April 2017

Confidence Interval Estimation of Cumulative Incidence for Clustered Competing Risks

Atul Sivaswamy
The University of Western Ontario

Supervisor
Neil Klar
The University of Western Ontario

Joint Supervisor
Merrick Zwarenstien
The University of Western Ontario

Graduate Program in Epidemiology and Biostatistics

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

Abstract

In a cluster randomized trial studying a primary outcome patients are sometimes exposed to competing events. These are risks that alter the probability of the primary outcome occurring. Traditional methods of estimating the cumulative incidence for an outcome and its associated confidence interval under competing risks do not account for the effect of clustering. This may cause incorrect estimation of confidence intervals because outcomes among patients from the same center are correlated. This thesis compared six nonparametric methods of confidence interval construction for cumulative incidence, four of which account for clustering effect, under competing risks via simulation study. Over the range of examined scenarios, if the clustering effect is mild (i.e. ICC = 0.01), estimators not accounting for clustering never have significantly worse coverage than those that do. However, in cases with a large clustering effect (i.e. ICC = 0.05), using confidence interval estimators accounting for clustering should be considered.

Keywords: Competing Risks; Clustering; Cumulative Incidence Function.

Acknowledgments

First, I thank Dr. Klar and Dr. Zwarenstein for their guidance and support. In particular, I would like to thank Dr. Klar for his careful assistance and insight during the preparation of this thesis.

I thank the Eastern Cooperative Oncology Group for providing with me their data.

Thanks to the faculty, staff, and students in the Department of Epidemiology and Biostatistics, who helped me throughout the completion of my degree. I thank my friends in the Department of Statistical and Actuarial Sciences. I thank the family of Allan Wilson for their friendship.

Finally, I thank my parents.

Contents

List of Figures

List of Tables

List of Abbreviations

	Definition	First Used
KM	Kaplan-Meier Estimator	2
CI(F)	Cumulative Incidence (Function)	2
ICC	Intracluster Correlation Coefficient	3
GVHD	Graft Versus Host Disease	4
Mult	Multinomial Based Variance Estimator	11
CP	Counting Processes Based Variance Estimator	11
Lin	Linearized Variance Estimator	14
Boot1	One-sample Bootstrap Variance Estimator	16
Boot2	Two-sample Bootstrap Variance Estimator	16
Jack	Jackknife Variance Estimator	17
EBMT	European Bone Marrow Transplant Group	55

Chapter 1

Introduction

1.1 The Competing Risks Setting

Competing risks data occur whenever study subjects are at risk of experiencing one of several mutually exclusive events (Kalbfleisch and Prentice, 2002). This situation contrasts with the classical setting of survival data where subjects are assumed to experience only one type of event or where an event is composite (e.g. the event is death due to any cause).

Methodology to analyze competing risks data goes back at least to 1776 when Daniel Bernoulli estimated the increase in life expectancy due to the elimination of smallpox mortality (Dietz and Heesterbeek, 2002). A brief summary of clinical and methodological questions that motivated the development of competing risks theory is provided by Birnbaum (1979). Prior to Kalbfleisch and Prentice (1980), most competing risks methodology used latent failure models that required specifying the degree of dependence between the hypothetical failure times of each competing risk. Such assumptions cannot be checked from the data, which only contains the failure time and the type of failure. Kalbfleisch and Prentice (1980) argued in favour of modeling based on directly estimable quantities from observed data alone. In this thesis, all modeling and estimation will follow this strategy.

In clinical studies with time-to-event data, one of the main interests is estimating the cumulative distribution function of a particular event, i.e. $P(T \leq t)$, the probability of the event

T occurring by time t (Satagopan, 2004). In the traditional survival setting, when there is only one type of event and patients are followed until they experience it or are censored, an estimator of $P(T \leq t)$ can be obtained using the Kaplan-Meier method (Kaplan and Meier, 1958). The Kaplan-Meier method can also be applied in a competing risks setting to determine the cumulative distribution by treating outcomes other than the risk of interest as censored. This is not always advisable. Indeed, several articles have been recently published cautioning against this strategy in the statistical analysis of survival data both in medical (Bakoyannis and Touloumi 2012, Koller et al. 2010, Wolkewitz 2014) and epidemiological research (Andersen, 2011). The main reason is that use of the traditional methods of survival analysis, such as the Kaplan-Meier method, for competing risks data, could lead to inconsistent estimators of the probability of experiencing the event. Such methods consistently estimate cumulative distribution under the assumption of the independence of the competing risks. This is not always an unreasonable assumption. There may sometimes be valid reasons to assume risks are independent. For example, when patients who enter a study at different times are administratively censored, we need not assume this alters their probability of experiencing an outcome of clinical interest. But there are many situations in which such an independence assumption is unreasonable. Consider for example, a trial interested in an event such as death due to a cancer occurring mostly in elderly patients. If patients enrolled in the trial die from some other chronic disease (e.g. stroke), it would be inappropriate to still use the Kaplan-Meier estimator and treat these deaths as censored, as the censoring is likely informative since both cancer and stroke incidence increase with age. To treat death from illnesses caused by aging as censoring would underestimate the probability of experiencing the event of interest (Wolkewitz, 2014). There are also examples where treating a competing risk as censoring can underestimate the probability of survival. Moeschberger (1995) points out as an example of this, the competing risk of patients who drop out of a study and move because they have been experiencing success with the therapy.

Just as one wishes to estimate the cumulative distribution function in the traditional survival

setting, in the competing risks setting one wishes to estimate the cumulative incidence function (CIF) of a particular risk. If there are k possible event types, the cumulative incidence for risk i is defined as $F_i(t) = P(T \leq t, R = i)$, where T denotes the random variable for the time to the first event and R is a random variable specifying the event type that occurred ranging from 1 to k. A nonparametric cumulative incidence estimator which accounts for competing events was proposed by Kalbfleisch and Prentice (1980).

Pocock (2007) observes that in the reporting of results from time-to-event studies, displays of statistical uncertainty in plots of Kaplan-Meir curves are often absent. This may lead to the over interpretation of treatment difference. Thus it should be encouraged when reporting the results of a study where competing risks data are analyzed to have not only an estimate of the cumulative incidence function, but also a value of the precision of the estimate. For example, in a study on contraceptive use in developing countries, the competing risks examined were three possible reasons for stopping contraceptives. It was of interest to estimate the risk of, and a confidence interval for, each of the reasons of discontinuation one year after administration of the contraceptive (Choudry, 2002). Methods for estimating the variance of an estimate of the cumulative incidence (and thus for specifying confidence intervals) have existed since at least Aalen (1978). However, only relatively recently have simulation studies been conducted comparing the performance of variance estimators for the cumulative incidence function (Choudry 2002, Braun and Yuan 2007, Iljon 2013). Iljon (2013) states that previous work has not addressed the question of one-arm estimation because researchers usually focus on comparing two treatment arms using tests that do not require estimation of cumulative incidence.

Moreover, the methods developed to estimate the variance of the cumulative incidence function assume that there is no correlation in the event times of observations, which may not always be the case. Consider the following scenarios:

1. **Antibiotic Treatment for Pneumonia in Stroke patients**: Patients from different centres are cluster randomized to be given antibiotics and standard treatment or standard

treatment alone. Patients were followed until they are discharged from hospital or experience the competing risk death (Kalra et al., 2015).

2. **Bone Marrow Transplantation for Leukemia**: Patients with Leukemia from different transplantation centres received a bone marrow transplant. Patients were followed until the development of graft versus host disease (GVHD), the primary outcome, or death from another cause, the competing event (Zhou et al., 2012).

3. **Hormonal Therapy for Prostate Cancer**: Patients from different centres are followed until death from prostate cancer, the primary outcome, or death from another cause, the competing event (Messing et al., 1999).

In cases where patients come from different centers, as in the examples above, there will likely be a correlation in response between patients from the same center. This is because individuals from the same center are more likely to be treated by the same clinician, or be from the same geographical area, and thus share environmental and lifestyle factors. For binary and quantitative outcomes, the degree of similarity among responses within a cluster, known as within-cluster correlation, can be measured by a parameter called the intracluster correlation coefficient (ICC) (Donner and Klar, 2000). For binary and quantitative outcomes, unbiased ICC estimators are easily derived using analysis of variance (ANOVA). This may not be true for time-to-event data however. Kalia (2015) showed ANOVA ICC estimators for time-to-event data are negatively biased for clustered outcomes.

Clustering effects are also of significant concern in cluster randomized trials when intact social units (e.g. hospitals, communities), called clusters, are randomized into treatment arms. The results of this thesis are particularly important for cluster randomized trials with competing risks endpoints, such as the antibiotic treatment for pneumonia in stroke patients trial described in Kalra et al. (2015).

Adjustments to competing risks methods have been made to account for clustering in the case of regression (Zhou et al., 2012), and at least one variance estimator for the cumulative

incidence function in a clustered setting has been developed (Chen et al., 2008). But very little has been written on how to estimate the variance of the cumulative incidence function when outcomes are clustered, and no study has been performed comparing methods of confidence interval construction in a clustered competing risks setting. This is despite the extension of the Consort statement (Campbell et al., 2012) to cluster randomized trials which stresses the importance of reporting 95% confidence intervals for summary measures of primary outcomes such as means, risks, and rates.

1.2 Focus of Thesis

We are concerned with confidence interval estimation of a cumulative incidence function constructed for one sample under clustering. While parametric methods exist to estimate the cumulative incidence function and confidence intervals, both in the presence and absence of clustering, parametric estimators require possibly incorrect assumptions about the underlying survival distribution. Thus the focus of this thesis will be limited to nonparametric methods of estimation. Confidence interval estimates for a point in time on the cumulative incidence function will be constructed from a point estimate and variance estimate for the cumulative incidence function. We focus on 95% intervals because these are frequently reported and Campbell et al. (2012) recommend their reporting.

We will estimate $F_i(t)$, the cumulative incidence function for risk i, using the nonparametric estimator described by Kalbfliesh and Prentice (1980). It has the advantage of always being less than the complement of the Kaplan-Meier estimator (1-KM) (Pintelle, 2006).

1.3 Objectives of Thesis

This thesis has two goals:

1. To compare methods of nonparametric variance estimation and methods of confidence interval construction based on robust variance estimators for the cumulative incidence function assuming clustered competing risks data. This comparison is based on a simulation study. Bias of variance estimators will also be compared. The confidence interval

estimators will be compared on coverage, percentage missed left and right of the interval, and interval length. Parameter values for the data generated for the simulation study will be informed by the example datasets: e.g. number of competing risks, the cluster size and number of clusters, and the degree of within-cluster correlation.

2. To apply methods of confidence interval construction to the analysis of the Bone Marrow Transplantation (Zhou et al., 2012) and Hormonal Therapy for Prostate Cancer (Messing et al., 1999) datasets introduced above.

1.4 Organization of Thesis

This thesis includes six chapters. This first chapter provides the rationale and thesis objectives introducing the key competing risks challenges for clustered data. Additional context and background for methods compared in the simulation study is provided in Chapter 2. Chapter 3 outlines the design of the simulation study while the results of the simulation study are described in Chapter 4. These results are illustrated in Chapter 5 using data from the trials introduced in Section 1.1. Thesis results are summarized in Chapter 6, where the limitations of this study and possible future research topics also are discussed.

Chapter 2

Literature Review

2.1 Introduction

This chapter describes in more detail the estimator for the cumulative incidence and associated variance estimators to be compared via simulation study in Chapter 4. Including this introductory section, there are six sections in this chapter. Section 2.2 describes the specific study design to which the methods will be applied, while Section 2.3 introduces the mathematical notation used in the thesis. Next, Section 2.4 presents estimators of cumulative incidence and variance estimators used to construct confidence intervals through the transformations described in Section 2.5. Finally, Section 2.6 summarizes the chapter.

2.2 Thesis Setting

The aim of this thesis is to compare methods of confidence interval construction for the one sample cumulative incidence function using competing risks data from patients treated at different centers. Only administrative censoring is assumed to occur and between center differences are assumed to vary at random (i.e. within-center times are positively correlated).

For the remainder of this thesis we will, for clarity of presentation, examine only estimating the cumulative incidence function of the first risk (i.e. study outcome of interest), and will be examining scenarios where there is a study outcome and one competing event. The discussion is generalizable to more than one competing risk and to the cumulative incidence function of

any event. Moreover, we make no assumptions about the dependence of the competing risks or degree of correlation between responses from the same cluster. That is, our focus is limited to nonparametric estimation of variance. We will exclude ANOVA estimators of variance which attempt to estimate intracluster correlation as these have been found to be biased for time-to-event data (Kalia, 2015).

2.3 Notation

Denoting the number of clusters in a study C, we let the number of subjects in cluster c be N_c at study entry, for $c \in \{1, ..., C\}$. The total sample size at study entry is denoted N and thus, $N = N_1 + ... + N_C$.

For $i \in \{1, ..., N_c\}$, the time-to-event and censoring variables for the i^{th} subject from cluster c is denoted: (t_{ci}, δ_{ci}), where t_{ci} is the time at which an event is observed or the observation is censored; $\delta_{ci} = 1$ if the study outcome occurred, $\delta_{ci} = 2$ if the competing event occurred, and $\delta_{ci} = 0$ if the event was censored.

Let M denote the number of unique event-times when subjects failed from either cause. All subjects are assumed to enter the study at the same time, prior to random assignment, denoted $t_0 = 0$. Let $t_1 < t_2 < ... < t_j < ... < t_M$ be the unique observed times at which (possibly more than one) failure occurs.

For $j \in \{1, ...M\}$, $c \in \{1, ...C\}$, and $i \in \{1, ...N_c\}$, let $d_{1ci}(t_j) = 1$ if subject i of cluster c fails at time t_j of the study outcome and $d_{1ci}(t_j) = 0$ otherwise; $d_{2ci}(t_j) = 1$ if subject i of cluster c fails at time t_j of the competing risk and $d_{2ci}(t_j) = 0$ otherwise; $n_{ci}(t_j) = 1$ if subject i of cluster c is still at risk at (or immediately before) time t_j and $n_{ci}(t_j) = 0$ otherwise.

Then at time t_j the number of failures from the study outcome is $d_{1j} = \sum_{c=1}^{C} \sum_{i=1}^{N_C} d_{1ci}(t_j)$; the number of failures from the competing risk is $d_{2j} = \sum_{c=1}^{C} \sum_{i=1}^{N_C} d_{2ci}(t_j)$; the number of failures of any type at time t_j is $d_j = d_{1j} + d_{2j}$; and the number of individuals still at risk is $n_j = \sum_{c=1}^{C} \sum_{i=1}^{N_C} n_{ci}(t_j)$.

Also note, the number of observations in the study is $N = \sum_{c=1}^{C} \sum_{i=1}^{N_C} n_{ci}(t_0)$, the total number of failures from the study outcome, $d_1 = \sum_{j=1}^{M} \sum_{c=1}^{C} \sum_{i=1}^{N_C} d_{1ci}(t_j)$, while the total number of

failures from the competing risk, $d_2 = \sum_{j=1}^{M} \sum_{c=1}^{C} \sum_{i=1}^{N_C} d_{2ci}(t_j)$. Note $M \leq d_1 + d_2 \leq N$.

2.4 Competing Risks Estimators

2.4.1 Estimating the Cumulative Incidence Function

In the absence of competing risks, the cumulative distribution function is defined as $P(T \leq t) = 1 - S(t)$, where $S(t)$ is the survival function. An estimator of the cumulative distribution function is 1-KM, where KM is the Kaplan-Meier survival estimator. We discussed in Chapter 1 that this estimator may be inappropriate in the presence of competing risks.

In a competing risks setting, the cumulative incidence function for the study outcome is defined as $F_1(t) = P(T \leq t, R = 1)$, where T is a random variable for the time to the first event and R is a discrete random variable specifying the type of event that occurred. $R = 1$ denotes the event of interest occurred, and $R = 2$ denotes the competing event occurred. Unlike a setting where there are no competing risks, the cumulative incidence function does not converge to 1 as $t \to \infty$, but instead, $F_1(\infty) = P(R = 1)$.

