than the variances estimated by the CP method and the KP method CIs completely contained the CP method CIs. The combination of higher CIF estimate and larger variance calculated by the KP method leads to higher estimates for the upper bound of the CIF and suggests that the KP method is more conservative than the CP method.

Bootstrapping was also used to investigate the historical data. For each method and each treatment group, the CIF and 95% CI for the CIF were calculated for 10,000 bootstrap replications. 95% BCa reference intervals were then determined for the lower and upper bounds of the 95% CI for the CIF. Additionally, a 95% BCa CI was calculated for the CIF estimate using each method. The bootstrap reference intervals for the lower and upper bounds of the 95% CI for the CIF did not overlap with the bootstrapped 95% CI for the CIF for the KP method; whereas, the intervals did overlap for the CP method. This demonstrates that the KP method bounds are

consistently more extreme than the CP method bounds for the 95% CI for the CIF. However, the widths of the bootstrapped 95% CIs for the CIF were similar for both methods. The width of the intervals was around 3.5% for both methods. When the bootstrap CIs for the CIF were compared to the CIs calculated by each method using CIF and variance estimates, the KP method CI was much wider. This suggests that the KP method may be over-estimating the variance. The Delta method, which is used to calculate the KP CIF and variance, includes a binomial approximation to the normal, whereas the CP method and bootstrap methods are based on exact values. This approximation may be questionable when the event rate is small, thus a potential cause for the large variances observed. Since the CP method CI was more similar to the bootstrap CI, it may provide a better approximation for the CI than the KP method in this case, despite the fact that the KP method is more conservative.

A simulation study was conducted to evaluate the performance of each method in a variety of scenarios. 32 scenarios were investigated where local recurrence rates and competing risk rates were varied. The rates investigated were 0.01, 0.10, 0.5 and 0.90. Two distributions for censoring times were also employed. Each scenario was replicated 100 times, with 600 subjects in each sample. The sample size was chosen to reflect the cohort study of interest and may be considered small in the context of some clinical trials, but it is relatively large for oncology trials. For each of the 100 samples in a scenario, the CIF, variance and 95% CI for the CIF were calculated using the KP and CP methods. For each scenario, the mean of each statistic was calculated.

The KP method estimated a higher CIF than the CP method in almost all cases. In the scenarios where the CP method estimated a lower CIF, the difference was very

small. In general, the KP method estimated the CIF more conservatively than the CP method. The differences between the two methods' CIF estimates were higher in scenarios with high local recurrence rates, high competing risk rates, and exponential censoring over scenarios with lower event rates and censoring times that mimicked the historical data. In scenarios with exponential censoring, a majority of the events and censoring occurred prior to the time point of interest, whereas, in the scenarios with historical censoring, a majority of the events and censoring occur after the time point of interest. Since the simulations with exponential censoring times had more censoring and events prior to the 10-year cut point, this allows more opportunity for the methods to diverge. The KP method CIFs appeared less sensitive to changes in competing risk rates than the CP method. The KP method estimated the same CIF for four scenarios with different competing risk rates. Alternately, the CP method was very sensitive to changes in competing risk rates. In particular, when competing risk rates were very high, the CP method CIF estimates were substantially lower than CIF estimates with lower competing risk rates.

The variances calculated by the KP method varied in magnitude more than the variances calculated by the CP method. The KP method variances were substantially larger than the CP method variances in every simulation. Both methods yielded larger variances in simulations with exponential censoring, demonstrating that an increase in censoring events leads to larger variability. As the competing risk rate increased, the CP method variances tended to decrease. With an increase in competing risk rate, the number of censoring events and the number of outcome events would decrease, regardless of the censoring distribution. As such, it is logical that the CP method variances decreased as the competing risk rate increased. Alternatively, the increase in

KP method variances suggests that this method is sensitive not only to the amount of censoring, but also to the number of competing risk events. When the local recurrence rate was very high, the CP method variances also decreased when the competing risk rate increased. In these cases, the associated CIF estimates were very close to 1. This trend may be a demonstration of how extreme CIF values may influence the variance calculation.

