Analysis of Complex Time-to-Event Data

Natalia A. Gouskova

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the
Department of Biostatistics.

Chapel Hill
2014

Approved by:

Jason P. Fine

Jianwen Cai

Stephen R. Cole

Michael G. Hudgens

Feng-Chang Lin

ABSTRACT
Natalia A. Gouskova: Analysis of Complex Time-to-Event Data
(Under the direction of Jason P. Fine)

The number needed to treat (NNT) is a tool often used in clinical settings to illustrate the effect of a treatment. It has been widely adopted in the communication of risks to both clinicians and non-clinicians. We introduced a definition of the NNT for time to event data with competing risks using the cumulative incidence function and suggest non-parametric and semi-parametric inferential methods for right censored time to event data in presence of competing risks.

In HIV-1 clinical trials the interest is often to compare how well treatments suppress the HIV-1 RNA viral load. We propose an endpoint based on the probability of the viral load being suppressed, and suggest that treatment differences be summarized using the mean restricted time a patient spends in the state of viral suppression.

In the standard analysis of competing risks data, proportional hazards models for cause-specific hazards are fit using the same time scale for all causes of failure. We propose estimating cumulative incidence function by fitting regression models for the cause-specific hazard functions using different time scales for each cause. We establish consistency and asymptotic normality of the proposed estimator and assess its performance in simulations. The method is illustrated with stage III colon cancer data obtained from the Surveillance, Epidemiology, and End Results (SEER) program of National Cancer Institute.

In competing risks setup, sometimes it is not possible to obtain information about the event type. We suggest a non-parametric method in which the probabilities of event types are first estimated using local polynomial regression, and then these estimates are used to estimate the cause-specific cumulative hazards and cumulative incidence functions. The method is illustrated using the data on infections in patients from the United States Cystic Fibrosis Foundation Patient Registry.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Literature Review

1.1 Introduction

The intent of this dissertation proposal is to address various problems related to analysis of complex time-to-event data, such as competing risks data and multi-state models. The proposed dissertation will consist of 4 parts. The first part provides a simple extension of the definition of the number needed to treat to competing risks data. The second part defines a new alternative endpoint for clinical trials where the outcome is the time spent by a patient in some transient state, such as viral suppression in HIV clinical trials. The third part proposes an estimator for cumulative incidence functions under the competing risks setup, when a natural choice of time scales is different for different types of event. The final fourth part is dedicated to estimation of cumulative incidence functions under competing risks setup, when the information on the cause of failure is possibly missing.

1.2 Overview of Competing Risks Data

The difference between the classic time to event and competing risks settings is that instead of considering only one type of event, we now recognize that a patient can experience events of several types. One of them is the event of interest, but other events can happen too and they can affect our event of interest. An example of such data can be the tamoxifen trial described by Cummings et al (Cummings et al., 1993). In this trial data, we observe events of two types, a relapse and a death prior to relapse, with a relapse being the event of interest. Death before relapsing precludes a relapse from happening, and thus death prior to relapse is said to be a competing risk with respect to relapse.

It is also possible that events of different types do not exclude one another, but can happen one after another to the same patient. One of an examples of such events could be suppression and then subsequent rebound of the HIV-1 RNA viral load. These are distinct events of two different types and both of them can happen to the same patient. In this particular example, the event of suppression always occurs first, and the event of rebound always follows it. We can view this series of events as various states that the patient is going through, with the initial state being the state prior to the occurrence of the first event (suppression), the next state being the stage after the event of suppression but prior to the event of rebound, and the final state being the state after the event of rebound has occurred. A very clear classification of possible relationships between events was given by Pepe in (Pepe, 1991).