The cumulative incidence function may be written in terms of the subhazards of the study outcome and competing risk (denoted $\lambda_1(t)$ and $\lambda_2(t)$ respectively). For $i = 1, 2$, the subhazard for risk i, the instantaneous event rate, is defined as:

$$\lambda_i(t) = \lim_{\Delta \to 0} \frac{P[t \leq T < t + \Delta | T \geq t, R = i]}{\Delta} \tag{2.1}$$

The overall survival function, that is, the probability of being event free from any cause, can then be written as:

$$S(t) = \exp\Big[- \int_0^t \Big(\lambda_1(v) + \lambda_2(v) \Big) dv \Big] \tag{2.2}$$

And the cumulative incidence function of the event of interest is:

$$F_1(t) = \int_0^t \lambda_1(u) S(u) du \tag{2.3}$$

Klein and Andersen (2005) note that the cause specific hazard and the cumulative incidence function require no assumptions about the dependence structure of the competing risks and that both can be estimated from the observed data, as opposed to being modeled from unverifiable assumptions. Because the focus of this thesis is on nonparametric methods of estimation, we will not discuss the model-based methods of estimating competing risks. For model-based methods of estimating competing risks see, for example, Chapter 4 of Birnbaum (1979).

As shown in equation (2.3) the cumulative incidence can be written in terms of $S(t)$, the event-free probability of survival, and $\lambda_1(t)$ the subhazard of risk 1, so an estimator for the cumulative incidence is given by an estimator for each function.

The formal derivation of these estimators is based on Aalen (1978) and Aalen and Johnson (1982). However, Pintelle (2006) and Gooley (1999) both provide heuristic justifications. That is, the Kaplan-Meier estimator can be used to estimate survival, and d_{1j}/n_j can be used to estimate the hazard of the study outcome at time t_j. Substituting these estimators into equation (2.3) gives the following estimator for the cumulative incidence for the first competing risk:

$$\hat{F}_1(t_j) = \sum_{p=1}^{j} \frac{d_{1p}}{n_p} \hat{S}(t_{p-1}) \tag{2.4}$$

where

$$\hat{S}(t_j) = \prod_{p=1}^{j} \left(1 - \frac{d_{1p} + d_{2p}}{n_p}\right) \tag{2.5}$$

and we define $\hat{S}(t_0) = 1$.

This estimator is consistent when there is no clustering and it is the nonparametric maximum likelihood estimator (Pintelle 2006; Section 4.2.2). The estimator is also available in standard statistical software (the R package "cmprsk", SAS 9.4 under "PROC phreg", and STATA 14.0 under "stcompet"). Chen et al. (2008) show that $\hat{F}_1(t)$ remains consistent for clustered event times, using the results of Speikerman and Lin (1998), as the number of clusters grows large, for a fixed cluster size. Since the three examples motivating this thesis presented in the introduction all have a large number of clusters relative to cluster size, we should not have too much concern about the bias of the point estimator in calculating confidence intervals for

our purposes. However, it is well known that the Kaplan-Meier estimator is biased under censoring, and thus the estimator shown in equation (2.4) is also biased, but Chen et al. (2008) showed the bias of this estimator is under < 0.02 absolute bias, under wide ranges of clustering and for cumulative incidence values between 0.1 to 0.5.

2.4.2 Variance Estimators Under Independence

Braun and Yuan (2007) performed a simulation study comparing six variance estimators for the cumulative incidence function where times to event were independent. Estimators were classified into two groups: those derived by counting process theory developed by Aalen (1978), and those based on the multinomial distribution. Explicit formulas of two estimators of each type are provided by Pintelle (2006; Section 4.2.4).

A variance estimator of the cumulative incidence function at time t_j (i.e. $\hat{F}_1(t_j)$) derived using counting process theory is:

$$\begin{aligned}\widehat{var}_{cp}(\hat{F}_1(t_j)) = &\sum_{p=1}^{j}[\hat{F}_1(t_j) - \hat{F}_1(t_p)]^2 \frac{d_p}{(n_p - 1)(n_p - d_p)} \\ &+ \sum_{p=1}^{j} \hat{S}(t_{p-1})^2 \frac{d_{1p}(n_p - d_{1p})}{(n_p - 1)n_p^2} \\ &- 2\sum_{p=1}^{j}[\hat{F}_1(t_j) - \hat{F}_1(t_p)]\hat{S}(t_{p-1}) \frac{d_{1p}(n_p - d_{1p})}{n_p(n_p - d_p)(n_p - 1)}\end{aligned} \tag{2.6}$$

while a variance estimator derived using a multinomial approach is:

$$\begin{aligned}\widehat{var}_{mult}(\hat{F}_1(t_j)) = &\sum_{p=1}^{j}[\hat{F}_1(t_j) - \hat{F}_1(t_p)]^2 \frac{d_p}{n_p(n_p - d_p)} \\ &+ \sum_{p=1}^{j} \hat{S}(t_{p-1})^2 \frac{d_{1p}(n_p - d_{1p})}{n_p^3} \\ &- 2\sum_{p=1}^{j}[\hat{F}_1(t_j) - \hat{F}_1(t_p)]\hat{S}(t_{p-1}) \frac{d_{1p}}{n_p^2}\end{aligned} \tag{2.7}$$

.

Equation (2.7) reduces to Greenwood's formula for the variance of the survival function when there are no competing risks (Pintelle, 2006). We notice only a slight difference in the three terms on the right hand side of equations (2.6) and (2.7). The first and second terms of equation (2.6) will always be greater than the first and second term in equation (2.7). This is because $1/(n_p - 1)$ is replaced by $1/(n_p)$ whenever it is found. This is also true of the third term, but in equation (2.6) there is the added multiplicand $(n_p - d_{1p})/(n_p - d_p)$. This term is always less than or equal to 1. These opposing scalar effects mean there is no statement we can make comparing the third terms of equations (2.6) and (2.7), and thus about the equations in general.

Past research comparing variance estimators favors those based on the counting processes method for the following reasons:

1. The simulation study by Braun and Yuan (2007) found that the counting process estimator given in equation (2.6) slightly overestimated the true variance of the cumulative incidence, and the multinomial estimator slightly underestimated the true variance (both estimations were within 6% of the empirical variance over all the scenarios they considered). Braun and Yuan (2007) denote these respectively as the Gray and Gaynor estimators.

2. The computing packages SAS 9.4 and R 3.2.0 (package "cmprsk") have implementations of a counting processes estimator. Indeed, Braun and Yuan (2007) note the estimator given in equation (2.7) is the only one available for general use.

However, because neither is computationally intensive, we will include both in our simulation study.

2.4.3 Extending Williams' Linearized Variance Estimator

In this section we derive a variance estimator for the cumulative incidence function. We will use a Taylor series linearized values approach (i.e. delta method), which is meant to account for clustered data. Williams (1995) uses such an approach to derive a variance estimator

for the Kaplan-Meier estimator applied to clustered time-to-event data. We will derive an analogous variance estimator for the cumulative incidence estimator given in (2.4) under clustered competing risks.

Williams (1995) used a robust variance estimator derived in two steps:

1. Calculating Taylor series linearized values for each cluster.

2. Applying a variance estimator accounting for clustering to these calculated values.

For the first step, Williams (1995) uses a technique developed by Woodruff (1971), which approximates a complex non-linear function (such as the Kaplan-Meier estimator or cumulative incidence estimator in Section 2.3.1) with a linear function based on a first order Taylor series expansion. This linear approximation is then used to estimate the variance of the non-linear function. When data are clustered, the variance of the Kaplan-Meier estimator $\hat{S}(t_j)$ is given by first calculating a linearized value of $\hat{S}(t_j)$ for each individual observation. Then, by linearity, the linearized values are accumulated by cluster.

In the second step, a between-cluster variance estimator uses the linearized values for each cluster to calculate the variance, by a cluster level variance estimator, as shown in equation (2.8).

Williams (2000) shows that this cluster level variance estimator is unbiased for clustered data for a linear statistic regardless of the setting, under the assumption that the C clusters were selected independently from a hypothetical infinite population of clusters. Rao and Colin (1991) show the between-cluster variance estimator is consistent when the number of clusters grows to infinity, for a fixed cluster size, when used with the Taylor series linearization of a non-linear statistic.

Applying the method used by Williams on the Kaplan-Meier estimator for clustered survival data to the cumulative incidence estimator in a clustered competing risks setting results in the formula for the variance estimator given in equation (2.8). The derivation of this estimator

is found in Appendix A and its formula is:

$$\widehat{var}_{lin}(\hat{F}_1(t_j)) = \frac{C}{C-1}\sum_{c=1}^{C}(z_c[\hat{F}_1(t_j)] - \bar{z}[\hat{F}_1(t_j)])^2 \tag{2.8}$$

where

$$\bar{z}[\hat{F}_1(t_j)] = \sum_{c=1}^{C} z_c[\hat{F}_1(t_j)]/C$$

and $z_c[\hat{F}_1(t_j)]$ are the linearized values of $\hat{F}_1(t_j)$ accumulated for cluster c defined in Appendix A.

This variance estimator has the advantage that it does not require any knowledge of the within-cluster correlation structure. Chen et al. (2008) attain a robust variance for the cumulative incidence function that accounts for within-cluster correlation based on William (1995), but the details of his derivation were not provided, nor an explicit formula for the variance estimator given. It is expected, but unknown if the estimator we derive is algebraically equivalent to the one given by Chen et al. (2008).

It is worth noting, Williams (2000) remarks that the Taylor series linearization approach used with a between-cluster variance estimator is closely related to the generalized estimating equation (GEE) robust variance approach of Liang and Zeger (1986) and, in some situations, the two approaches are the same when assuming working independence. The GEE approach attempts to improve estimation by including assumptions about the within-cluster correlation structure in the estimating equations. We forgo the inclusion of a GEE approach in the present study.

2.4.4 Bootstrap Variance Estimators

The bootstrap is a general method introduced by Efron (1979) that can be used to estimate variance for any study outcome. Several different bootstrap algorithms have been developed. The bootstrap method we consider is nonparametric, so it has the advantage of making weak distributional assumptions.

Efron (1981) applied a nonparametric bootstrap to obtain a variance estimator of $KM(t)$. This paper, samples with replacement from typical survival data of the form: $(t_1, \epsilon_1), \ldots (t_i, \epsilon_i)$,

$\ldots(t_n, \epsilon_n)$, where $\epsilon_i = 1$ if the event occurred $\epsilon_i = 0$ if the event was censored, and t_i is the time of the event or censoring. Bootstrap samples, $(t_1^*, \epsilon_1^*), \ldots (t_i^*, \epsilon_i^*), \ldots (t_n^*, \epsilon_n^*)$ are then obtained and the Kaplan-Meier formula is applied to find a point estimate for each bootstrap sample. Then the variance is estimated by determining the sample variance of the Kaplan-Meier estimates across the bootstrap samples.

Bootstrap algorithms that account for clustering have been proposed by Davison and Hinckley (1997), and their performance has been studied when applied to clustered survival data by Xiao and Abrahamowicz (2010). Xiao and Abrahamowicz (2010) report the results of a simulation study comparing two bootstrap methods of estimating the variance of Cox model regression coefficients. They also constructed and compared Wald based confidence intervals for estimated hazard ratios.

The two bootstrap approaches of estimating variance differed in how they built their bootstrap samples. If there were C clusters, the first method builds a bootstrap sample by resampling with replacement from these C clusters, and including those observations from the sampled clusters in the bootstrap sample. For example, suppose there were 3 clusters and we sampled clusters 1, 2, and 1. Then the bootstrap sample would include all the observations from cluster 1 twice, the observations from cluster 2 once, and no observations from cluster 3.

The second method however, resamples both at the cluster and individual level. One first obtains C clusters by sampling with replacement, then for each sampled cluster, c, one samples with replacement from the N_c observations, to obtain the final bootstrap sample. Thus the second method is computationally more demanding than the first.

Xiao and Abrahamowicz (2010) found for both cases, non-informative right censoring does not affect the large sample properties of either bootstrap estimator. They calculated Wald based confidence intervals from each estimator. They found the second approach resulted in more conservative coverage. We will include both methods of constructing bootstrap samples in our simulation study. Xiao and Abrahamowicz (2010) did not notice a difference in performance between 100 and 500 bootstrap resamples so we will generate 200 bootstrap resamples,

$B_1, \ldots, B_{200}$. Denoting the cumulative incidence estimator for bootstrap sample i at a particular point t, $\hat{F}_1(t)_i$, our bootstrap variance estimators will be:

$$\widehat{var}_{boot}(\hat{F}_1(t)) = \frac{1}{199}\sum_{i=1}^{200}[\hat{F}_1(t)_i - \bar{\hat{F}}_1(t)_i]^2 \tag{2.9}$$

where

$$\bar{\hat{F}}_1(t)_i = \sum_{i=1}^{200}\frac{\hat{F}_1(t)_i}{200}$$

.

When the first method of resampling is used, we will denote the estimator $\widehat{var}_{boot1}(\hat{F}_1(t))$ and when the second method of resampling is used, we will denote our estimator $\widehat{var}_{boot2}(\hat{F}_1(t))$.

One limitation of all bootstrap estimators is that they give different results with each analysis. Differences should be small with 200 samples however.

2.4.5 Jackknife Variance Estimator

Quenouille (1949) developed the Jackknife method to correct the bias of estimators, but it was named and first used to calculate variance by Tukey (1956). Like nonparametric bootstrap methods it makes no distributional assumptions, but unlike the bootstrap, it is not computationally intensive.

Jackknife variance estimation is similar to bootstrap estimation in that the variance is calculated from point estimates applied to samples taken from an original dataset. Jackknife samples select portions of the data from the sample. The Jackknife has also been applied to estimate variance in settings where data is clustered (Gladen, 1979). In clustered data, with C clusters, there are C Jackknife samples, each including $C - 1$ of the available clusters.

The Jackknife has also been used to obtain variance estimates in correlated survival data. Again, a theme in reviewing the literature on variance estimation of correlated survival data is that much attention has been paid to the estimation of the variance of the regression coefficients from the Cox model. For example, Lipsitz et al. (1994) derive variance estimates for these coefficients using a Jackknife method. For the simulation study described in Chapter 3, we

use "a delete a group" Jackknife approach, described in Lipsitz et al. (1994) appropriate for clustered data. The estimator is described below.

Let $\hat{F}_1(t_j)_{-c}$ be the estimator of $F_1(t_j)$ using all the data except for that from cluster c. Then the Jackknife variance estimator for $\hat{F}_1(t_j)$ is:

$$\widehat{var}_{jack}(\hat{F}_1(t_j)) = \frac{C-1}{C}\sum_{c=1}^{C}(\hat{F}_1(t_j)_{-c} - \hat{F}_1(t_j))^2 \tag{2.10}$$

2.5 Confidence Interval Estimates

We will limit attention to Wald based procedures for simplicity. As mentioned in the introduction, we estimate 95% confidence interval estimates from the variance estimators described above. Two methods of confidence interval construction will be considered.

First, we can estimate a $1-\alpha$ confidence interval for the cumulative incidence function of event of interest, $F_1(t)$, based on the consistency and asymptotic normality of the cumulative incidence estimator in Section 2.3.1 as:

$$\hat{F}_1(t) \pm 1.96\sqrt{\widehat{var}(\hat{F}_1(t))} \tag{2.11}$$

We will denote this method as the linear confidence interval method, following the terminology of the confidence interval options of PROC LIFETEST in SAS 9.2. It has the advantage of being simple to construct and well-known. However, it may provide values outside the possible range of $F(t)$, $[0,1]$. Thus we will also consider the method of confidence interval construction developed by Kalbfleisch and Prentice (1980) termed the log-log method. Here, the confidence interval will be given by $\hat{F}_1(t)^{exp[\pm A]}$, where A is:

$$A = \frac{1.96\sqrt{\widehat{var}(\hat{F}_1(t))}}{\hat{F}_1(t)\log(\hat{F}_1(t))} \tag{2.12}$$

2.6 Summary

This literature review shows relatively little work has been done on nonparametric methods of variance estimation for the cumulative incidence in clustered competing risks data. Along with two existing methods of variance estimation for the cumulative incidence function, we have chosen four nonparametric methods of variance estimation used for correlated survival data and extended them to the clustered competing risks setting. The six variance estimators in Table 2.1 along with confidence intervals constructed from them will be compared via simulation study. The design of this study is described in Chapter 3 and the results are presented in Chapter 4.

Table 2.1: Variance Estimators

Notation	Variance Estimator
$\widehat{var}_{mult}(\hat{F}_1(t_j))$	Multinomial Based Estimator
$\widehat{var}_{cp}(\hat{F}_1(t_j))$	Counting Processes Based Estimator
$\widehat{var}_{lin}(\hat{F}_1(t_j))$	Linearized Estimator
$\widehat{var}_{jack}(\hat{F}_1(t_j))$	Jackknife Estimator
$\widehat{var}_{boot1}(\hat{F}_1(t_j))$	Bootstrap 1 Estimator
$\widehat{var}_{boot2}(\hat{F}_1(t_j))$	Bootstrap 2 Estimator

Chapter 3

Design of the Simulation Study

3.1 Introduction

In this chapter we describe the design of the simulation study, making sure to include all the relevant components discussed by Burton et al. (2006). There are seven sections in this chapter, including this introduction. Section 3.2 justifies selection of parameters for the generation of data used in the simulation study. Then Section 3.3 outlines the study design while Section 3.4 and 3.5 describe the model and method by which datasets are generated. Section 3.6 describes the criteria used to evaluate the methods of confidence interval construction and finally, Section 3.7 briefly summarizes the key elements of study design.