The 95% CIs for the CIF calculated by the KP method were much wider than those calculated by the CP method. The KP method estimated much larger variances than the CP method so this is sensible. In cases with small estimates, the CP method CIs are not very wide. As such, bootstrapping may be a more suitable method of CI estimation when event rates are low. This was further illustrated using the CIF plots for the historical data. In cases where event rates are high and censoring follows an exponential distribution, the KP method CIs are wider and higher than the CP method CIs. As such, in cases with high event rates or lots of early censoring, it is important that investigators consider the differences between the two methods prior to implementation. It may be beneficial to investigate the bias associated with each of the estimates in order to decide which method should be implemented in these cases. In order to evaluate bias, a different simulation method would need to be employed, whereby a CIF curve is generated and data is simulated from the curve. The results would only be applicable to that particular curve. Aalen (1978) proved that the CP method CIF is the minimum variance unbiased estimator for the CIF. To the best of my knowledge, bias for the CIF computed by the KP method has not been derived.

The simulations covered a variety of scenarios; however, the generalizability of these results is not clear. Although several different rates were investigated, only

two sample sizes were studied. All of the trends that were observed for simulations with 600 subjects per sample were also observed for simulations with 1000 subjects per sample. The variances for both methods were smaller when the sample size was larger, leading to narrower CIs. Similar trends are expected for trials with smaller sample sizes; however, in trials with much larger samples, the two calculation methods may yield more comparable results because of the decrease in variability. In the simulation study, only two censoring distributions were investigated. Although the exponential distribution is commonly employed for survival data, other distributions like the Weibull distribution are also suitable. It is not clear if these results would be similar with other censoring distributions; however, it is hypothesized that distributions that result in lots of early censoring will yield similar results to the exponential distribution. Likewise, distributions that result in later censoring might be expected to yield similar results to the historical distribution, which tended to result in narrower CIs than those calculated in the simulations with exponential censoring times. Finally, these results were generated using right-censored data and are not generalizable to left- or interval-censored data.

The planned single-arm cohort study is expected to have moderate event rates (5-10%) that would be comparable to the event rates in the historical data. The CP method calculated 95% CIs that were similar to the bootstrap CIs for the CIF, suggesting that the estimated variance is reasonable. The variances estimated by the CP method also fluctuated less than the KP method, particularly in cases with historical censoring times. While the KP method yields a more conservative CI for the CIF, it's estimates are much less consistent than those estimated by the CP method. Furthermore, the majority of main computing packages, including SAS, R, and SPSS,

are written to calculate the CIF using the CP method. Thus, it is recommended that the CP method be used for analysis of the cohort study.

Appendix A

R Code

Organizing data into vectors for analysis

This function will receive a matrix or data frame of times and a corresponding matrix or data frame of statuses. It returns a data frame containing times and corresponding statuses that indicate the first event to occur and the event type. 0 indicates no event and the corresponding time indicates the censoring time. 1 indicates an event of interest (for this particular analysis, it indicates a local recurrence), and 2 indicates a competing risk event.

```
crv<-function(t_cr, c_cr)
{
for (i in 1:nrow(t_cr)) # loop through each patient row
{
  for (j in 2:ncol(c_cr)) # loop through each competing risk event
  {
    if (c_cr[i,1]==1) #has there been a local recurrence (lr)?
```

```
{
  if (c_cr[i,j]==1) #has there been a competing risk event (cr)?
  {
    if (t_cr[i,1]>t_cr[i,j]) #did the cr happen before the lr?
    {
      t_cr[i,1]<-t_cr[i,j] # if so, change the time to match cr
      c_cr[i,1]<-2         # and change the status to 2
    }
  }
}
else
{
  if (c_cr[i,j]==1) #has there been a cr?
  {
    if (c_cr[i,1]==2) # has there already been a cr recorded?
    {
      if (t_cr[i,1]>t_cr[i,j]) # which event happened first?
      {
        t_cr[i,1]<-t_cr[i,j] #record time of the earlier cr event
      }
    }
    else
    {
      t_cr[i,1]<-t_cr[i,j] # change the time to match the cr
```

```
          c_cr[i,1]<-2         # change the status to 2
        }
      }
    }
  }
}
# return the vector of times and statuses
cr<-as.data.frame(cbind(t_cr[,1], c_cr[,1]))
names(cr)<-c("time", "status")
return(cr)
}
```