Though the problem of dealing with competing risks data is a relatively old one, it is fair to say that there is no consensus about which function to choose to describe the probability of the event of interest. The two approaches most frequently used for this purpose are the cumulative incidence estimator and the complement of the Kaplan-Meier estimator (Kaplan and Meier, 1958). In our competing risks illustration above, the cumulative incidence function for relapse is $F_{relapse}(t) = Pr(T \leq t \land event\,type = relapse)$ and it is the probability of relapse in the existing conditions with competing risks present (Kalbfleisch and Prentice, 1980). The Kaplan-Meier estimator $\hat{S}_{relapse}(t)$ estimates the function $S_{relapse}(t) = \exp\left\{ -\int_0^t \lambda_{relapse}(u)du \right\}$, where $\lambda_{relapse}(t)$ is the cause-specific hazard function for relapse, formally defined below. Failures due to competing risks are treated by this estimator as censoring. In our example, a patient who died prior to experiencing a relapse would be considered censored.

Using the complement of the Kaplan-Meier estimator $1 - \hat{S}(t)$ still remains arguably the most common way to quantify the probability of the event of interest, even though there is much debate in literature about when it is appropriate. Some discussion on this topic and further references can be found in (Pepe and Mori, 1993) and (Gooley et al., 1999). It could be interpreted as the probability of the event of interest in a hypothetical situation when all the competing risks are removed, assuming that the events are independent (Tsiatis, 1975). That means that removing the mechanism which causes deaths prior to relapse would not affect the mechanism which causes relapse and hence the probability of relapse.

In general, $1 - \hat{S}_{relapse}(t)$ is larger than the probability of the event of interest in the settings when all competing risks operate (Gooley et al., 1999). The reason why the results will be biased is that the Kaplan-Meier estimator treats patients who failed from competing causes the same way it treats censored, even though patients who were censored and those who failed from competing risks are very different on one particular respect. Patients who were censored can still experience the event of interest after being censored. In our example, if a patient was censored they still can experience a relapse in future, and we do take this possibility into account when we compute the probability of relapse. Patients who failed from a competing risk, however, cannot experience the event of interest any longer. A patient who died prior to relapse will never have a relapse. When we treat them the same way as we treat those who were censored and assume that they too have some non-zero probability to experience a relapse in future, we end up overestimating the probability of relapse when a competing risk of death prior to relapse is present. A formal yet very understandable mathematical explanation of this fact is given by Gooley in (Gooley et al., 1999).

Let's introduce some notation and formally describe the competing risks setup. Let ϵ

denote the type of event, $\epsilon \in \{1, ..., J\}$, T the time of event, and C – the censoring time. Also, let $\mathbf{Z}$ be a $p \times 1$ vector covariates, possibly time-dependent. The distribution of time to the event of interest may depend on the covariates, so we may be interested in being able to adjust for covariate values during estimation.

Suppose we have N patients, indexed by $i = 1, \ldots, N$. The data we observe for the i-th patient is $(X_i, \delta_i, \epsilon_i, \mathbf{Z_i})$, where $X_i = min(T_i, C_i)$, $\delta_i = I(X_i = T_i)$, and ϵ_i – the event type. The random variable ϵ_i is not observed if $\delta_i = 0$. We assume that $(X_i, \delta_i, \epsilon_i, \mathbf{Z_i})$,, are independent and identically distributed, and that the censoring mechanism is independent of the mechanisms that cause events, conditionally on covariates values in $\mathbf{Z}$.

We define the cumulative incidence function, or subdistribution, for an event of type j, $j = 1, ..., J$, as $F_j(t; \mathbf{Z}) = Pr(T \leq t, \epsilon = j \mid \mathbf{Z})$, which is the probability an event of type j occurs by the time t.

There are two types of hazard functions which we can use to describe the competing risks data. One is the cause-specific hazards, defined as:

$$\tilde{\lambda}_j(t; \mathbf{Z}) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr\{t \leq T \leq t + \Delta t, \epsilon = j \mid T \geq t, \mathbf{Z}\}$$

$$(1.1)$$

This function can be viewed as an instantaneous rate of event type j given a patient being at risk, with 'being at risk' defined as not having experienced any event by the time t.