3.2 Parameters

The simulated datasets are designed to replicate an arm from a cluster randomized trial where patients from different centers are followed for 5 years until one of two competing events occurs or they are administratively censored due to trial completion. That is, all clusters and subjects are recruited prior to random assignment. This is done for simplicity of data generation and to avoid potential for selection bias. We are interested in estimating the cumulative incidence of the event of interest at study completion along with variances and 95% confidence intervals. Where possible, values for parameters will be motivated by data from the two studies mentioned in the Introduction: bone marrow transplant data discussed by Zhou et al. (2012)

and prostate cancer data described by Messing et al. (1999). The parameters specified to generate the simulated datasets are: the cumulative incidence, the hazard rate for the event of interest and for the competing risk, the percentage of patients censored, the intracluster correlation coefficient (ICC), the number of patients enrolled in the study, the size of each cluster. We will generate datasets based on concrete values for these parameters and calculate the cumulative incidence at 5 years.

The cumulative incidence is determined by specifying the hazards of the event and competing risk and the percentage of censoring. Based on our choices for these parameters, the simulation was conducted for cumulative incidence values of approximately: 0.25, 0.33, 0.37, 0.4, 0.54, and 0.6.

The hazard of a patient experiencing any event is determined by the percentage of events censored, fixing the sum of the hazards of the study outcome and the competing event. In specifying hazard rates of the event of interest and the competing risk, we wish to model scenarios where, like our motivating datasets described in the introduction, the event of interest occurs more often than the competing event. A constant hazard rate for both risks is also assumed. This may not be a large disadvantage; Chen et al. (2008) conducted a simulation study examining the bias of a cumulative incidence estimator under clustering and found the results with time varying hazards didn't significantly differ from those with fixed hazards. The ratio of patients experiencing the event of interest to the competing risk is between 1 and 3 in both motivating datasets and is completely determined by the ratio of their hazards. We will consider scenarios where the ratio is either 1, 2, or 3.

Attention is limited to administrative censoring, that is, censoring due to not all patients experiencing an event at 5 years when the cumulative incidence is calculated. We assume that all clusters and patients within clusters are enrolled prior to the start of the study. We adjusted the range of the censoring variable so that the percentage of observations censored is on average 20% or 50%.

Small values of the ICC are usually observed in cluster randomized trials where patients

are from different centers (Donner and Klar 2000). Recall an ICC = 0 indicates no intracluster correlation while an ICC = 1 means total dependence between responses from the same cluster. Thus we considered ICC values of 0.01, 0.05 and as a check an ICC of 0 as well.

We are interested in scenarios where there are either 400 or 100 patients enrolled in a study. In the motivating studies described in the Introduction, because the diseases are rare in the general population, the number of patients recruited per center is low and the number of centers is high. For simplicity, we considered cases where the same number of patients is recruited from each cluster. For the case when 100 patients are enrolled, cluster sizes of 5, 10, and 20 are chosen. This will give us scenarios where there are 5, 10, and 20 clusters from which patients are recruited. For the case when 400 patients are enrolled, cluster sizes of 10, 40, and 80 will be used. This will give scenarios where there are 5, 10, and 40 clusters from which patients are recruited.

The values of the parameters to be investigated are summarized in Table 3.1. Using these parameter values there are $3 \times 2 \times 3 \times 2 \times 3 = 3^3 2^2 = 108$, scenarios to investigate.

3.3 Simulation Procedures

For each scenario, 1000 independent datasets were generated and each of the methods of variance construction are applied to each dataset. The estimated cumulative incidence and the estimated variance estimators for each dataset were stored.

In the exceptional case that a dataset has only censored data (i.e. when there are no events in a dataset), it will be replaced. The probability of this happening is less than $(1/2)^{100}$ because the probability of an individual observation being censored is less than $1/2$.

The simulation study was conducted using R version 3.2.0 (R Core Team, 2015). The generation of random numbers was done using the "runif", "rbin", "rexp", and "rgamma" functions from the base package. A starting seed was specified for each scenario to ensure that scenarios are independent of each other and can be replicated. To further aid in replication, the code used for data generation and calculating estimators is provided in Appendix B.

Table 3.1: Simulation Parameters

Parameter	Definition	Values	Unique No. of Parameters
$F_1(t)$	Cumulative incidence	0.25, 0.33, 0.37, 0.4, 0.54, 0.6	-
λ_1/λ_2	Ratio of the hazard of outcome vs. competing event	1, 2, 3	3
P	Percentage of observations censored	20%, 50%	2
ρ	Intracluster correlation coefficient	0, .01, .05	3
N	Number of patients recruited	400, 100	2
s	Cluster size	5, 10, 20 (100 patients case) 10, 40, 80 (400 patient case)	3
C	Number of clusters	5, 10, 20 (100 patients case) 5, 10, and 40 (400 patient case)	-
t	Time confidence interval is determined	5 years (trial completion)	-

3.4 Modeling the Data

We will model the clustering effect using a shared gamma frailty model. The relationship between the ICC and the frailty parameter is given by Jeong and Jung (2006). For our purposes, it suffices to note that given C clusters, frailties $v_1, ..., v_C$ generated from a single-parameter gamma distribution with shape ρ and scale $1/\rho$, model a within cluster correlation of ρ. Thus we generate C frailties $v_1, ..., v_C$ from the model shown in Equation (3.1).

$$f(v;\rho) = \frac{v^{\rho-1}e^{\rho v}}{\Gamma(\rho)(1/\rho^{\rho})} \qquad v > 0, \quad \rho > 0 \tag{3.1}$$

Then for each cluster c, independent event times follow an exponential distribution with hazard $v_c(\lambda_1 + \lambda_2)$, conditional on the shared frailty.

3.5 Data Generation Procedure

Some investigators simulate competing risks data using a latent failure time model (e.g. Illjon (2013), Braun and Yuan (2007)). As described in the Introduction, we use only observable quantities for the analysis of competing risks, to minimize model assumptions. Thus we prefer the approach of Beyersmann (2009) over the latent failure approach. Beyersmann (2009)'s approach is described in our data generation algorithm below.

The data are generated so observations from the same cluster are correlated. The procedure for data generation based on the specified parameters is as follows:

1. Generate frailties, $v_1, ..., v_C$, from a single-parameter gamma distribution with shape ρ and scale $1/\rho$, which implies a within cluster correlation with the ICC, ρ.

2. We wish to generate data so that at time $t = 5$ years on average P percent experience neither event (i.e are censored). To do this, we must determine a value for the hazard of experiencing either event (i.e. $\lambda_1 + \lambda_2$). Because we have a fixed time $t = 5$, a fixed probability that P percentage of events are greater than 5, and event times follow an exponential distribution, we can determine a value for $\lambda_1 + \lambda_2$. Then we can determine values of λ_1 and λ_2 based on the ratio we specified for λ_1/λ_2.

3. For each cluster c generate s event times from an exponential distribution with hazard $v_c(\lambda_1 + \lambda_2)$. We now have observations $(t_{c1}, ..., t_{cN_c})$ where t_{ci} is the time of the event, for $i \in \{1, ..., N_c\}$.

4. After event times are generated for each cluster, assign to each event time an associated risk $\delta_{c1}, ..., \delta_{cN_c}$, where these can either take the value 1 or 2 to indicate the time is associated with the event or the competing risk respectively. These are generated from a binomial random variable, with the event assigned probability $\lambda_1/(\lambda_1 + \lambda_2)$ and the competing event assigned probability $\lambda_2/(\lambda_1 + \lambda_2)$. We now have observations $(t_{c1}, \delta_{c1}), ..., (t_{cN_c}, \delta_{cN_c})$.

5. If a time t_{ci} exceeds 5, replace its value with 5 and replace the value for δ_{c1} with 0, to indicate it has been administratively censored.

6. Return observations $(t_{c1}, \delta_{c1}), ..., (t_{cN_c}, \delta_{cN_c})$ for $c \in \{1, ..., C\}$.

3.6 Evaluation Criteria

The primary purpose of the simulation study is to evaluate the confidence interval estimators constructed from the variance estimators described in Section 2.3, but we will also test the performance of the variance estimators, as these are the variable component of the different methods of confidence interval construction.

3.6.1 Criteria to Assess Variance Estimators

Following Braun and Yuan (2007) and Chen et al. (2008), we compare the mean variance of each estimator to the empirical variance.

3.6.1.1 Percentage Bias of Estimated Variance

Given 1000 simulated datasets, and the prespecified time $t = 5$, we obtain estimators of the cumulative incidence $(\hat{F}_1(t)_i)_{i=1}^{1000}$. The empirical variance will be given by the formula:

$$EMV := \sum_{i=1}^{1000} [\hat{F}_1(t)_i - \bar{\hat{F}}_1(t)_i]^2/999 \tag{3.2}$$

where $\bar{\hat{F}}_1(t)_i = \sum_{i=1}^{1000} \frac{\hat{F}_1(t)_i}{1000}$.

The percentage bias is then approximated by the percentage difference between the empirical variance and the mean of a variance estimator over the different simulations. It is given by:

$$\frac{[\frac{1}{1000}\sum_{i=1}^{1000} \widehat{var}(\hat{F}_1(t)_i)] - EMV}{EMV} \tag{3.3}$$

We consider a percentage bias of magnitude over 15% to be a poor performance by a variance estimator.

3.6.2 Criteria to Assess Confidence Interval Estimators

We evaluate the confidence interval estimators on coverage and width. The latter is done by examining the width of the estimated confidence intervals, and the former by whether the true value of the cumulative incidence was in the confidence interval or missed to the left or right of the interval. Because the confidence interval estimators are for the cumulative incidence function, which in general is bounded by $[0, 1]$, if we estimate a lower bound below 0 or an upper bound above 1 they will be truncated to 0 or 1 respectively. This will have to be done with the linear method of confidence interval construction only, not the log-log method, which always gives results within these bounds.

3.6.2.1 Coverage

The confidence interval estimators will be examined by determining the proportion of times the confidence interval contained the true value of the cumulative incidence function at the specified point in time. The true value of the cumulative incidence function will be given by

$$F_1(t) = \int_{u=0}^{t} \lambda_1 S(u) du \tag{3.4}$$

where $S(t) = \exp\Big(-\int_{u=0}^{t}(\lambda_1 + \lambda_2)du\Big)$.

For a given scenario, variance estimator and method of confidence interval construction, let $[L_i, U_i]$ be the the confidence interval estimator for dataset i for $i = 1, ..., 1000$. The proportion

of times $F_1(t)$ is in the confidence interval is:

$$\frac{\#\{F_1(t) \in [L_i, U_i] : i = 1, ..., 1000\}}{1000} \tag{3.5}$$

A better performance is, all things equal, indicated by the proportion being closer to 95%. More specifically, following the method of Bradley (1978), we consider coverage outside the bounds set by $0.95 \pm 1.96\sqrt{0.95 \times (0.05/1000)}$, which is approximately between 93.5% and 96.5% to be poor.

We also consider the proportion of times the true value of the cumulative incidence function misses to the left and to the right. These are respectively:

$$\frac{\#\{F_1(t) < [L_i, U_i] : i = 1, ..., 1000\}}{1000} \tag{3.6}$$

and

$$\frac{\#\{F_1(t) > [L_i, U_i] : i = 1, ..., 1000\}}{1000} \tag{3.7}$$

It is preferable to have symmetry in proportion of misses to the left and right.

3.6.2.2 Width

To assess the precision of the confidence interval estimators we calculate their average length over the simulated datasets. This is:

$$\frac{1}{1000}\sum_{i=1}^{1000}(U_i - L_i) \tag{3.8}$$

The shorter the length of an estimator, the more precise the estimator will be. It is expected that methods that account for clustering, due to this, will have greater coverage but also greater width. If two estimators perform similarly in coverage, but one has a shorter width on average, it is be preferred.

3.7 Summary

The simulation study to be performed compares six variance estimators of the cumulative incidence, listed in Table 2.1, and the twelve confidence interval estimators built from these variance estimators. It compares them over a range of scenarios described in Section 3.2. The scenarios are motivated by the datasets described in Chapter 1 and analyzed in detail in Chapter 5.

Data will be simulated by the method given by Beyersmann (2009) to limit attention only to the modeling of observable quantities.

The variance estimators are compared on their percentage bias and the methods of confidence interval construction are compared on their coverage and width.

Chapter 4

Results of the Simulation Study

4.1 Introduction

Section 4.2 provides general remarks about the simulation and describes the manner in which the results are presented. Section 4.3 discusses the percentage bias performance of the variance estimators while Section 4.4 discusses the performance of the confidence interval estimators. Section 4.5 summarizes the main findings of this chapter and Section 4.6 contains tables of the simulation results.

4.2 Description of Simulation Results

Recall that 1000 replications were performed for each of the 108 scenarios distinguished by the parameters listed in Table 3.1. For each scenario, the cumulative incidence and the six variance estimators summarized in Table 2.1 were estimated. These variance estimators were the Multinomial Based Estimator, Counting Processes Based Estimator, Linearized Estimator, Jackknife Estimator, a cluster-sample Bootstrap Estimator, and a two-step Bootstrap Estimator. They are denoted Mult, CP, Lin, Jack, Boot1, and Boot2 respectively and are described in Chapter 2.

The performance of these variance estimators and the confidence intervals formed from them was evaluated as described in Section 3.6. First, the percentage bias of the variance estimators was calculated. Then, for both the linear and log-log methods of confidence interval

construction, and each of the six variance estimators, coverage, left and right misses, and mean confidence interval width were calculated. Recall from Section 3.3 that we would resimulate a dataset in the event that it contained only censored observations, as this would make the use of a cumulative incidence estimator impossible. This simulation had no occurrence of a dataset containing only censored observations and thus no resimulation was needed.

All simulation results are presented at the end of this chapter in Section 4.6. The simulation results for the percentage bias of the variance estimators are presented in Tables 4.1 to 4.6. While we discuss both the linear and log-log methods of confidence interval construction in this chapter, only the coverage of the confidence intervals (including left and right misses) formed using the log-log method are presented. These are given in Tables 4.7 to 4.12. The reason we do not present the tables for the linear method of confidence interval construction is given in Section 4.4. We also present in Tables 4.13 to 4.18, the average width of the log-log estimators.

Thus in Section 4.6, where we present the results of the simulation, there are three sets of tables for the 108 scenarios: one set for the percentage bias results, one set for the coverage of the log-log interval estimators, and one set for the width of the log-log interval estimators. For readability, each set of results is presented in six tables containing 18 scenarios (6x18=108), organized as follows: the first three tables are simulations where the total sample size was 100, the latter three where the total sample size was 400. For a fixed total sample size, each table presents the results for those scenarios where the ICC was 0, 0.01, or 0.05 respectively.

Thus a table contains, for a fixed total sample size and ICC, the results for the 18 scenarios created by taking all possible values of the remaining parameters of data generation. These were the value of the cumulative incidence at $t = 5$ (i.e. $F_1(5)$), the number of clusters C, the percentage of observations censored, and the ratio of the hazards of the event of interest and the competing risk (i.e. λ_1/λ_2).

Finally, we note one particularity of the results. Despite the Linearized and Jackknife variance estimators not being algebraically equivalent, in this simulation their average percentage

biases were indistinguishable and their coverage performance were identical regardless of use of the linear or log-log method of confidence interval construction.

4.3 Percentage Bias Results

This section discusses the the percentage bias simulation results of the variance estimators. These results are displayed in the tables Table 4.1 to Table 4.6. In these tables, those cases where a method had above 15% absolute bias were bolded.

We first examine the number of scenarios in which a variance estimator has the lowest absolute percentage bias. The Linearized and Jackknife variance estimators achieved the lowest absolute percentage bias the most number of times, in 54 of the 108 scenarios, while the Counting Process estimator achieved it the second most at 43. The Multinomial estimator achieved it under 10 times and the Bootstrap 1 estimator 14 times. Bootstrap 2 never had the lowest percentage bias.

Next we examine the percentage of scenarios for which an estimator has absolute percentage bias $\leq 15\%$. The Linearized and Jackknife estimator satisfy this 100% of the time. The Counting Processes estimator 81% and Bootstrap 1 at 69%, and Multinomial estimator 44% of times. Bootstrap 2 never achieves absolute bias under 15%.