Calculating CI for CIF using the KP Method

This function receives a vector of times and a vector of corresponding statuses. It returns a data frame containing the CIF, variance, transformed lower bound, and transformed upper bound from the KP method.

```
cif.kp<-function(time, stat)
{
  #estimate indicates the times of interest for the CIF and variance.
  #estimate can be a vector of any length
  estimate<-10
  #sort the times and corresponding censors
  o<-order(time)
  time<-time[o]
```

```
stat<-stat[o]
all<-as.data.frame(cbind(time, stat))
names(all)<-c("time", "status")
N<-length(time)

# risk group
# data is sorted by time
# everyone in the time past has had an event or been censored
n<-rep(NA, N)
for (i in 1:N)
{
 n[i]<-N-i+1
}

#Patients who have had a local recurrence at that specific time
#This is not a cumulative count, just for that single step
#d will take the value 1 or 0
d<-rep(NA, N)
for (i in 1:N)
{
  cr<-all$status[i]
  d[i]<-as.numeric(cr==1)
}
```

```
#Survival function
s<-rep(NA, N)
s[1]<-1-d[1]/n[1]
for (i in 2:N)
{
 s[i]<-s[i-1]*(1-d[i]/n[i])
}

#CIF
F<-rep(NA, N)
F[1]<-d[1]/n[1]
for (i in 2:N)
{
  F[i]<-s[i-1]*(d[i]/n[i])+F[i-1]
}

#indices for times of interest based on the vector estimate
#This loop looks for the smallest absolute difference between
#the time of interest and times listed in the vector.
#If the index chosen results in a positive difference, it means
# that it is past the time of interest, so the previous index
# is selected
j=1
index<-rep(NA, length(estimate))
```

```
  for (i in estimate)
  {
  ind<-which.min(abs(time-i))
  if (time[ind]-i>0)
    {
    ind<-ind-1
    }
  index[j]=ind
  j=j+1
  }


 #Calculate the variance and 95\% confidence intervals for
# the times of interest
 kp.l<-kp.u<-v<-rep(NA, length(estimate))
  k=1
  for (i in index)
  {
    #Variance of CIF
    vterms<-rep(NA, i)
    #specify first term because S[0]=1 is not in the S vector
    vterms[1]<-(F[i]-F[1])^2*d[1]/(n[1]*(n[1]-d[1]))
         +(n[1]-d[1])/(n[1]^3)-2*(F[i]-F[1])*1*d[1]/n[1]^2
    if (i>1)
    {
```

```
      for (j in 2:i)
      {
        a<-(F[i]-F[j])^2*d[j]/(n[j]*(n[j]-d[j]))
        b<-(s[j-1])^2*(n[j]-d[j])/(n[j]^3)
        c<-2*(F[i]-F[j])*s[j-1]*d[j]/n[j]^2
        vterms[j]<-a+b-c
      }
    }
    v[k]<-sum(vterms)

    #lower and upper bounds of transformed confidence interval
    a<-sqrt(v[k])/(F[i]*log(F[i]))
    b<-log(-log(F[i]))-qnorm(0.025, lower.tail=F)*a/sqrt(N)
    c<-log(-log(F[i]))+qnorm(0.025, lower.tail=F)*a/sqrt(N)
    kp.l[k]<-exp(-exp(b))
    kp.u[k]<-exp(-exp(c))

    k=k+1
  }
    f<-F[index]

result<-as.data.frame(cbind(estimate, f, v, d.l, d.u, kp.l, kp.u))
names(result)<-c("time", "CIF", "var", "kp.lower", "kp.upper")
return(result)
```

```
}
```

Calculating CI for CIF using Counting Process Method

This function calculates the CI for the CIF using the Counting Process method from the R package `cmprsk`. It receives a vector of failure times and a vector of corresponding statuses. It returns a data frame containing estimates of CIF, variance, lower bound, and upper bound from the Counting Process method.

```
cif.cpm<-function(time, stat)
{
#estimate is a vector of the times of interest
estimate<-10
n<-length(time)
cpm<-cuminc(time, stat, cencode=0)
cpm.t<-timepoints(cpm, estimate)
cpm.CIF<-rep(NA, length(estimate))
cpm.var<-rep(NA, length(estimate))
cpm.l<-rep(NA, length(estimate))
cpm.u<-rep(NA, length(estimate))
j=1
#timepoints returns values for both events - 1 and 2
#we are interested in the rates for event type 1
for (i in estimate)
  {
    cpm.CIF[j]<-cpm.t$est[1,j]
```

```
    cpm.var[j]<-cpm.t$var[1,j]
#note that qnorm gives -1.96
    cpm.l[j]<-cpm.CIF[j]+qnorm(0.025)*sqrt(cpm.var[j])
    cpm.u[j]<-cpm.CIF[j]-qnorm(0.025)*sqrt(cpm.var[j])
    j=j+1
  }
as.data.frame(cbind(cpm.CIF, cpm.var, cpm.l, cpm.u))
}
```