We now consider the influence varying each parameter has on the absolute percentage bias of each variance estimator. As the ICC, ρ, increases the bias of Multinomial and Counting Process estimators increase, which is to be expected as these estimators do not account for clustering. The Linearized, Jackknife, and Bootstrap 1 estimators all have no noticeable change in bias, while the bias of the Bootstrap 2 estimator decreases as clustering increases.

As N, the total sample size, increases from 100 to 400, the Multinomial and Counting Process estimators increase in percentage bias, while other estimators show no noticeable change. This is likely due to an increase in the clustering effect, as those cases when $N = 400$ generally have larger clusters.

As $F_1(5)$ increases only the bias in the Multinomial estimator increases, other estimators seem to be unaffected.

As the number of clusters C increases, the absolute percentage bias generally decreases for all but the Bootstrap 2 estimator. The bias for the Linearized and Jackknife estimators, seems unaffected by the number of clusters increasing.

The percentage censored changing from 20% to 50%, appears not to have an effect except to decrease the mean absolute percentage bias for the Multinomial estimator. Changing the ratio of the hazards of the event of interest and competing risk has no clear effect on the percentage bias results.

4.4 Confidence Interval Results

This section discusses the performance of the confidence intervals formed from the variance estimators. First we compare the performance of the linear and log-log methods of confidence interval construction. Because the log-log method provides better coverage, we will consider in greater detail the coverage and width performance of the confidence interval estimators using the log-log method.

The coverage of the log-log confidence intervals for all scenarios are displayed in Table 4.7 to Table 4.12. In these tables, as discussed in Section 3.6.2.1, those cases where the coverage was approximately between 93.5% and 96.5% were thought to have good coverage. Instances outside this range are bolded.

4.4.1 Difference Between Linear and Log-Log Methods of Confidence Interval Construction

We compared the size of the difference in overall coverage between the linear and log-log confidence interval methods, for each of the 108 scenarios and 6 variance estimators. Thus there were $6 \times 108 = 648$ total comparisons of coverage performance. The difference in coverage between them exceeds 2% only 19 times in total (approx 3% of the cases). In no case did the difference in coverage between the linear and log-log exceed 3%.

For each method of variance estimation, we compared the number of scenarios where a variance estimator had acceptable coverage (i.e. between 93.5% and 96.5%) when a confidence

interval is constructed using the linear versus using the log-log method. This nearly always increased when the log-log method was used; increasing by 8 for the Multinomial estimator, by 4 for the Linearized and Jackknife estimator and by 5 when Bootstrap 1 is used. The number of scenarios with acceptable coverage decreased by 1 when Bootstrap 2 is used. Because of it's better performance, we only present tables of coverage and width for the log-log method of confidence interval construction.

4.4.2 Confidence Interval Performance

We now examine in greater detail, for only the log-log method of confidence interval construction, the performance of the different confidence interval estimators.

In 77 of 108 scenarios the Counting Process method provided coverage in our acceptable range. It is the only method that had acceptable coverage in over 50% of the scenarios. Of course, a third of our scenarios had no clustering effect whatsoever. Indeed, in scenarios with ICC of 0.05, it often performs significantly worse than Bootstrap 2.

Thus we will consider the influence varying the parameters ICC, sample size, number of clusters, and the value of the cumulative incidence has on the coverage of each variance estimator.

When the ICC is 0 the Counting Process estimator should clearly be preferred over the other estimators. Although when $F_1(5)$ is smaller, the Multinomial estimator has slightly better coverage. As the ICC, ρ, increases from 0 to 0.01, there is no great difference between coverage in any of the estimators. However as the ICC increases from 0.01 to 0.05, the number of scenarios with good coverage drops for both the Multinomial and Counting Processes estimators, which is to be expected, as these do not account for clustering. Even when $\rho = 0.05$, the Linearized, Jackknife and Bootstrap 1 estimators never achieve the required number of clusters and sample size to outperform the Counting Processes estimator when the sample size is 100. However, when the sample size $N = 400$ and $\rho = 0.05$, these estimators sometimes offer better coverage than Counting Process estimator. The Bootstrap 2 performs well in 4 scenarios where the rest of the estimators preformed poorly. All these scenarios are when $\rho = 0.05$, $N = 400$,

and $C = 5$. This suggests that in scenarios where there is a small number of large clusters and heavy clustering, the Bootstrap 2 may be the preferred estimator.

When the number of clusters C is less than or equal to 10, the Linearized, Jackknife and Bootstrap 1 estimators rarely have acceptable coverage, but when they are above 10, they have good coverage a number of times comparable to the Counting Process method, even under no clustering.

For estimators accounting for clustering there was no noticeable pattern that increasing the value of $F_1(5)$ had on coverage performance. However the Multinomial estimator performed better for lower values of $F_1(5)$ and the Counting Process estimator performed better for higher values.

The coverage of the confidence interval estimators varied significantly. As such, differences in left and right misses between estimators is not as valuable a consideration as if the performance of estimators was closer.

The average width of estimators corresponds with the coverage performance of the estimators. As expected, increasing the number of clusters does not change the average width of a confidence interval estimators not account for clustering. The width of the Bootstrap 2 had the widest average width over all scenarios, whose coverage was also always the largest. The Bootstrap 1 width was always lower than the Linearized and Jackknife widths.

4.5 Summary of Simulation Results

This simulation study found that the percentage bias of the Linearized and Jackknife estimators was lowest, but this did not correspond to their having good confidence interval coverage. This is perhaps because we examined many scenarios with a number of clusters C too small for their asymptotic performance to be noticeable.

The log-log method of confidence interval construction is to be preferred to the linear method.

The Counting Process estimator was clearly the best performer under no clustering and minimal clustering of 0.01. It even performs well when ICC = 0.05 when the sample size

is 100. However, under clustering of 0.05 and a sample size of 400, methods accounting for clustering showed better coverage. With a sample size of 400 and a low number of large clusters (i.e. 5 clusters of 40 patients each), the Bootstrap 2 method performs the best.

4.6 Simulation Results

Table 4.1: Percentage Bias for Variance Estimators Applied to Simulated Data

Sample Size = 100, ICC = 0

				Percentage Bias of Variance Estimator					
$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Mult	CP	Lin	Jack	Boot1	Boot2
0.25	20	0.5	1	-10	-5	-7	-7	-12	**64**
	10			1	7	5	5	-5	**91**
	5			0	6	4	4	**-17**	**83**
0.33	20		2	-5	1	3	3	-2	**77**
	10			-11	-6	-7	-7	**-16**	**68**
	5			-6	-1	-2	-2	**-21**	**73**
0.37	20		3	1	6	6	6	1	**85**
	10			-6	-1	-2	-2	-12	**77**
	5			2	7	8	8	-14	**87**
0.40	20	0.2	1	-15	4	3	3	-2	**80**
	10			**-23**	-5	-6	-6	**-16**	**69**
	5			**-20**	-1	-2	-2	**-22**	**72**
0.53	20		2	**-18**	4	1	1	-4	**79**
	10			-14	9	9	9	-2	**95**
	5			**-18**	3	3	3	**-18**	**79**
0.60	20		3	-14	9	8	8	3	**89**
	10			**-20**	1	0	0	-10	**81**
	5			**-21**	0	0	0	**-20**	**74**

Table 4.2: Percentage Bias for Variance Estimators Applied to Simulated Data

Sample Size = 100, ICC = 0.01

				Percentage Bias of Variance Estimator					
$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Mult	CP	Lin	Jack	Boot1	Boot2
0.25	20	0.5	1	-12	-7	-7	-7	-11	**62**
	10			-14	-10	-8	-8	**-17**	**63**
	5			-2	4	10	10	-12	**85**
0.33	20		2	-3	2	2	2	-3	**79**
	10			-4	2	4	4	-7	**84**
	5			-4	1	6	6	-15	**81**
0.37	20		3	-4	1	1	1	-4	**77**
	10			-10	-5	-3	-3	-12	**72**
	5			-9	-5	2	2	**-18**	**71**
0.40	20	0.2	1	**-22**	-5	-6	-6	-10	**65**
	10			**-18**	1	2	2	-9	**82**
	5			**-21**	-3	-1	-1	**-21**	**70**
0.53	20		2	**-27**	-8	-10	-10	-15	**59**
	10			**-20**	0	3	3	-8	**81**
	5			**-25**	-5	-1	-1	**-20**	**68**
0.60	20		3	**-25**	-5	-7	-7	-12	**63**
	10			**-17**	5	7	7	-3	**89**
	5			**-19**	2	7	7	-14	**81**

Table 4.3: Percentage Bias for Variance Estimators Applied to Simulated Data

Sample Size = 100, ICC = 0.05

				Percentage Bias of Variance Estimator					
$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Mult	CP	Lin	Jack	Boot1	Boot2
0.25	20	0.5	1	-2	3	6	6	1	**82**
	10			-6	-1	5	5	-6	**82**
	5			**-21**	**-16**	-6	-6	**-24**	**54**
0.33	20		2	-15	-10	-6	-6	-10	**59**
	10			-13	-8	0	0	-10	**71**
	5			**-18**	-14	6	6	**-16**	**66**
0.37	20		3	-7	-2	2	2	-3	**74**
	10			-10	-5	7	7	-4	**81**
	5			**-27**	**-23**	-4	-4	**-24**	**49**
0.40	20	0.2	1	-15	3	5	5	-1	**80**
	10			**-18**	0	2	2	-8	**80**
	5			**-28**	-11	-4	-4	**-24**	**60**
0.53	20		2	**-18**	2	5	5	0	**81**
	10			**-21**	-1	5	5	-6	**81**
	5			**-35**	**-19**	-7	-7	**-26**	**50**
0.60	20		3	**-25**	-7	-4	-4	-9	**65**
	10			**-33**	**-17**	-11	-11	**-20**	**53**
	5			**-37**	**-21**	-3	-3	**-22**	**50**

Table 4.4: Percentage Bias for Variance Estimators Applied to Simulated Data

Sample Size = 400, ICC = 0

				Percentage Bias of Variance Estimator					
$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Mult	CP	Lin	Jack	Boot1	Boot2
0.25	40	0.5	1	-13	-9	-8	-8	-10	**72**
	10			-8	-4	-6	-6	-15	**78**
	5			2	6	10	10	-13	**91**
0.33	40		2	0	4	6	6	3	**96**
	10			0	4	4	4	-7	**94**
	5			-6	-2	-5	-5	**-24**	**73**
0.37	40		3	-4	0	0	0	-2	**86**
	10			-4	0	0	0	-11	**87**
	5			0	4	3	3	**-18**	**84**
0.40	40	0.2	1	**-19**	-2	-2	-2	-4	**83**
	10			-14	3	4	4	-6	**93**
	5			**-17**	0	-1	-1	**-20**	**79**
0.53	40		2	**-20**	-1	0	0	-3	**86**
	10			**-21**	-2	-4	-4	-14	**82**
	5			**-24**	-5	-6	-6	**-25**	**68**
0.60	40		3	**-18**	2	1	1	-2	**90**
	10			**-19**	1	3	3	-8	**90**
	5			**-25**	-8	-10	-10	**-28**	**63**

Table 4.5: Percentage Bias for Variance Estimators Applied to Simulated Data

Sample Size = 400, ICC = 0.01

				Percentage Bias of Variance Estimator					
$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Mult	CP	Lin	Jack	Boot1	Boot2
0.25	40	0.5	1	-6	-2	0	0	-3	**85**
	10			-12	-8	-3	-3	-12	**77**
	5			**-19**	**-16**	-3	-3	**-22**	**60**
0.33	40		2	-3	1	3	3	0	**90**
	10			-12	-8	3	3	-7	**82**
	5			**-25**	**-22**	-9	-9	**-27**	**50**
0.37	40		3	0	3	6	6	3	**96**
	10			-10	-6	2	2	-8	**82**
	5			**-21**	**-18**	0	0	**-20**	**61**
0.40	40	0.2	1	**-19**	-2	-2	-2	-4	**84**
	10			**-23**	-7	-3	-3	-12	**77**
	5			**-20**	-4	5	5	**-16**	**79**
0.53	40		2	**-23**	-5	-5	-5	-7	**79**
	10			**-24**	-6	-1	-1	-11	**80**
	5			**-25**	-7	8	8	-14	**77**
0.60	40		3	**-17**	3	4	4	2	**93**
	10			**-25**	-7	4	4	-7	**83**
	5			**-34**	**-18**	-6	-6	**-25**	**55**

Table 4.6: Percentage Bias for Variance Estimators Applied to Simulated Data

Sample Size = 400, ICC = 0.05

				Percentage Bias of Variance Estimator					
$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Mult	CP	Lin	Jack	Boot1	Boot2
0.25	40	0.5	1	-11	-8	-2	-2	-5	**78**
	10			**-27**	**-24**	-4	-4	-14	**59**
	5			**-39**	**-37**	3	3	**-17**	**44**
0.33	40		2	-8	-4	6	6	3	**89**
	10			**-34**	**-31**	1	1	-10	**57**
	5			**-51**	**-49**	3	3	**-17**	**32**
0.37	40		3	-7	-3	7	7	4	**91**
	10			**-31**	**-28**	10	10	0	**67**
	5			**-55**	**-54**	-1	-1	**-20**	**25**
0.40	40	0.2	1	**-16**	0	5	5	2	**91**
	10			**-39**	**-27**	-13	-13	**-22**	**49**
	5			**-41**	**-29**	6	6	-15	**53**
0.53	40		2	**-19**	-1	6	6	3	**91**
	10			**-42**	**-29**	-5	-5	-15	**54**
	5			**-51**	**-40**	2	2	**-18**	**40**
0.60	40		3	**-21**	-3	6	6	3	**88**
	10			**-44**	**-31**	1	1	-9	**56**
	5			**-56**	**-46**	5	5	**-16**	**37**

Table 4.7: Coverage of log-log Confidence Intervals

Sample Size = 100, ICC = 0

$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Percentage Coverage of Confidence Interval																
				Mult			CP			Lin			Jack			Boot1			Boot2		
				Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right
0.25	20	0.5	1	3	94	4	2	96	2	2	94	4	2	94	4	3	**93**	4	0	**99**	1
	10			2	95	3	1	**97**	2	3	**92**	4	3	**92**	4	4	**91**	5	1	**99**	1
	5			3	94	4	2	**97**	2	6	**87**	7	6	**87**	7	7	**84**	9	0	**99**	1
0.33	20		2	3	94	4	2	95	4	2	94	4	2	94	4	3	**93**	4	0	**99**	1
	10			3	**93**	4	3	94	4	4	**91**	5	4	**91**	5	5	**89**	6	1	**98**	1
	5			3	94	3	2	95	3	6	**88**	6	6	**88**	6	8	**84**	8	0	**99**	1
0.37	20		3	2	95	3	2	96	3	2	95	3	2	95	3	2	94	3	1	**99**	1
	10			3	95	3	2	95	3	3	**92**	5	3	**92**	5	4	**91**	5	1	**99**	1
	5			2	95	3	1	96	3	5	**88**	6	5	**88**	6	6	**86**	7	0	**98**	2
0.40	20	0.2	1	4	**93**	4	2	96	2	3	94	3	3	94	3	3	94	3	0	**99**	1
	10			5	**92**	3	2	96	2	5	**92**	4	5	**92**	4	5	**90**	5	1	**99**	1
	5			4	**92**	4	2	96	2	5	**88**	7	5	**88**	7	6	**85**	9	0	**99**	1
0.53	20		2	4	**93**	4	2	96	3	3	94	4	3	94	4	3	**93**	4	0	**99**	1
	10			3	94	3	2	**97**	1	3	**93**	4	3	**93**	4	4	**92**	4	0	**99**	1
	5			4	**92**	4	2	95	2	5	**87**	8	5	**87**	8	7	**83**	10	1	**98**	1
0.60	20		3	3	**93**	4	2	96	3	2	95	3	2	95	3	3	94	3	0	**99**	1
	10			4	**92**	5	2	95	3	3	**92**	5	3	**92**	5	4	**91**	5	0	**99**	1
	5			4	**91**	4	3	95	3	5	**87**	8	5	**87**	8	7	**83**	10	0	**99**	1

Table 4.8: Coverage of log-log Confidence Intervals

Sample Size = 100, ICC = 0.01

$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Percentage Coverage of Confidence Interval																
				Mult			CP			Lin			Jack			Boot1			Boot2		
				Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right
0.25	20	0.5	1	4	**92**	4	2	95	3	4	**93**	4	4	**93**	4	4	**92**	4	1	**99**	1
	10			3	**92**	5	3	95	3	4	**90**	6	4	**90**	6	5	**88**	7	1	**99**	1
	5			3	94	4	1	**97**	2	7	**87**	6	7	**87**	6	8	**85**	8	1	**98**	1
0.33	20		2	3	95	2	2	96	2	2	95	3	2	95	3	2	94	3	1	**99**	1
	10			3	94	3	1	96	3	3	**93**	4	3	**93**	4	4	**92**	5	0	**99**	1
	5			2	94	4	1	95	4	5	**88**	6	5	**88**	6	7	**85**	8	0	**98**	2
0.37	20		3	2	95	3	2	95	3	2	95	4	2	95	4	2	94	4	0	**99**	1
	10			3	94	3	3	94	3	4	**91**	5	4	**91**	5	4	**90**	6	1	**99**	1
	5			3	94	3	3	94	3	7	**87**	6	7	**87**	6	8	**85**	7	1	**98**	1
0.40	20	0.2	1	4	**91**	5	2	95	4	3	**93**	4	3	**93**	4	3	**93**	4	1	**99**	1
	10			2	**93**	5	1	96	3	3	**93**	4	3	**93**	4	4	**91**	5	0	**99**	1
	5			4	**92**	5	1	95	3	5	**87**	8	5	**87**	8	6	**84**	10	0	**99**	1
0.53	20		2	4	**91**	5	2	95	3	3	**93**	5	3	**93**	5	4	**91**	5	1	**98**	1
	10			4	**92**	4	2	95	3	3	**93**	5	3	**93**	5	3	**92**	5	0	**99**	1
	5			4	**92**	4	3	95	3	5	**87**	7	5	**87**	7	6	**85**	9	0	**99**	1
0.60	20		3	4	**91**	5	2	94	4	2	94	4	2	94	4	3	**93**	4	0	**99**	1
	10			2	**92**	6	1	95	4	2	**93**	5	2	**93**	5	2	**92**	6	0	**99**	1
	5			4	**92**	5	2	95	3	5	**89**	7	5	**89**	7	6	**86**	8	0	**99**	1

Table 4.9: Coverage of log-log Confidence Intervals

Sample Size = 100, ICC = 0.05

$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Percentage Coverage of Confidence Interval																
				Mult			CP			Lin			Jack			Boot1			Boot2		
				Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right
0.25	20	0.5	1	3	**93**	4	2	96	2	2	94	4	2	94	4	3	**93**	5	0	**99**	1
	10			2	**93**	5	1	96	2	3	**93**	5	3	**93**	5	3	**91**	5	0	**99**	1
	5			4	**90**	7	3	**93**	4	5	**87**	8	5	**87**	8	6	**84**	10	1	**97**	2
0.33	20		2	4	94	3	3	95	3	4	**92**	4	4	**92**	4	4	**92**	4	1	**98**	1
	10			3	94	3	2	95	3	3	**93**	4	3	**93**	4	4	**91**	5	0	**99**	1
	5			3	**92**	4	2	94	4	4	**89**	8	4	**89**	8	7	**84**	10	1	**98**	1
0.37	20		3	2	94	4	2	95	4	2	95	3	2	95	3	3	**93**	4	0	**99**	1
	10			1	95	4	1	95	4	3	**93**	4	3	**93**	4	3	**92**	6	0	**99**	1
	5			3	**91**	6	3	**92**	6	5	**88**	7	5	**88**	7	7	**84**	9	1	**98**	2
0.40	20	0.2	1	3	**93**	4	1	96	3	1	95	4	1	95	4	2	94	4	0	**99**	1
	10			2	**92**	6	1	95	4	2	**92**	6	2	**92**	6	3	**90**	7	0	**99**	1
	5			4	**90**	6	2	94	4	6	**87**	7	6	**87**	7	7	**85**	9	0	**98**	1
0.53	20		2	3	**93**	5	1	95	3	1	94	5	1	94	5	1	94	5	0	**99**	1
	10			3	**91**	6	2	95	4	2	**92**	6	2	**92**	6	2	**91**	7	0	**98**	1
	5			5	**89**	6	3	**92**	5	5	**88**	7	5	**88**	7	6	**84**	9	0	**98**	2
0.60	20		3	3	**90**	7	2	**93**	5	2	**93**	5	2	**93**	5	2	**93**	5	0	**99**	1
	10			3	**89**	8	2	**92**	7	2	**91**	7	2	**91**	7	3	**89**	8	0	**97**	3
	5			4	**86**	10	3	**90**	7	5	**87**	9	5	**87**	9	6	**83**	11	0	**97**	3

Table 4.10: Coverage of log-log Confidence Intervals

Sample Size = 400, ICC = 0

$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Percentage Coverage of Confidence Interval																
				Mult			CP			Lin			Jack			Boot1			Boot2		
				Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right
0.25	40	0.5	1	3	94	3	3	94	3	3	**93**	4	3	**93**	4	3	**93**	4	0	**99**	1
	10			2	95	4	2	95	4	4	**92**	4	4	**92**	4	4	**91**	5	0	**99**	1
	5			2	96	3	2	96	3	4	**89**	7	4	**89**	7	7	**86**	8	0	**99**	1
0.33	40		2	2	95	3	2	96	3	3	95	3	3	95	3	3	95	3	0	**99**	1
	10			2	95	3	1	96	3	3	**92**	4	3	**92**	4	4	**91**	5	0	**99**	0
	5			3	95	3	2	95	3	5	**87**	8	5	**87**	8	7	**84**	9	1	**98**	1
0.37	40		3	3	94	3	3	95	2	3	94	3	3	94	3	3	94	3	1	**99**	1
	10			2	95	3	2	96	2	3	**92**	5	3	**92**	5	4	**90**	6	0	**99**	1
	5			2	96	3	2	96	2	5	**89**	6	5	**89**	6	7	**86**	8	0	**99**	1
0.40	40	0.2	1	3	**92**	5	2	94	4	2	**93**	4	2	**93**	4	3	**93**	5	0	**99**	1
	10			4	**93**	3	2	96	2	4	**93**	4	4	**93**	4	4	**92**	4	1	**99**	0
	5			3	**93**	4	2	95	3	6	**87**	7	6	**87**	7	7	**84**	9	1	**99**	1
0.53	40		2	4	**92**	4	2	95	3	2	95	3	2	95	3	3	94	3	0	**99**	0
	10			4	**92**	4	3	94	3	5	**91**	4	5	**91**	4	5	**89**	5	1	**99**	1
	5			3	**92**	4	2	95	4	6	**87**	7	6	**87**	7	8	**83**	9	0	**98**	2
0.60	40		3	4	**93**	3	2	96	2	2	95	3	2	95	3	2	95	3	0	**100**	0
	10			5	**92**	3	3	96	2	4	**93**	3	4	**93**	3	5	**92**	4	1	**99**	0
	5			5	**90**	5	3	**93**	4	7	**85**	8	7	**85**	8	9	**82**	9	1	**98**	1

Table 4.11: Coverage of log-log Confidence Intervals

Sample Size = 400, ICC = 0.01

$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Percentage Coverage of Confidence Interval																
				Mult			CP			Lin			Jack			Boot1			Boot2		
				Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right
0.25	40	0.5	1	2	95	3	2	95	3	2	95	3	2	95	3	3	94	4	0	**99**	0
	10			3	94	3	3	94	3	3	**92**	5	3	**92**	5	4	**90**	6	1	**98**	1
	5			4	**92**	4	4	**92**	4	6	**87**	7	6	**87**	7	8	**83**	9	1	**98**	1
0.33	40		2	3	94	3	2	95	3	3	94	3	3	94	3	3	94	3	0	**99**	0
	10			2	94	4	2	94	4	3	**92**	5	3	**92**	5	5	**90**	6	1	**99**	1
	5			4	**91**	6	3	**92**	6	5	**88**	7	5	**88**	7	7	**85**	8	1	**97**	2
0.37	40		3	2	95	3	2	96	2	2	96	2	2	96	2	2	96	2	0	**100**	0
	10			2	95	3	2	95	3	3	**93**	4	3	**93**	4	4	**92**	5	1	**99**	1
	5			3	**92**	5	3	**93**	5	5	**88**	7	5	**88**	7	7	**85**	9	1	**98**	1
0.40	40	0.2	1	4	**93**	4	3	95	3	3	94	3	3	94	3	3	94	3	1	**99**	1
	10			4	**91**	5	3	**93**	4	3	**93**	4	3	**93**	4	4	**91**	5	1	**99**	1
	5			4	**92**	4	3	94	3	6	**89**	6	6	**89**	6	6	**86**	8	1	**99**	1
0.53	40		2	3	**91**	6	2	94	4	3	94	4	3	94	4	3	**93**	4	0	**99**	1
	10			4	**91**	5	2	94	5	4	**91**	5	4	**91**	5	4	**89**	7	0	**99**	1
	5			3	**91**	6	2	**93**	5	5	**88**	7	5	**88**	7	7	**85**	8	0	**98**	1
0.60	40		3	3	**93**	4	1	96	3	1	96	3	1	96	3	2	96	3	0	**99**	1
	10			3	**92**	5	2	95	3	4	**92**	5	4	**92**	5	4	**91**	5	0	**99**	1
	5			5	**90**	5	3	94	3	7	**87**	6	7	**87**	6	9	**83**	8	2	**98**	1

Table 4.12: Coverage of log-log Confidence Intervals

Sample Size = 400, ICC = 0.05

$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Percentage Coverage of Confidence Interval																
				Mult			CP			Lin			Jack			Boot1			Boot2		
				Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right	Left	Within	Right
0.25	40	0.5	1	3	**93**	4	3	**93**	4	2	**93**	5	2	**93**	5	2	**93**	5	0	**99**	1
	10			3	**92**	5	3	**92**	5	3	**91**	6	3	**91**	6	4	**90**	7	1	**98**	1
	5			4	**89**	8	4	**89**	8	5	**88**	7	5	**88**	7	6	**85**	10	1	**97**	3
0.33	40		2	3	**93**	4	2	94	4	2	95	3	2	95	3	2	94	4	0	**99**	1
	10			5	**89**	7	4	**89**	7	3	**92**	5	3	**92**	5	3	**90**	7	1	**98**	2
	5			7	**83**	10	6	**84**	10	6	**88**	7	6	**88**	7	7	**85**	8	2	95	3
0.37	40		3	2	94	4	2	95	3	2	95	3	2	95	3	2	94	4	0	**99**	1
	10			3	**89**	8	3	**89**	7	2	**92**	6	2	**92**	6	3	**91**	6	1	**98**	1
	5			9	**79**	12	9	**81**	11	6	**87**	7	6	**87**	7	8	**84**	8	2	95	3
0.40	40	0.2	1	2	**92**	7	1	94	5	1	94	5	1	94	5	1	**93**	5	0	**99**	1
	10			4	**87**	9	3	**89**	8	3	**90**	7	3	**90**	7	4	**88**	8	1	**97**	2
	5			4	**86**	10	3	**88**	9	4	**87**	9	4	**87**	9	5	**83**	12	1	**97**	3
0.53	40		2	2	**91**	8	1	**93**	6	1	**93**	6	1	**93**	6	2	**93**	6	0	**99**	1
	10			4	**85**	11	2	**88**	10	3	**89**	8	3	**89**	8	3	**88**	9	0	**97**	3
	5			6	**82**	12	4	**86**	11	4	**88**	8	4	**88**	8	6	**85**	10	1	96	3
0.60	40		3	2	**90**	8	1	**93**	5	1	**93**	6	1	**93**	6	1	**93**	7	0	**99**	1
	10			4	**83**	12	3	**88**	9	2	**91**	7	2	**91**	7	2	**90**	8	0	**97**	3
	5			6	**79**	15	3	**85**	12	4	**89**	8	4	**89**	8	5	**86**	10	1	95	4

Table 4.13: Width of log-log Confidence Intervals

Sample Size = 100, ICC = 0

				Average Length of Confidence Interval					
$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Mult	CP	Lin	Jack	Boot1	Boot2
0.25	20	0.5	1	0.164	0.168	0.164	0.164	0.160	0.218
	10			0.163	0.168	0.162	0.162	0.154	0.221
	5			0.163	0.168	0.156	0.156	0.140	0.216
0.33	20		2	0.178	0.183	0.183	0.183	0.179	0.240
	10			0.178	0.183	0.178	0.178	0.169	0.241
	5			0.178	0.183	0.171	0.171	0.153	0.237
0.37	20		3	0.183	0.188	0.186	0.186	0.181	0.245
	10			0.183	0.188	0.183	0.183	0.173	0.248
	5			0.183	0.188	0.176	0.176	0.158	0.243
0.40	20	0.2	1	0.173	0.191	0.188	0.188	0.183	0.249
	10			0.173	0.191	0.185	0.185	0.175	0.252
	5			0.173	0.191	0.179	0.179	0.160	0.247
0.53	20		2	0.174	0.195	0.191	0.191	0.186	0.254
	10			0.174	0.196	0.191	0.191	0.181	0.258
	5			0.174	0.196	0.183	0.183	0.164	0.252
0.60	20		3	0.171	0.193	0.189	0.189	0.185	0.252
	10			0.171	0.193	0.186	0.186	0.176	0.255
	5			0.172	0.193	0.181	0.181	0.162	0.250

Table 4.14: Width of log-log Confidence Intervals

Sample Size = 100, ICC = 0.01

				Average Length of Confidence Interval					
$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Mult	CP	Lin	Jack	Boot1	Boot2
0.25	20	0.5	1	0.163	0.168	0.166	0.166	0.162	0.219
	10			0.163	0.167	0.164	0.164	0.155	0.220
	5			0.164	0.168	0.162	0.162	0.145	0.220
0.33	20		2	0.178	0.183	0.181	0.181	0.177	0.240
	10			0.178	0.183	0.180	0.180	0.171	0.242
	5			0.178	0.183	0.176	0.176	0.157	0.239
0.37	20		3	0.183	0.187	0.185	0.185	0.180	0.245
	10			0.183	0.188	0.185	0.185	0.176	0.249
	5			0.183	0.188	0.183	0.183	0.164	0.246
0.40	20	0.2	1	0.173	0.191	0.188	0.188	0.183	0.249
	10			0.172	0.191	0.187	0.187	0.177	0.253
	5			0.172	0.191	0.181	0.181	0.162	0.247
0.53	20		2	0.174	0.195	0.191	0.191	0.186	0.254
	10			0.174	0.196	0.193	0.193	0.183	0.260
	5			0.174	0.195	0.187	0.187	0.168	0.255
0.60	20		3	0.172	0.193	0.189	0.189	0.184	0.251
	10			0.172	0.193	0.191	0.191	0.181	0.257
	5			0.172	0.193	0.187	0.187	0.167	0.253

Table 4.15: Width of log-log Confidence Intervals

Sample Size = 100, ICC = 0.05

				Average Length of Confidence Interval					
$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Mult	CP	Lin	Jack	Boot1	Boot2
0.25	20	0.5	1	0.163	0.167	0.167	0.167	0.163	0.220
	10			0.163	0.167	0.168	0.168	0.159	0.223
	5			0.163	0.167	0.166	0.166	0.149	0.221
0.33	20		2	0.178	0.183	0.185	0.185	0.181	0.241
	10			0.178	0.182	0.185	0.185	0.176	0.244
	5			0.178	0.182	0.189	0.189	0.169	0.246
0.37	20		3	0.183	0.188	0.189	0.189	0.184	0.247
	10			0.183	0.187	0.193	0.193	0.183	0.254
	5			0.183	0.187	0.196	0.196	0.175	0.253
0.40	20	0.2	1	0.173	0.191	0.190	0.190	0.185	0.249
	10			0.173	0.190	0.188	0.188	0.178	0.252
	5			0.173	0.191	0.186	0.186	0.166	0.250
0.53	20		2	0.175	0.196	0.196	0.196	0.191	0.258
	10			0.175	0.196	0.196	0.196	0.186	0.261
	5			0.175	0.195	0.196	0.196	0.176	0.260
0.60	20		3	0.173	0.193	0.194	0.194	0.189	0.255
	10			0.173	0.193	0.195	0.195	0.184	0.259
	5			0.173	0.193	0.202	0.202	0.181	0.261