Single Iteration of Simulation

This function runs a single iteration of a simulation. This function receives probabilities of local recurrence and competing risk, sample size, and the type of data that should be simulated (exponential or following the historical data). It returns a data frame containing the CIF, variance, lower bound, and upper bound calculated by both the KP method and CP method.

```
simulate<-function(problr, probcr, n, stype)
{
  es<-10
#generate data based on simulation parameters
  time_lr<-rexp(n, rate=1/3)
  status_lr<-rbern(n, problr)
  time_cr<-rexp(n, rate=1/3)
  status_cr<-rbern(n, probcr)
```

```
#to generate data from whelan, replace the times for the status=0
  if (stype=="whelan")
  {
  lr<-as.data.frame(cbind(time_lr, status_lr))
  cr<-as.data.frame(cbind(time_cr, status_cr))
  for (i in 1:n)
    {
    if (lr$status_lr[i]==0)
    {
      lr$time_lr[i]<-rtime(runif(1), p)
    }
    if (cr$status_cr[i]==0)
    {
      cr$time_cr[i]<-rtime(runif(1), p)
    }
    }
  time_lr<-lr$time_lr
  time_cr<-cr$time_cr
  status_lr<-lr$status_lr
  status_cr<-cr$status_cr
  }

#make the vectors of failure times and statuses
times<-as.data.frame(cbind(time_lr, time_cr))
```

```
status<-as.data.frame(cbind(status_lr, status_cr))
cr.data<-crv(times, status)

#get estimates and variance from KP method
kp.data<-cif.kp(cr.data$time, cr.data$censor)

#get estimates and variance from counting process method
cpm.data<-cif.cpm(cr.data$time, cr.data$censor)

as.data.frame(cbind(kp.data, cpm.data))
}
```

Multiple Iterations of Simulation

This function runs many iterations of a simulation scenario. It receives the probabilities of local recurrence and competing risk, the sample size, the number of simulations to run, and the type of data that should be simulated. It returns a data frame containing the probabilities of local recurrence and competing risk, the CIF, variance, lower and upper bounds calculated by both the KP method and Counting Process method. This function uses the function `simulate`, given above.

```
simulaterep<-function(problr, probcr, n, reps, stype)
{
  es<-10
 store<-matrix(rep(NA, 11*reps), nrow=reps)
 for (i in 1:reps)
```

```
{
  sim<-simulate(problr, probcr, n, stype)
  store[i,1]<-sim$time[1]
  store[i,2]<-sim$CIF[1]
  store[i,3]<-sim$var[1]
  store[i,4]<-sim$t.lower[1]
  store[i,5]<-sim$t.upper[1]
  store[i,6]<-sim$cpm.CIF[1]
  store[i,7]<-sim$cpm.var[1]
  store[i,8]<-sim$cpm.lower[1]
  store[i,9]<-sim$cpm.upper[1]
}
store<-as.data.frame(store)
names(store)<-c("time", "CIF", "var", "kp.lower", "kp.upper",
"cpm.CIF", "cpm.var", "cpm.lower", "cpm.upper")

kp.v<-mean(store$var, na.rm=T)
kp.c<-mean(store$CIF, na.rm=T)
kp.l<-mean(store$kp.lower, na.rm=T)
kp.u<-mean(store$kp.upper, na.rm=T)
cpm.v<-mean(store$cpm.var)
cpm.c<-mean(sim$cpm.CIF)
cpm.l<-mean(sim$cpm.lower)
cpm.u<-mean(sim$cpm.upper)
```

```
 sims<-as.data.frame(cbind(problr, probcr, kp.c,kp.v, kp.l,
 kp.u, cpm.c, cpm.v, cpm.l, cpm.u))
 names(sims)<-c("Problr","Probcr", "kp.c","kp.v", "kp.l",
 "kp.u", "cpm.c", "cpm.v", "cpm.l", "cpm.u")
 return(sims)
}
```