Table 4.16: Width of log-log Confidence Intervals

Sample Size = 400, ICC = 0

				Average Length of Confidence Interval					
$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Mult	CP	Lin	Jack	Boot1	Boot2
0.25	40	0.5	1	0.083	0.085	0.084	0.084	0.083	0.116
	10			0.083	0.085	0.082	0.082	0.077	0.114
	5			0.083	0.085	0.081	0.081	0.072	0.112
0.33	40		2	0.090	0.092	0.092	0.092	0.091	0.126
	10			0.090	0.092	0.090	0.090	0.085	0.125
	5			0.090	0.092	0.086	0.086	0.076	0.121
0.37	40		3	0.093	0.095	0.094	0.094	0.093	0.129
	10			0.093	0.094	0.092	0.092	0.087	0.128
	5			0.093	0.095	0.088	0.088	0.079	0.124
0.40	40	0.2	1	0.087	0.096	0.095	0.095	0.094	0.130
	10			0.087	0.096	0.094	0.094	0.089	0.130
	5			0.087	0.096	0.090	0.090	0.080	0.126
0.53	40		2	0.088	0.098	0.098	0.098	0.096	0.133
	10			0.088	0.098	0.094	0.094	0.089	0.132
	5			0.088	0.098	0.091	0.091	0.081	0.128
0.60	40		3	0.086	0.096	0.095	0.095	0.094	0.131
	10			0.086	0.096	0.094	0.094	0.089	0.131
	5			0.087	0.096	0.090	0.090	0.080	0.126

Table 4.17: Width of log-log Confidence Intervals

Sample Size = 400, ICC = 0.01

				Average Length of Confidence Interval					
$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Mult	CP	Lin	Jack	Boot1	Boot2
0.25	40	0.5	1	0.083	0.085	0.085	0.085	0.084	0.116
	10			0.083	0.085	0.084	0.084	0.080	0.116
	5			0.083	0.085	0.085	0.085	0.076	0.115
0.33	40		2	0.090	0.092	0.092	0.092	0.091	0.126
	10			0.090	0.092	0.095	0.095	0.090	0.128
	5			0.090	0.092	0.094	0.094	0.084	0.125
0.37	40		3	0.093	0.095	0.095	0.095	0.094	0.129
	10			0.093	0.095	0.096	0.096	0.091	0.130
	5			0.093	0.094	0.098	0.098	0.088	0.130
0.40	40	0.2	1	0.087	0.096	0.096	0.096	0.094	0.131
	10			0.087	0.096	0.096	0.096	0.091	0.131
	5			0.087	0.096	0.094	0.094	0.084	0.129
0.53	40		2	0.088	0.098	0.098	0.098	0.096	0.134
	10			0.088	0.098	0.098	0.098	0.092	0.134
	5			0.088	0.098	0.099	0.099	0.088	0.133
0.60	40		3	0.087	0.096	0.096	0.096	0.095	0.131
	10			0.087	0.096	0.099	0.099	0.094	0.134
	5			0.087	0.096	0.097	0.097	0.087	0.131

Table 4.18: Width of log-log Confidence Intervals

Sample Size = 400, ICC = 0.05

				Average Length of Confidence Interval					
$F_1(5)$	C	Per. Cen.	λ_1/λ_2	Mult	CP	Lin	Jack	Boot1	Boot2
0.25	40	0.5	1	0.083	0.084	0.086	0.086	0.085	0.116
	10			0.083	0.084	0.092	0.092	0.087	0.121
	5			0.082	0.084	0.101	0.101	0.090	0.124
0.33	40		2	0.090	0.092	0.096	0.096	0.095	0.128
	10			0.090	0.092	0.108	0.108	0.102	0.137
	5			0.090	0.092	0.122	0.122	0.110	0.144
0.37	40		3	0.093	0.094	0.099	0.099	0.097	0.132
	10			0.093	0.094	0.114	0.114	0.108	0.142
	5			0.092	0.094	0.130	0.130	0.116	0.150
0.40	40	0.2	1	0.088	0.096	0.097	0.097	0.095	0.131
	10			0.088	0.096	0.101	0.101	0.096	0.135
	5			0.087	0.096	0.109	0.109	0.098	0.138
0.53	40		2	0.089	0.098	0.101	0.101	0.100	0.136
	10			0.088	0.098	0.110	0.110	0.104	0.142
	5			0.088	0.098	0.120	0.120	0.108	0.146
0.60	40		3	0.087	0.097	0.100	0.100	0.099	0.134
	10			0.087	0.096	0.114	0.114	0.108	0.144
	5			0.087	0.096	0.127	0.127	0.113	0.150

Chapter 5

Examples

5.1 Introduction

This chapter illustrates the use of confidence interval estimators for the cumulative incidence function, evaluated by simulation in Chapter 4. The data used in this chapter come from two multicentre studies with competing risks outcomes, which were highlighted in the Introduction. The first dataset is from a prospective and observational study (Zhou et al., 2012) and the second dataset is the control arm of an individually randomized trial (Messing et al., 1999).

Data from the control arm of a cluster randomized trial with competing risks outcomes would be the ideal but we were unable to obtain such a dataset. Nevertheless, analysis of the datasets here still provides insight into the performance of the estimators compared previously.

Section 5.2 and 5.3 provide context and statistical analysis for the data from each respective study. In the first study, patients with leukemia undergoing bone marrow transplant are followed until the occurrence of graft versus host disease. In the second study, patients in remission from prostate cancer having nodal metastasis were randomized to either immediately receive hormonal therapy or to only receive therapy after the disease progressed sufficiently. For purposes of illustration, attention is limited only to estimating the incidence of death from prostate cancer in the control arm of the study. Section 5.4, the final section, summarizes what can be learned about the variance and confidence interval estimators from these applications.

5.2 Bone Marrow Transplantation Dataset

5.2.1 Background

The bone marrow transplantation data we will use for illustration are a random subsample of 400 patients from a dataset analyzed by Zhou et al. (2012). These data are obtained from the R package "crrSC".

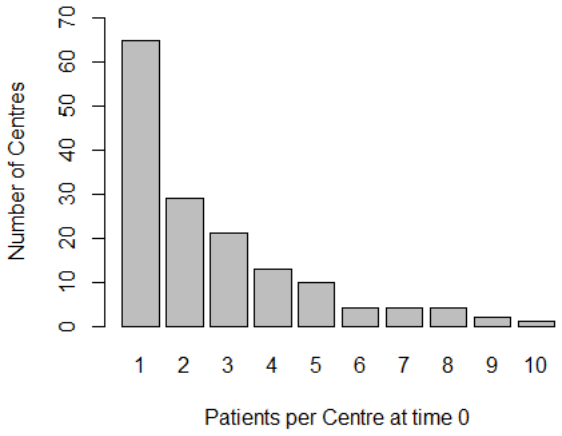
Data are from a bone marrow transplant registry kept by the European Blood and Bone Marrow Transplant Group (EBMT). The EBMT is a working group where greater than 450 transplant centres volunteer to once a year report stem cells transplantations and follow-up. Bone marrow transplantation, a common treatment for leukemia, has a possible side effect known as graft versus host disease and occurs following a bone marrow graft when the graft cells (from the donor) attack the host cells. It can lead to a quicker death than leukemia itself (Klingebiel and Schlegel, 1998). The data follows patients in remission from leukemia recruited between January 1, 1994 and December 31, 2004, followed until July 2008. Follow-up begins from the time of the bone marrow graft to the occurrence of either the event of interest, graft versus host disease (GVHD), or the competing risks of death (from any cause) or cancer relapse without GVHD. Events were administratively censored at July 2008.

Details on the inclusion criteria of the study subjects can be found in Katashian et al. (2006), the most important being: patients must be above 16, in full remission from acute myeloid leukemia, and that centers with only one patient were excluded (the reason for this was not given). Katashian et al. (2006) also notes that patient populations are very different across the centers, and that patients from the same center may share traits. Thus clustering effects may cause underestimation of variance estimates should one incorrectly assume independence of patients from the same center.

5.2.2 Descriptive Summary of Data

The data has three variables: time of event (in days), event indicator (GVHD, death without GVHD, or censored) and the center ID.

Figure 5.1: Number of Patients per Center in EBMT Dataset



Zhou et al. (2012) analyzed data for 2996 patients from 244 centers. The random sample we analyze has 400 patients from 153 centers. Figure 5.1 is a chart displaying the number of patients recruited per center. Unlike the dataset analyzed by Zhou et al. (2012), our random subsample contains centers with only one patient. The mean number of patients per center was approximately 2.6, and the median was 2.

There were 194 patients with an occurrence of GVHD (48.5%), 74 experienced a competing risk (18.5%), and 132 patients were censored (33%). The median length of follow-up was approximately six months while the maximum follow-up was approximately 14 years. Unlike our simulated data in Chapter 4, patients entered the population at different times and thus patients could be censored at different lengths of follow-up.

5.2.3 Confidence Interval Estimation of the Cumulative Incidence

For illustration we graph estimates of the cumulative incidence for all times annually up to the last occurrence of GVHD which happens approximately after 7 years follow-up. Because a clinician may be particularly interested in GVHD cumulative incidence rates for GVHD within the first 5 years, Table 5.1 presents these values and compares the associated confidence intervals formed by the estimators described in Chapter 4 using the log-log method.

Figure 5.2: Cumulative Incidence of Graft Versus Host Disease Accounting for Competing Events due to Death or Relapse and associated 95% log-log Counting Process Confidence Intervals at Years 1 to 5

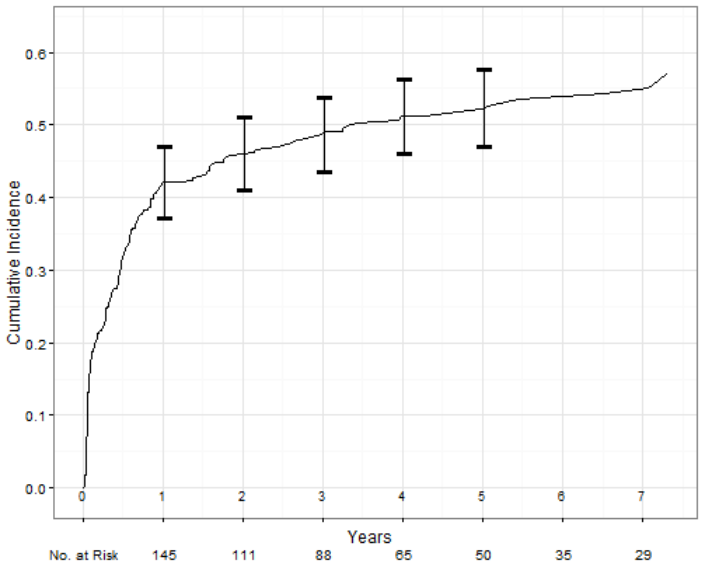


Figure 5.2 shows the cumulative incidence function for GVHD, estimated using Formula 2.4, with the associated 95% log-log Counting Process confidence interval. Following Pocock (2007), we include the number of patients still at risk in our graph. We chose to provide the log-log Counting Process confidence interval because it performed well under similar scenarios in our simulation study, though, as we see from Table 5.1, the choice of the Counting Process estimator does not make much difference. We see the cumulative incidence estimate sharply rises over the first year, until it reaches above 0.4, where over half the population is no longer at risk, then gradually increases from then until year 7 from approximately 0.4 to 0.6.

The 95% confidence interval estimates obtained at 1 to 5 years of follow up are given in Table 5.1. We present the confidence intervals calculated using the log-log confidence interval method because it performed better than the linear confidence interval method in our simulation study, although applied to the present data, their difference did not exceed the second decimal place. Only Bootstrap 2 is slightly wider than other intervals, it being wider by at most 4%. Other intervals preform nearly always identically.

Table 5.1: Comparison of 95% log-log Confidence Interval Methods for Cumulative Incidence at 1 to 5 Years

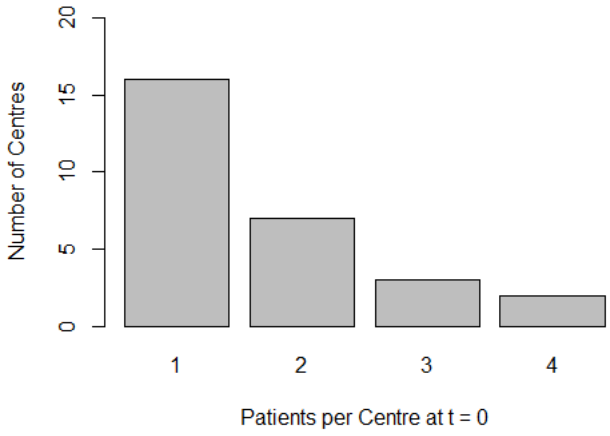
t	No. at Risk	$F(t)$	Multi	CP	Lin	Jack	Boot1	Boot2
1	145	0.42	[0.37, 0.47]	[0.37, 0.47]	[0.36, 0.47]	[0.36, 0.47]	[0.36, 0.47]	[0.35, 0.49]
2	111	0.46	[0.41, 0.51]	[0.41, 0.51]	[0.40, 0.51]	[0.40, 0.51]	[0.40, 0.52]	[0.39, 0.52]
3	88	0.49	[0.44, 0.54]	[0.43, 0.54]	[0.43, 0.54]	[0.43, 0.54]	[0.43, 0.54]	[0.42, 0.55]
4	65	0.51	[0.46, 0.56]	[0.46, 0.56]	[0.46, 0.56]	[0.46, 0.56]	[0.46, 0.56]	[0.44, 0.58]
5	50	0.52	[0.47, 0.57]	[0.47, 0.57]	[0.47, 0.57]	[0.47, 0.57]	[0.46, 0.58]	[0.46, 0.59]

5.3 Hormonal Therapy for Prostate Cancer Dataset

5.3.1 Background

The second dataset we consider is from a randomized trial conducted by the Eastern Cooperative Oncology Group of the United States testing the effectiveness of early hormonal therapy for the treatment of prostate cancer (Messing et al., 1999). It was conducted to determine if early hormonal therapy prolonged the survival of men with prostate cancer. Earlier intervention had not been thought to do so. The study participants were 98 men recruited between 1988 and 1993, having undergone prostatectomy and diagnosed as free of prostate cancer but as having residual cancer cells due to nodal metastases. They were randomized to either immediately receive hormonal therapy or to only receive therapy after the disease progressed sufficiently. Patients were followed from the date of randomization to the date of either death from prostate

Figure 5.3: Number of Patients per Centre in Immediate Arm of ECOG Dataset



cancer, death due to another reason, or censoring. For purposes of illustration, we will estimate the cumulative incidence of death due to prostate cancer for the immediate treatment group and associated 95% confidence intervals.

5.3.2 Descriptive Summary of Data

The 98 patients were recruited from 43 centers. 32% died due to prostate cancer, 14% died from another cause, and 53% of patients were censored. The minimum follow-up time was 1.3 years, the median follow up time was 10.6 years, and the maximum follow-up time was 14.5 years.

Only 47 patients from 28 centers were recruited from the immediate treatment arm. Other summary statistics for the immediate treatment arm are similar, with the minimum follow up time being 2.1 years, median follow-up time being 11.2 years, and maximum follow-up time 14.5 years. Figure 5.3 is a histogram of the number patients recruited from each hospital for the immediate treatment arm. In the immediate treatment arm, 34% of patients were the only one from their centre. The mean number of patients per center was approximately 1.7, and the median was 1.

Table 5.2: Comparison 95% log-log Confidence Interval Methods for Cumulative Incidence of Death due to Prostate Cancer at 10 Years

t	No. at Risk	$F(t)$	Multi	CP	Lin	Jack	Boot1	Boot2
10	34	0.13	[0.05, 0.24]	[0.05, 0.24]	[0.05, 0.26]	[0.05, 0.26]	[0.04, 0.26]	[0.04, 0.27]

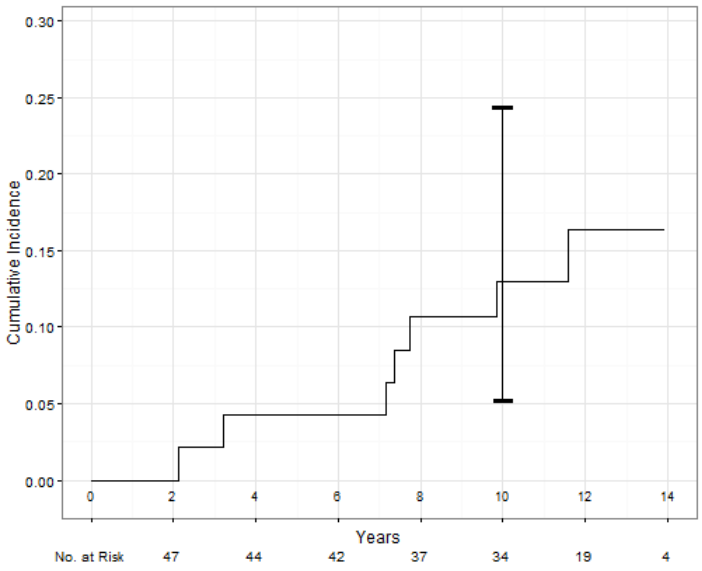
5.3.3 Confidence Interval Estimation of the Cumulative Incidence

For illustration we graph estimates the cumulative incidence for all times up to 14 years. Because a clinician may be interested in the cumulative incidence at 10 years for death due to prostate cancer with immediate treatment, we will present a table with this rate and the associated confidence intervals formed by the estimators we studied in Chapter 4.

Figure 5.4 shows the cumulative incidence of death due to prostate cancer in the immediate treatment arm. As before, the cumulative incidence is estimated using Formula 2.4, with the associated 95% log-log Counting Process confidence interval. We also follow Pocock (2007), and include the number patients still at risk in our graph. We see the cumulative incidence estimate remains much lower than in our previous example, with the maximum value being under 0.2.

To compare the performance of the different confidence interval estimators, the 95% confidence interval estimates obtained at 10 years of follow up are given in Table 5.2. As in the previous example, we present the confidence intervals calculated using the log-log method, which is advantageous here because the cumulative incidence estimates are near 0. Indeed, using the linear confidence interval method would require truncating lower range of the Bootstrap 2 at 0. The Linearized, Jackknife, and Bootstrap 1 methods are wider than the Multinomial and Counting Process methods by 2%, and narrower than Bootstrap 2 by 1%.

Figure 5.4: Cumulative Incidence of Death due to Prostate Cancer in the Immediate Treatment Arm Accounting for Other Causes of Death and Associated 95% log-log Counting Processes Confidence Interval at Year 10



5.4 Summary

The application of the methods of confidence interval estimation to the two datasets here shows a much lower difference in the application these methods than the results of the simulation study would suggest.

Of course, this is not to be unexpected, because the examples do not cover as wide a range of scenarios as the simulation study. Furthermore, as both datasets were used for illustrative purposes only, and did not arise from a cluster randomized trial, the data is likely much less affected by clustering than data coming from an arm in a cluster randomized trial. This may also account for the similarity in performance between the different estimators.

In the analysis of the first EBMT dataset, because the data is a random sample of a much larger original dataset, we have some justification to think our analysis on this sample will be similar to analysis done on the full dataset. However, an important difference between this subset and the original data must be kept in mind: in the original dataset, centers with only one

patient were excluded, but in the random sample, 65 of the 153 centers have only one patient. Also, in the second ECOG dataset, 34% of patients were the only one from their center. The effect of singleton clusters on the similarity of results of the confidence interval estimators, especially those which account for clustering, will be discussed in the next chapter.

Chapter 6

Discussion

6.1 Introduction

This final chapter outlines the results of this thesis and contextualizes its findings within past and future research. Including this introduction the chapter contains five sections. Section 6.2 compares the results of our simulation study to similar simulation studies. The key findings of the thesis are presented in Section 6.3. Section 6.4 looks at the limitations of this work and opportunities for future research, and Section 6.5 summarizes this chapter.

6.2 Comparison to Previous Studies

One of the main goals of this simulation study was to determine how, for cluster randomized trials with competing risks endpoints, clustering affects the performance of traditional confidence interval estimators and if estimators accounting for clustering would provide better coverage. Results from Chapter 4 showed clustering does negatively affect the performance of the traditional Multinational and Counting Process estimators. They also show under heavy clustering estimators accounting for clustering have better coverage.

We will contextualize our results by comparing them with other studies looking at similar problems. One common feature of the papers we compare our study to is, if they looked at clustering, their simulations modeled much higher levels of correlation between subjects from the same cluster than we did. This is often because the setting for their simulations

was sometimes different than cluster randomized trials of patients from different centers. We modeled smaller values because small values of the ICC are observed in cluster randomized trials with patients from different centers (Donner and Klar, 2000).

First we compare our results to two papers comparing variance estimators for the cumulative incidence under competing risks with no clustering. Both included a Multinomial estimator and a Counting Processes based estimator. The first, Braun and Yuan (2007), found the Counting Process estimator given in Equation 2.6 slightly overestimated the empirical variance of the cumulative incidence, and the Multinomial estimator slightly underestimated the empirical variance. In our simulation study, when there was no clustering, that is, when the ICC, $\rho = 0$, our bias results confirm this pattern - the bias of the Counting Process estimator was always above that of the Multinomial estimator, and positive in the majority of cases. The Multinomial estimator was positive in only 5 of the 36 scenarios when ICC = 0. Illjon (2013), comparing the Counting process and Multinomial estimator, made similar conclusions to Braun and Yaun (2007).

Next we compare our results to two simulations looking at clustered survival data, but outside the context of competing risks. Williams (1995) conducted a simulation study comparing the Greenwood estimator for the survival function to a Linearized variance estimator for the survival function. This Linearized estimator was meant to account for the clustering in the survival data. Motivated partly by a repeated measurement study, Williams considers ICCs of 0.5, 0.3 and 0.1. These were much larger than what were examined in our study, where when clustering existed, it was generated using an ICC of either 0.01 and 0.05. He presented results for scenarios where cluster sizes were 5 and 15 and the number of clusters was 50, finding the Linearized estimator had significantly better coverage than Greenwood's estimator. The scenario that most closely matched his in our study on the parameters of sample size, number of clusters, and ICC was when when the sample size was 400, the number of clusters was 40, and ICC = 0.05. In these scenarios, like Williams, our Linearized estimator always performed better than the Multinomial estimator (which is the extension of Greenwood's formula to the

competing risk setting). However, in scenarios with a smaller number of clusters or lower clustering, no estimator was definitively superior .

The second paper looking at clustered survival data we compare our results to is a simulation by Xiao and Abrahamowicz (2010). They performed a simulation applying the two bootstrap methods we considered to study the coverage of confidence intervals for covariate effect estimates in the Cox model. They only considered scenarios with a sample size of 2000 with 50 clusters each of size 40. Like us, they found the two-step bootstrap performed very poorly when the ICC = 0 (The one scenario they considered with ICC = 0 has its coverage at above 99%). All the other scenarios they considered had an ICC above 0.07. Our results were similar to theirs for Bootstrap 1 in the scenarios when the total sample size was 400, ICC = 0.05, and the number of clusters was 40. When the number of clusters was 10 or lower, in our simulation study, Bootstrap 1 generally had poor coverage. Bootstrap 2 however, did not perform as well for us as it did for them in these scenarios. This may be attributable to the smaller sample size we considered. It is also possible that the poor performance of our Bootstrap 2 estimator may be due to a low number of bootstrap replications. We used 200 replications, but Booth and Sarkar (1998) suggest a minimum of 800 replications. Although, Xiao and Abrahamowicz (2010) used 100 and 500 and did not notice a change in the accuracy of their estimators.

Caution should be taken in comparing the results of the previous two studies to ours because, not only were the methods of data generation different, but the settings did not include competing risks.

Finally we compare our results with Chen et al. (2008). They obtained a robust variance estimator for the cumulative incidence function that accounted for within-cluster correlation and performed a simulation study assessing the bias of this variance estimator. Their setting was exactly ours: competing risks with clustering. Their robust estimator was similar to the Linearized estimator we considered. Like them, we found that the bias performance of the Linearized estimator was better than any other method. However, because they did not study

confidence intervals based on this estimator, and the good bias performance of the Linearized estimator did not carry over to good performance of confidence interval coverage in many of the scenarios we considered, we cannot comment of the confidence interval performance of their estimator compared to ours.

6.3 Key Findings

Our simulation study confirmed the results of Braun and Yuan (2007), that when no clustering exists, the Counting Processes based estimator has lower bias than the Multinomial based estimator. We also found confidence intervals based on the Counting Process estimator generally had better coverage than those based on the Multinomial estimator.

When the clustering effect was small (i.e. ICC = 0.01), we found a disadvantage to using any of estimators accounting for clustering we studied, compared to the Counting Process estimator. Coverage of these was poorer especially in scenarios with a smaller number of clusters, but the Linearized, Jackknife, and Bootstrap 1 had comparable performance in scenarios with above 20 clusters.

This may be because, as noted by Donner and Klar (2000), there is an inverse relationship between the sample size from a cluster and the degree of intracluster correlation. If only a small sample of cluster members are included in studies randomizing large clusters then the degree of variance inflation may be quite small. Thus the benefit of having an estimator accounting for clustering may not be noticed when the size of the sample from a cluster is small, especially when the number of clusters has not reached a sufficient number to perform as well as they would asymptotically.

This is consistent with our results in cases where ICC = 0.05, and the largest sizes of patients from a cluster (80 patients), the Bootstrap 2 estimator preforms best, while the Counting Processes estimator performs poorly. And in cases where ICC = 0.05 and there were 40 clusters, the Linearized, Jackknife, and Bootstrap 1 estimators matched or outperformed the Counting Process estimator.

6.4 Limitations and Future Research

As mentioned in Chapter 3, the focus of this thesis was limited to:

- Administrative censoring, that is, censoring due to not all patients experiencing an event when the cumulative incidence is calculated
- All patients being accrued prior to follow-up beginning
- Fixed hazard rates for both competing events over time
- Clusters of equal size

Thus, future simulation studies examining wider types of right censoring may produce more insight into the performance of these estimators in a less idealized scenario. It may also be beneficial if future studies examined different kinds of patient and cluster accrual. In all our scenarios, both clusters and patients within clusters were assumed to be enrolled prior to randomization. However, cluster randomized trials can proceed with recruitment of clusters well beyond the study start date, and allow patient accrual up to the study end.

One question of practical importance that arose during Chapter 5, where we applied our methods to two example datasets, was that some cluster randomized trials may have clusters with only one patient, and a wide range of cluster sizes. A future simulation study examining the effect of differences in cluster sizes on confidence interval estimates would aid future researchers.

Furthermore, our approach of data generation described in Chapter 3 - using a gamma frailty model to simulate clustering and Beyersmann's approach to simulate competing risks - is one of many possible options. Burton (2006) notes that the method of data generation can affect the simulation results. Researchers would benefit from studies considering different data generation procedures.

A final technical question raised by the simulation which would be a subject for future research is the reason for the similarity between the performance of Linearized and Jackknife

estimators.

6.5 Summary

This thesis examined simulated data modeling cluster randomized trials with competing risks endpoints. It found heavy clustering (ICC = 0.05) can affect the coverage of traditional confidence interval estimators not accounting for clustering. Under small amounts of clustering (ICC = 0.01), especially with a sample size of 100 patients, large numbers of clusters, and few patients per cluster, methods of confidence interval estimation for the cumulative incidence not accounting for clustering perform comparably to those that do. In cases with there is heavy clustering (ICC = 0.05), methods accounting for clustering can perform better.

Bibliography

[1] Aalen O. Nonparametric inference for a family of counting processes. *Annals of Statistics* 1978; **6**:701–726.

[2] Aalen O, Johansen S. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics* 1978; **5**:141–150.

[3] Andersen PK, Geskus R, Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. *International Journal of Epidemiology* 2012; **41**:861–870.

[4] Bakoyannis G, Touloumi G. Practical methods for competing risks data: A review. *Statistical Methods in Medical Research* 2010; **21**(3):257–272.

[5] Beyersmann J, Latouche A, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. *Statistics in Medicine* 2009; **28**:956–971.

[6] Birnbaum, Z. On the mathematics of competing risks. *Vital and Health Statistics* 1979; **2**(77).

[7] Booth J, Sarkar S. Monte Carlo Approximation of Bootstrap Variances. *The American Statistician* 1998; **52**(4):354-357.

[8] Bradley, J. Robustness? *British Journal of Mathematical and Statistical Psychology* 1978; **31**(2):144-152.

[9] Braun T, Yuan Z. Comparing the small sample performance of several variance estimators under competing risks. *Statistics in Medicine* 2007; **26**:1170–1180.

[10] Burton A, Altman D, Royston P, Holder R. The design of simulation studies in medical statistics. *Statistics in Medicine* 2006; **25**:4279–4292.

[11] Campbell M, Piaggio G, Elbourne D, Altman D; CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. *British Medical Journal* 2012; **4**:345.

[12] Chen B, Kramer J, Greene M, Rosenberg P. Competing risks analysis of correlated failure time data. *Biometrics* 2008; **64**(1):172–179.

[13] Choudhury J. Non-parametric confidence interval estimation for competing risks analysis: application to contraceptive data. *Statistics in Medicine* 2002; **21**:1129–1144.

[14] Davison A, Hinkley D. *Bootstrap Methods and Their Application*. Cambridge University Press, 1997.

[15] Dietz K, Heesterbeek J. Daniel Bernoulli's epidemiological model revisited. *Mathematical Biosciences* 2002; **180**:1–21.

[16] Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. Arnold, London, UK, 2000.

[17] Efron B. Bootstrap methods: another look at the jackknife. *Annals of Statistics* 1979; **7**(1):1-26.

[18] Efron B. Censored data and the bootstrap. *Journal of the American Statistical Association* 1981; **76**(374):312-319

[19] Fine J, Gray R. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 1999; **94**(446):496-509.

[20] Iljon T. *Calculating confidence intervals for the cumulative incidence function while accounting for competing risks: comparing the Kalbfleisch-Prentice Method and the Counting Process Method*. Master's Thesis, 2013.

[21] Jeong J, Jung S. Rank tests for clustered survival data when dependent subunits are randomized. *Statistics in Medicine* 2006; **25**:361–373.

[22] Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York: Wiley, 1980.

[23] Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data, 2nd ed.* New York: Wiley, 2002.

[24] Kalia S. On the estimation of intracluster correlation for time-to-event outcomes in cluster randomized trials. Master's Thesis, 2015.

[25] Kalra L, Irshad S, Hodsoll J, Simpson M, Gulliford M, Smithard D, Patel A, Rebollo-Mesa I. Prophylactic antibiotics after acute stroke for reducing pneumonia in patients with dysphagia (STROKE-INF): a prospective, cluster-randomised, open-label, masked endpoint, controlled clinical trial. *Lancet* 2015; **386**(10006):1835–1844.

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

[27] Katsahian S, Resche-Rigon M, Chevret S, Porcher R. Analysing multicentre competing risks data with a mixed proportional hazards model for the subdistribution. *Statistics in Medicine* 2006; **25**:4267–4278.

[28] Klein J, Andersen P. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* 2005; **61**:223-229.

[29] Klingebiel T and Schlegel PG. GVHD: overview on pathophysiology, incidence, clinical and biological features. *Bone Marrow Transplant* 1998; **21**:S45–S49.

[30] Koller M, Raatz H, Steyerberg E, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? *Statistics in Medicine* 2012; **31**:1089–1097.

[31] Liang K and Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**(1):13-22.

[32] Lin DY. Non-parametric Inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine* 1997; **16**:901-910.

[33] Lipsitz S, Dear K, and Zhao L. Jackknife estimators of variance for parameter estimates from estimating equations with applications to clustered survival data. *Biometrics* 1994; **50**(3):842-846.

[34] Messing E, Manola J, Yao J, Kiernan M, Crawford D, Wilding G, di'SantAgnese PA, Trump D. Immediate hormonal therapy compared with observation after radical prostatectomy and pelvic lymphadenectomy in men with node-positive prostate cancer. *New England Journal of Medicine* 1999; **341**(24):1781-1788

[35] Moeschberger ML, Klein J. Statistical methods for dependent competing risks. *Lifetime Data Analysis* 1995; **1**:195-204.

[36] Pintilie M. *Competing Risks: A Practical Perspective*. Wiley, 2006.

[37] Pocock S, Travison T, Wruck L. Figures in clinical trial reports: current practice and scope for improvement. *Trials* 2007; **8**:36.

[38] Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics* 1978; **34**:541-554.

[39] Quenouille, M. Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society B* 1949; **11**:68-84.

[40] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015; URL http://www.R-project.org/.

[41] Rao J, Colin D. Fitting dose-response models and hypothesis testing in teratological studies. *Statistics in Toxicology* Gordon and Breach: New York, 1991.

[42] Satagopan JM. A note on competing risks in survival data analysis. *British Journal of Cancer* 2004; **91**:1229-1235.

[43] Spiekerman C, Lin D. Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association* 1998; **93**:443.

[44] Tukey J. Bias and confidence in not quite large samples. *The Annals of Mathematical Statistics* 1958; **29**:614–623.

[45] Williams RL. Product-limit survival functions with correlated survival times. *Lifetime Data Analysis* 1995; **1**:171-186.

[46] Williams RL. A note on robust variance estimation for cluster-correlated data. *Biometrics* 2000; **56**:645–646.

[47] Wolkewitz M. Interpreting and comparing risks in the presence of competing events. *British Medical Journal* 2014; **349**:5060.