Bootstrap statistic function

This is the function used for the bootstrapping confidence intervals. It is called by the function `boot` from the R package `boot`. It receives a data frame of failure times and statuses and returns bootstrap samples of the CIF, variance, lower and upper bounds calculated by the KP method and Counting Process method.

```
boot.all<-function(b, i)
{
  b<-b[i,]
  kp<-cif.kp(b$time, b$status)
  kp.c<-kp$CIF
  kp.v<-kp$var
  kp.u<-kp$t.upper
  kp.l<-kp$t.lower
  cpm<-cif.cpm(b$time, b$status)
  cpm.c<-cpm$cpm.CIF
  cpm.v<-cpm$cpm.var
  cpm.l<-cpm$cpm.lower
```

```
  cpm.u<-cpm$cpm.upper
  cbind(kp.c, kp.v, kp.l, kp.u, cpm.c, cpm.v, cpm.l, cpm.u)
}
```

Sampling random times from the censoring distribution of the historical data

This code was used to determine the cumulative probabilities of censoring from the historical data.

```
t<-cr$time[cr$status==0]
f<-rep(NA, 15)
for (i in 1:15)
  {
  t1<-t[t<i]
  f[i]<-sum(i-1<=t1)
  }
p<-f/sum(f)
p<-cumsum(p)
```

This function samples random times from the censoring distribution based on the historical data. It receives a vector of random numbers from the Uniform distribution on the interval $[0, 1]$, and a vector containing the cumulative probabilities of censoring in each year. It returns a vector of random times from the censoring distribution based on the historical data.

```
rtime<-function(x,p)
```

```
{
  t<-rep(NA, length(x))
  for (i in 1:length(x))
  {
    if (0<= x[i] && x[i]<=p[1])
      {
      t[i]<-runif(1, min=0, max=1)
      }
    else
      {
      for (j in 2:15)
        {
        if (p[j-1]<= x[i] && x[i]<=p[j])
          {
          t[i]<-runif(1, min=j-1, max=j)
          }
        }
      }
  }
  return(t)
}
```

Appendix B

Derivation of KP Method Variance

The CIF is a function of the survival function $\hat{S}(t)$:

$$\hat{F}_k(t) = \sum_{t_j \leq t} \hat{S}(t_{j-1}) \frac{d_{kj}}{n_j}$$

To find the variance of the CIF, the variance of $\hat{S}(t)$ must be calculated first. Recall the formula for $\hat{S}(t)$:

$$\hat{S}(t) = \prod_{t_j \leq t} \left(1 - \frac{d_j}{n_j}\right)$$

The variance of a product is difficult to calculate, so using the log transformation can simplify the problem.

$$\ln(\hat{S}(t)) = \sum_{t_j \leq t} \ln\left(\frac{n_j - d_j}{n_j}\right) \tag{B.1}$$

Let $p_j = \frac{n_j - d_j}{n_j}$. Then the observations of survival among the n_j subjects that are at risk are independent Bernoulli trials with constant probability p_j. The estimator and variance of this Bernoulli experiment are thus $\hat{p}_j$ and $\frac{\hat{p}_j(1-\hat{p}_j)}{n_j}$, respectively.

The Delta Method may be used to estimate the standard error of a transformed parameter. If X is a random variable with mean μ, and the function g is differentiable, then

$$g(X) \approx g(\mu) + (X - \mu)g'(\mu)$$

An approximation of the variance is

$$Var(g(X)) \approx Var(X)[g'(\mu)]^2$$

Thus the variance of a single term from Equation B.1 can be approximated with the Delta Method with the transformation $ln(X)$ and the parameter $\hat{p_j}$:

$$Var(\ln(\hat{p}_j)) \approx \frac{\hat{p}_j(1-\hat{p}_j)}{n_j}\left[\frac{1}{\hat{p}_j}\right]^2 = \frac{(1-\hat{p}_j)}{n_j\hat{p}_j} = \frac{\frac{n_j-(n_j-d_j)}{n_j}}{n_j\frac{(n_j-d_j)}{n_j}} = \frac{d_j}{n_j(n_j-d_j)}$$

The complete variance can be written as a sum of variances since each term is an independent Bernoulli experiment:

$$Var(\ln(\hat{S}(t))) \approx Var\left(\sum_{t_j \le t}\ln(p_j)\right) = \sum_{t_j \le t} Var(\ln(p_j)) = \sum_{t_j \le t}\frac{d_j}{n_j(n_j-d_j)}$$