[48] Woodruff R. A simple method of approximating the variance of a complicated estimate. *Journal of the American Statistical Association* 1971; **66**(334):411-414.

[49] Xiao Y. *Bootstrap-based Inference for Cox's Proportional Hazards Analyses of Clustered Censored Survival Data* Master's Thesis, 2005.

[50] Xiao Y, Abrahamowicz M. Bootstrap-based methods for estimating standard errors in Cox's regression analyses of clustered event times *Statistics in Medicine* 2010; **29**:915-923.

[51] Zhang X. *Generalized Estimating Equations for Clustered Survival Data*. PhD Thesis, 2006.

[52] Zhou B. *Contributions to Competing Risks Regression*. PhD Thesis, 2010.

[53] Zhou B, Fine J, Latouche A, Labopin M. Competing risks regression for clustered data. *Biostatistics* 2012; **13**(3):371–383.

Appendix A

Derivation of Linearized Variance Estimator

The following appendix presents a derivation of a Taylor series estimator for the variance of the cumulative incidence function using the linearized value approach by Woodruff (1971). This derivation is analogous to the derivation the variance for the survival function in Williams (1995). Refer to Section 2.3 for the definition of notations.

Recall, the estimator we use for the cumulative incidence function for the first competing risk is:

$$\hat{F}_1(t_j) = \sum_{l=1}^{j} \frac{d_{1l}}{n_l} \prod_{p=1}^{l-1} (1 - \frac{d_{1p} + d_{2p}}{n_p})$$

Now, it is possible to consider $\hat{F}_1(t_j)$ as a function of the d_{il}'s and n_l's. For every $l = 1...j$ there is an associated d_{1l}, d_{2l}, and n_l. Thus the function has a domain that is a subset of $\mathbb{R}^{3j}$, and can be written as:

$$\hat{F}_1(t_j) : \mathbb{R}^{3j} \longrightarrow \mathbb{R}$$

$$(d_{11}, d_{21}, n_1, ..., d_{1l}, d_{2l}, n_l, ..., d_{1j}, d_{2j}, n_j) \longmapsto \hat{F}_1(t_j)$$

We now take the partial derivatives of this function. Let $q_l = d_{1l}/n_l$. The partial derivative with respect to d_{1l}, for $l = 1...j$ is:

$$\frac{\hat{F}_1(t_j)}{\partial d_{1l}} = \frac{1}{n_l}\hat{S}(t_{l-1}) + \sum_{l<l'\leq t_j} \frac{d_{1l'}}{n_{l'}} \frac{\hat{S}(t_{l'-1})}{q_l}(-\frac{1}{n_l})$$

The partial derivative with respect to d_{2l}, for $l = 1...j$ is:

$$\frac{\hat{F}_1(t_j)}{\partial d_{2l}} = \sum_{l<l'\leq t_j} \frac{d_{1l'}}{n_{l'}} \frac{\hat{S}(t_{l'-1})}{q_l}(-\frac{1}{n_l})$$

The partial derivative with respect to n_l, for $l = 1...j$ is:

$$\frac{\hat{F}_1(t_j)}{\partial n_l} = -(\frac{d_{1l}}{n_l^2})\hat{S}(t_{l-1}) + \sum_{l<l'\leq t_j} \frac{d_{1l'}}{n_{l'}} \frac{\hat{S}(t_{l'-1})}{q_l}(\frac{d_{1l}+d_{2l}}{n_l^2})$$

Thus the linearized value of $\hat{F}_1(t_j)$, with respect to the k^{th} observation in cluster c is given below.

$$z_{ck}[\hat{F}_1(t_j)] = \sum_{l=1}^{j} \frac{\hat{F}_1(t_j)}{\partial d_{1l}} d_{1ck}(t_l) + \frac{\hat{F}_1(t_j)}{\partial d_{2l}} d_{2ck}(t_l) + \frac{\hat{F}_1(t_j)}{\partial n_l} n_{ck}(t_l)$$

And using the Woodruff (1971) linearization approach, the linearized value of $\hat{F}_1(t_j)$, with respect to cluster c becomes:

$$z_c[\hat{F}_1(t_j)] = \sum_{k=1}^{N_c} z_{ck}[\hat{F}_1(t_j)]$$

Then applying the cluster variance estimator discussed in Williams (1995) and Williams (2000), to these linearized cluster values, we obtain the following variance estimator for the cumulative incidence:

$$var(\hat{F}_1(t_j)) = \sum_{c=1}^{C} (C/(C-1))(z_c[\hat{F}_1(t_j)] - \bar{z}[\hat{F}_1(t_j)])^2 \tag{A.1}$$

where

$$\bar{z}[\hat{F}_1(t_j)] = \sum_{c=1}^{C} z_c[\hat{F}_1(t_j)]/C$$

Appendix B

R Code

B.1 Data Generation

```
#gen.data.beyer.admin is a function
#simulating clustered competing risks data using a gamma frailty model
#to generate cluster hazards and Beyersmann's method to assign
#an event to a generated time.


#s = Cluster size
#N = Number of events
#ICC = Intracluster correlation coefficient
#lambda = Vector containing hazards for the two competing risks
#per.cen = Percentage censored
#(Only administrative censoring is considered)


gen.data.beyer.admin <- function(lambda, per.cen, ICC, s, N) {

  events<-s
```

```
C<-N/s   #C is the number of clusters

#Generating frailties for the k clusters
if (ICC == 0)
  {frailties<-rep(1, times = C)
}  else {frailties<-rgamma(C, shape = 1/ICC, scale = ICC)
}

#Hazards for the ith cluster and jth competing risk.
#Rows are clusters, columns are hazards
hazards<-outer(frailties, lambda)

#ccds = name of generated data frame
ccds <-NULL

#Generating ccds
for (h in 1:C) {
  #Generating Failure Times
  gentimes<-rexp(events, rate = hazards[h,1]+hazards[h,2])

  #Assigning a risk to failure times
  risk <-rbinom(events,1, hazards[h,1]/(hazards[h,1]+hazards[h,2]))
  risk <- replace(risk, risk==0, 2)

  #Generating the point of administrative censoring
  len.of.stud<-qexp((1-per.cen), rate = (lambda[1]+lambda[2]))
```

```
    #Censoring failure times to account for administrative censoring
    risk <- replace(risk, len.of.stud  <= gentimes, 0)
    gentimes <- sapply(1:events,
                       function(i)
                         if (risk[i] == 0)
                         {len.of.stud}
                       else gentimes[i])



    gentimes<- cbind(gentimes, risk, rep(h,events))
    ccds <- rbind(ccds,gentimes)
  }
  #Resimulating data if all events are censored
  if (all(risk == 0))
  {gen.data.beyer.admin(lambda, per.cen, ICC, s, N)
   warning("All events are censored, resimulating.")
  }  else {return(ccds)
  }
}
```

B.2 Code for Cumulative Incidence and Variance Estimators

```
#Loading the "parallel" package.
library(parallel)


#This document contains the code for the cumulative incidence function
#and six variance estimators of the cumulative incidence.
```

```
#The input for every function here is a dataframe containing
#clustered competing risk observations
#and a time at which the function is to be calculated.

#The dataframe's first column specifies a time,
#the second a status associated with the time,
#and the third a cluster.



#cif calculates the cumulative incidence.
cif<-function(crd, t) {

  df <- aggregate(crd[,2], list(crd[,1]),
                  function(status) c(sum(status==1), sum(status==2),
                                     sum(status==0)))

  df<-setNames(df, c("times", "status"))

  #unqiue events times
  times <- df$times
  #d1
  d1 <- df$status[,1]
  #d2
  d2 <- df$status[,2]
  #censored
  c  <- df$status[,3]
```

```
#number remaining
n <- rev(cumsum(rev(d1+d2+c)))

df<-cbind(times, d1,d2,n)
df<-subset(df, df[,1]<=t)

#warning if no events occour before the time we wish to calculate
if(nrow(df)==0) {warning("No values below t")}

df2<-subset(df, df[,2]>=1|df[,3]>=1)

#warning if no competing risk or censoring occours
if(nrow(df2)==0) {warning("No competing risk or censoring occours")
                  return(df)}

df<-df2
times<-df[,1]
d1<- df[,2]
d2<- df[,3]
n <- df[,4]
q <- 1-((d1+d2)/n)
#survival function
s <- cumprod(q)
sm1 <- c(1,s[-length(s)])
#cif values
f <- cumsum((d1/n)*sm1)
```

```
  #returning the value at the desired time
  if(t >= tail(times,1)) {
    return( tail(f,1))
  }
  else {
    return(f[which(tail(times[times <= t], 1) == times)])
  }
}

#ci.cp.lin function calculating the Multinomial, Counting Process,
#and Linearzed estimators.
ci.cp.lin<-function(crd, t) {

  df <- aggregate(crd[,2], list(crd[,1]),
                  function(status) c(sum(status==1), sum(status==2),
                                     sum(status==0)))

  df<-setNames(df, c("times", "status"))

  #unqiue events times
  times <- df$times
  #d1
  d1 <- df$status[,1]
  #d2
  d2 <- df$status[,2]
  #censored
  c  <- df$status[,3]
```

```
#number remaining
n <- rev(cumsum(rev(d1+d2+c)))

df<-cbind(times, d1,d2,n)
df<-subset(df, df[,1]<=t)

#warning if no events occour before the time we wish to calculate
if(nrow(df)==0) {warning("No values below t")}

df2<-subset(df, df[,2]>=1|df[,3]>=1)

#warning if no competing risk or censoring occours
if(nrow(df2)==0) {warning("No competing risk or censoring occours")
                  return(df)}
df<-df2
times<-df[,1]
lf<-length(times)
d1<- df[,2]
d2<- df[,3]
d<- d1+d2
n <- df[,4]
q <- 1-((d1+d2)/n)
#survival function
s <- cumprod(q)
sm1 <- c(1,s[-length(s)])

#cumulative incidence
```

```
f <- cumsum((d1/n)*sm1)


ftj<-tail(f,1)


#multivariate variance
mult.var <- (sum((ftj-f[1:lf])^2 * (d1[1:lf]/
                                      (n[1:lf]*(n[1:lf]-d1[1:lf]))))
+ sum(sm1[1:lf]^2 * ((d1[1:lf]*(n[1:lf]-d1[1:lf]))/(n[1:lf]^3)))
- 2 * sum((ftj-f[1:lf])* sm1[1:lf] * (d1[1:lf]/(n[1:lf]^2))))


#counting process variance
cpvar <- (sum((ftj-f[1:lf])^2 * (d[1:lf]/((n[1:lf]-1)
                                          *(n[1:lf]-d[1:lf]))))
        + sum(sm1[1:lf]^2 * ((d1[1:lf]*(n[1:lf]-d1[1:lf]))
                             /(n[1:lf]^2*(n[1:lf]-1))))
        - 2 * sum((ftj-f[1:lf])* sm1[1:lf]
                  *((d1[1:lf]*(n[1:lf]-d1[1:lf]))
                     /(n[1:lf]*(n[1:lf]-d[1:lf])
                       *(n[1:lf]-1)))))


#Derivative of cif with respect to d1
Dd1 <- unlist(sapply(1:lf-1,
                     function(i) {(((1/n[i])*(sm1[i])) +
                                     (-1/(q[i]*n[i]))*
                                     sum(((d1*sm1)/n)[(i+1):lf])}))
#Derivative of cif with respect to d2
Dd2 <- unlist(sapply(1:lf-1,
```

```
                    function(i) {(-1/(q[i]*n[i]))*
                                    sum(((d1*sm1)/n)[(i+1):lf])}))

#Derivative of cif with respect to n
Ddn <- unlist(sapply(1:lf-1,
                    function(i) {((-d1[i]/(n[i]^2))*(sm1[i]) +
                                    ((d[i])/(q[i]*(n[i]^2)))*
                                    sum((((d1*sm1)/n)[(i+1):lf]))}))

fDd1 <- c(Dd1, ((1/n[lf])*(sm1[lf])))

fDd2 <- c(Dd2, 0)

fDn <- c(Ddn, (-d1[lf]/(n[lf]^2))*(sm1[lf]))

zg<-NULL

clusters<-as.integer(levels(factor(crd[,3])))
C<-length(clusters)

for (g in clusters) {

  g.obs<-subset(crd, crd[,3] == g)

  n.g<-sapply(1:lf, function(i) c(sum(g.obs[,1]==times[i]
                                    & g.obs[,2]==1),
                                  sum(g.obs[,1]==times[i]
```

```
                                        & g.obs[,2]==2),
                                    sum(g.obs[,1]>=times[i])))

    n.g<-t(n.g)

    zg<-c(zg, sum(fDd1*n.g[,1], fDd2*n.g[,2], fDn*n.g[,3]))
  }

  zhat <- mean(zg)

  #the linearized variance
  linvar <- (C/(C-1)) * sum((zg-zhat)^2)

  c(ftj, mult.var, cpvar, linvar)
}

#jackknife calculates the Jackknife variance estimator
jackknife <- function(crd,t) {
  #finding clusters
  clusters<-as.integer(levels(factor(crd[,3])))
  #defining number of clusters
  C<-length(clusters)
  #estimating the value of the cif at time t
  cift <- cif(crd, t)

  #the jackknife estimator
  ((C-1)/C)*sum(sapply(1:C,
```

```
                        function(i) {
                          fmc<-subset(crd, crd[,3] != clusters[i])
                          (cif(fmc, t)
                           -cift)^2}
  ))
}

#clust.boot calculates the onesample cluster bootstrap estimator Boot1
clust.boot <- function(crd, t) {
  #finding clusters
  clusters<-as.integer(levels(factor(crd[,3])))
  #defining number of clusters
  C<-length(clusters)

  #this subfunction creates a bootstrap sample and applies the
  #cumulative incidence function to this sample at time t
  bootrep.clustsam<-function(crdf) {
    bootsamp<-NULL
    bootclustsamp<-sample(clusters, C, TRUE)
    for(i in 1:C) {
      samp<-subset(crdf, crdf[,3] == bootclustsamp[i])
      bootsamp<-rbind(bootsamp, samp)
    }
    if (all(bootsamp[,2] == 0))
    {bootrep.clustsam(crdf)
    }  else {cif(bootsamp, t)
    }
```

```
  }

  # Calculate the number of cores
  no_cores <- detectCores() - 1
  # Initiate cluster
  cl <- makeCluster(no_cores)
  clusterExport(cl, c("crd", "C", "clusters", "bootrep.clustsam",
                      "cif", "t", "ac.pt"), envir = environment())

  num_boot_rep <- 200

  #performs bootstrap sample
  bootrepsamp<-parSapply(cl, 1:num_boot_rep, function(x) bootrep.clustsam(crd))

  stopCluster(cl)
  #estimate of variance
  sum((bootrepsamp-mean(bootrepsamp))^2)/(num_boot_rep-1)
}



#Bootstrap Calculates the Two-sample cluster bootstrap estimator Boot2
bootstrap <- function(crd, t) {
  #finding clusters
  clusters<-as.integer(levels(factor(crd[,3])))
  #defining number of clusters
  C<-length(clusters)
```

```
#this subfunction creates a bootstrap sample and applies the
#cumulative incidence function to this sample at time t
bootrep<-function(crdf) {
  bootsamp<-NULL
  bootclustsamp<-sample(clusters, C, TRUE)
  for(i in 1:C) {
    samp<-subset(crdf, crdf[,3] == bootclustsamp[i])
    lsam<-length(samp[,1])
    samp[sample.int(lsam, replace =TRUE),]
    bootsamp<-rbind(bootsamp, samp[sample.int(lsam, replace =TRUE),])
  }
  if (all(bootsamp[,2] == 0))
  {bootrep(crdf)
  }  else {cif(bootsamp, t)
  }
}
#Calculate the number of cores
no_cores <- detectCores() - 1
#Initiate cluster
cl <- makeCluster(no_cores)
clusterExport(cl, c("crd", "C", "clusters", "bootrep", "cif",
                    "t", "ac.pt"), envir = environment())

num_boot_rep <- 200

#performs bootstrap sample
bootrepsamp<-parSapply(cl, 1:num_boot_rep, function(x) bootrep(crd))
```

```
  stopCluster(cl)
  #estimate of variance
  sum((bootrepsamp-mean(bootrepsamp))^2)/(num_boot_rep-1)
}
```

Curriculum Vitae

Name:	Atul Sivaswamy
Post-secondary Education and Degrees:	University of Western Ontario London, ON 2017 M.Sc.
	University of Waterloo Waterloo, ON 2011 B.Math
Related Work Experience:	Teaching Assistant University of Western Ontario 2015
	Research Assistant University of Western Ontario 2014 - 2016