Using the Delta Method again, the variance of $\hat{S}(t) = \exp(\ln(\hat{S}(t)))$ is

$$Var(\hat{S}(t)) \approx \sum_{t_j \le t}\left[\frac{d_j}{n_j(n_j-d_j)}\right]\left[\exp(\ln(\hat{S}(t)))\right]^2 = \left[\hat{S}(t)\right]^2\sum_{t_j \le t}\frac{d_j}{n_j(n_j-d_j)}$$

Finally, using the Delta Method again, the variance of $\hat{F}(t)$ is

$$Var(\hat{F}_k(t)) = \sum_{t_j \leq t} \left\{ [\hat{F}_k(t) - \hat{F}_k(t_j)]^2 \frac{d_j}{n_j(n_j - d_j)} + [\hat{S}(t_{j-1})]^2 \frac{n_j - d_{kj}}{n_j^3} \right.$$
$$\left. -2[\hat{F}_k(t) - \hat{F}_k(t_j)][\hat{S}(t_{j-1})] \frac{d_k j}{n_j^2} \right\}.$$

For further details, see Hosmer *et al.* (2008).

The Delta Method is applied once again when this variance is transformed using the $\log(-\log)$ transformation to ensure that the CI for the CIF estimate is in the interval $[0, 1]$.

Appendix C

Results of simulations with 1,000 subjects per trial

	p	q	Exponential Censoring			Historical Censoring		
			Mean KP	Mean CP	Mean Difference	Mean KP	Mean CP	Mean Difference
1	0.01	0.01	0.032	0.031	0.001	0.010	0.010	0.000
2	0.01	0.10	0.031	0.027	0.004	0.010	0.009	0.000
3	0.01	0.50	0.021	0.011	0.010	0.010	0.008	0.002
4	0.01	0.90	0.012	0.007	0.005	0.010	0.008	0.002
5	0.10	0.01	0.282	0.278	0.004	0.098	0.098	0.000
6	0.10	0.10	0.267	0.231	0.036	0.099	0.094	0.005
7	0.10	0.50	0.199	0.109	0.089	0.098	0.074	0.024
8	0.10	0.90	0.119	0.057	0.061	0.098	0.055	0.043
9	0.50	0.01	0.811	0.803	0.007	0.489	0.487	0.002
10	0.50	0.10	0.798	0.729	0.070	0.490	0.466	0.024
11	0.50	0.50	0.726	0.462	0.264	0.488	0.369	0.120
12	0.50	0.90	0.559	0.283	0.275	0.484	0.273	0.211
13	0.90	0.01	0.950	0.945	0.005	0.873	0.869	0.004
14	0.90	0.10	0.949	0.897	0.052	0.873	0.831	0.042
15	0.90	0.50	0.941	0.692	0.250	0.873	0.662	0.212
16	0.90	0.90	0.908	0.496	0.412	0.871	0.491	0.379

Table C.1: Mean CIF estimates calculated for simulated data by the KP method and CP method.

p denotes the probability of a local recurrence. q denotes the probability of a competing risk event. Scenarios in the first column had censoring times that followed the exponential distribution with mean 1/3. Scenarios in the second column had censoring times that mimicked the censoring times for the historical data. Every combination was simulated 1000 times with $n = 1000$.

	p	q	Exponential Censoring		Historical Censoring	
			KP	CP	KP	CP
1	0.01	0.01	0.02615	0.00023	0.00018	0.00001
2	0.01	0.10	0.02639	0.00015	0.00029	0.00001
3	0.01	0.50	0.02652	0.00002	0.00124	0.00001
4	0.01	0.90	0.02725	0.00001	0.00745	0.00001
5	0.10	0.01	0.01657	0.00134	0.00025	0.00009
6	0.10	0.10	0.01724	0.00089	0.00036	0.00009
7	0.10	0.50	0.02041	0.00017	0.00127	0.00007
8	0.10	0.90	0.02418	0.00006	0.00713	0.00005
9	0.50	0.01	0.00219	0.00049	0.00034	0.00026
10	0.50	0.10	0.00248	0.00048	0.00044	0.00025
11	0.50	0.50	0.00489	0.00035	0.00118	0.00024
12	0.50	0.90	0.01234	0.00021	0.00522	0.00020
13	0.90	0.01	0.00018	0.00007	0.00013	0.00012
14	0.90	0.10	0.00025	0.00011	0.00020	0.00014
15	0.90	0.50	0.00075	0.00023	0.00066	0.00023
16	0.90	0.90	0.00312	0.00025	0.00262	0.00025

Table C.2: Mean variances calculated for simulated data by the KP method and CP method.

p denotes the probability of a local recurrence. q denotes the probability of a competing risk event. Scenarios in the first column had censoring times that followed the exponential distribution with mean 1/3. Scenarios in the second column had censoring times that mimicked the censoring times for the historical data. Every combination was simulated 0 times with $n = 1000$.

	p	q	KP Method			CP method		
			Mean CIF	Mean Variance	95% CI	Mean CIF	Mean Variance	95% CI
1	0.01	0.01	0.033	0.02615	(0.000, 0.835)	0.031	0.00023	(0.005, 0.058)
2	0.01	0.10	0.031	0.02639	(0.000, 0.841)	0.027	0.00015	(0.005, 0.048)
3	0.01	0.50	0.021	0.02652	(0.000, 0.915)	0.011	0.00002	(0.003, 0.020)
4	0.01	0.90	0.012	0.02725	(0.000, 0.973)	0.007	0.00001	(0.002, 0.012)
5	0.10	0.01	0.282	0.01657	(0.079, 0.536)	0.278	0.00134	(0.207, 0.349)
6	0.10	0.10	0.267	0.01724	(0.067, 0.529)	0.231	0.00089	(0.173, 0.289)
7	0.10	0.50	0.199	0.02041	(0.024, 0.510)	0.109	0.00017	(0.083, 0.135)
8	0.10	0.90	0.119	0.02418	(0.002, 0.532)	0.058	0.00006	(0.043, 0.072)
9	0.50	0.01	0.811	0.00219	(0.698, 0.885)	0.803	0.00049	(0.760, 0.847)
10	0.50	0.10	0.798	0.00248	(0.678, 0.877)	0.729	0.00048	(0.686, 0.771)
11	0.50	0.50	0.726	0.00489	(0.561, 0.837)	0.463	0.00035	(0.426, 0.498)
12	0.50	0.90	0.558	0.01234	(0.322, 0.741)	0.283	0.00021	(0.255, 0.312)
13	0.90	0.01	0.950	0.00018	(0.915, 0.971)	0.945	0.00007	(0.929, 0.961)
14	0.90	0.10	0.949	0.00025	(0.907, 0.972)	0.897	0.00011	(0.877, 0.918)
15	0.90	0.50	0.941	0.00075	(0.854, 0.976)	0.692	0.00023	(0.662, 0.721)
16	0.90	0.90	0.908	0.00312	(0.702, 0.971)	0.496	0.00025	(0.464, 0.527)

Table C.3: Mean lower and upper bounds for the 95% CI for the CIF for simulations with exponential censoring times

p denotes the probability of a local recurrence. q denotes the probability of a competing risk event. Censoring times followed the exponential distribution with mean 1/3. Each scenario was simulated 1000 times with $n = 1000$.

	p	q	KP Method			CP method		
			Mean CIF	Mean Variance	95% CI	Mean CIF	Mean Variance	95% CI
1	0.01	0.01	0.010	0.00018	(0.001, 0.088)	0.010	0.00001	(0.004, 0.016)
2	0.01	0.10	0.010	0.00029	(0.000, 0.128)	0.010	0.00001	(0.004, 0.015)
3	0.01	0.50	0.010	0.00124	(0.000, 0.422)	0.008	0.00001	(0.003, 0.013)
4	0.01	0.90	0.010	0.00745	(0.000, 0.876)	0.008	0.00001	(0.004, 0.013)
5	0.10	0.01	0.100	0.00025	(0.071, 0.132)	0.098	0.00009	(0.079, 0.117)
6	0.10	0.10	0.099	0.00036	(0.066, 0.139)	0.094	0.00009	(0.075, 0.112)
7	0.10	0.50	0.098	0.00127	(0.043, 0.181)	0.074	0.00007	(0.057, 0.090)
8	0.10	0.90	0.098	0.00713	(0.009, 0.328)	0.055	0.00005	(0.041, 0.069)
9	0.50	0.01	0.490	0.00034	(0.453, 0.525)	0.487	0.00026	(0.456, 0.518)
10	0.50	0.10	0.490	0.00044	(0.449, 0.531)	0.466	0.00025	(0.435, 0.498)
11	0.50	0.50	0.488	0.00118	(0.419, 0.554)	0.369	0.00024	(0.339, 0.399)
12	0.50	0.90	0.484	0.00522	(0.338, 0.616)	0.273	0.00020	(0.245, 0.300)
13	0.90	0.01	0.873	0.00013	(0.849, 0.894)	0.869	0.00012	(0.848, 0.890)
14	0.90	0.10	0.873	0.00020	(0.843, 0.898)	0.831	0.00014	(0.808, 0.854)
15	0.90	0.50	0.873	0.00066	(0.813, 0.915)	0.661	0.00023	(0.632, 0.691)
16	0.90	0.90	0.871	0.00262	(0.726, 0.941)	0.491	0.00025	(0.460, 0.522)

Table C.4: Mean lower and upper bounds for the 95% CI for the CIF for simulations with historical censoring times.

p denotes the probability of a local recurrence. q denotes the probability of a competing risk event. Censoring times mimicked the censoring times of the historical distribution. Each scenario was simulated 1000 times with $n = 1000$.

Bibliography

Aalen, O. (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. *The Annals of Statistics*, **6**, 534–545.

Altman, D. G. and Bland, J. (1998). Time to event (survival) data. *British Medical Journal*, **317(7156)**, 468–469.

Boos, D. D. and Nychka, D. (2012). *Rlab: Functions and Datasets Required for ST370 class*. R package version 2.15.1.

Canty, A. and Ripley, B. D. (2012). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-7.

Dahl, D. B. (2013). *xtable: Export tables to LaTeX or HTML*. R package version 1.7-1.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**(446), 496–509.

Gooley, T., Leisenring, W., Crowley, J., Storer, B., *et al.* (1999). Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine*, **18**(6), 695–706.

Gray, B. (2013). *cmprsk: Subdistribution Analysis of Competing Risks*. R package version 2.2-4.

Hosmer, D., Lemeshow, S., and May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data, Second Edition*. John Wiley & Sons, Ltd.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, Ltd., 2nd edition.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**(282), 457–481.

Kim, H. T. (2007). Cumulative incidence in competing risks data and competing risks regression analysis. *Clinical Cancer Research*, **13**(2), 559–565.

Lau, B., Cole, S. R., and Gange, S. J. (2009). Competing risk regression models for epidemiologic data. *American Journal of Epidemiology*, **170**, 244–256.

Lin, D. *et al.* (1997). Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine*, **16**(8), 901–910.

MacMahon, S., Rogers, A., Neal, B., and Chalmers, J. (1997). Blood pressure lowing for the secondary prevention of myocardial infarction and stroke. *Hypertension*, **29**, 537–538.

Parpia, S., Thabane, L., Julian, J., Whelan, T., and Levine, M. (2013). Empirical comparison of methods for analyzing multiple time-to-event outcomes in a non-inferiority trial: a breast cancer study. *BMC Medical Research Methodology*, **13**(1), 44.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Satagopan, J. M., Ben-Porat, L., Berwick, M., Robson, M., Kutler, D., and Auerbach, A. D. (2004). A note on competing risks in survival data analysis. *Br J Cancer*, **91**(7), 1229–1235.

Tai, B.-C., Machin, D., White, I., and Gebski, V. (2001). Competing risks analysis of patients with osteosarcoma: a comparison of four different approaches. *Statistics in Medicine*, **20**(5), 661–684.

Whelan, T. and Smith, S. (????). A prospective cohort study evaluating risk of local recurrence following breast conserving surgery and endocrine therapy in low risk luminal a breast cancer (lumina).

Whelan, T. J., Pignol, J.-P., Levine, M. N., Julian, J. A., MacKenzie, R., Parpia, S., Shelley, W., Grimard, L., Bowen, J., Lukka, H., Perera, F., Fyles, A., Schneider, K., Gulavita, S., and Freeman, C. (2010). Long-term results of hypofractionated radiation therapy for breast cancer. *The New England Journal of Medicine*, **362:6**, 513–520.

Zhang, M.-J., Zhang, X., and Sheike, T. H. (2008). Modeling cumulative incidence function for competing risks data. *Expert Review of Clinical Pharmacology*, **1:3**, 391–400.