A subdistributional hazard function $\lambda_j(t; \mathbf{Z})$ for the event of type j is

$$\lambda_j(t; \mathbf{Z}) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr\{t \leq T \leq t + \Delta t, \epsilon = j \mid T \geq t \lor (T \leq t \land \epsilon \neq j), \mathbf{Z}\}$$
$$= \frac{\frac{d}{dt} F_j(t; \mathbf{Z})}{1 - F_j(t; \mathbf{Z})}$$

Note that for the purpose of this second definition 'being at risk' defined as not having experienced an event of type j. Thus a patient who has experienced a competing event of type 2 still is considered to be at risk for an event of type 1.

The cumulative incidence function can be expressed via cause-specific hazard functions as:

$$F_j(t; \mathbf{z_0}) = \int_0^t S(u) \tilde{\lambda}_j(t; \mathbf{Z})) du\}.$$

where $S(u)$ is the overall survival function. Alternatively, it can be expressed via subdistri-

3

butional hazard function as

$$F_j(t; \mathbf{z_0}) = 1 - \exp\{-\int\limits_0^t \lambda_j(u; \mathbf{z_0})du\}.$$

1.2.1 Non-Parametric Methods

Let $t_1 \leq t_2 \leq \ldots \leq t_n$ be the ordered observed times when an event of any type occurred. Let d_{ij} be the number of events of type j that occurred at time t_i. Let y_i be the number of patients who are still at risk just prior to time t_i, that is those who haven't experienced any event and haven't been censored yet. Let $S(t) = Pr(T \geq t)$ denote the overall survival function, which is the probability of surviving to the time t without experiencing an event of any type. The non-parametric estimate of the cumulative incidence function is given by the Aalen-Johansen estimator (Aalen and Johansen, 1978):

$$\hat{F}_1(t) = \sum_{t_i \leq t} \hat{S}(t_i-)\frac{d_{i1}}{y_i}$$

where $\hat{S}(\cdot)$ is the Kaplan-Meier estimator of the overall survival function.

There exist a number of estimators for the variance of $\hat{F}_1(t)$. A comparison of their empirical performance in small samples is given in (Braun and Yuan, 2007). The following variance estimator is obtained by simplifying the Gray's estimator (Gray, 1988) for the case of competing events of two types:

$$\widehat{Var}(\hat{F}_1(t)) = \sum_{t_i \leq t} \{\hat{F}_1(t) - \hat{F}_1(t_i) - \hat{S}(t_i)\}^2 \frac{d_{i1}}{y_i^2}$$
$$+ \sum_{t_i \leq t} \{\hat{F}_1(t) - \hat{F}_1(t_i)\}^2 \frac{d_{i2}}{y_i^2}.$$

1.2.2 Semi-Parametric Methods

When it is desirable to look at the effect of covariates on the treatment effect, a semi-parametric approach can be used. One can model either the cause-specific hazard functions, or subdistributional hazards.

Cause-specific hazards can be estimated by fitting the standard Cox model to the time of the event of interest and treating the competing events as censored observations.

Fine and Gray in (Fine and Gray, 1999) developed the proportional subdistributional hazards model which states that

$$\lambda_1(t; \mathbf{Z}) = \lambda_{10}(t) \exp\{\mathbf{Z}^T(t)\beta_0\}$$

where $\lambda_{10}(t)$ is the baseline subdistributional hazard function and β_0 is a $p \times 1$ parameter vector. If the covariate vector $\mathbf{Z}(t) = \mathbf{Z}$ does not depend on time, the subdistributional hazards are proportional for all values of t. In the general case, we allow for time × covariate interactions, and hence for time-dependent covariates. The cumulative incidence $F_1(t; \mathbf{Z})$ can then be expressed as

$$F_1(t; \mathbf{Z}) = 1 - \exp\Big[- \int_0^t \lambda_{10}(s) \exp\{\mathbf{Z}^T(s)\beta_0\}ds\Big],$$

from which one can estimate the parameter vector $\hat{\beta}$ using the modified score function from the partial likelihood for $F_1(t; \mathbf{Z})$ (Fine and Gray, 1999), and the baseline integrated hazard for subdistribution $\hat{\Lambda}_{10}(t)$ by a modification of Breslow's estimator (Breslow, 1972).

For a given value of a covariate vector $\mathbf{Z} = \mathbf{z_0}$ we can compute the integrated hazard

$$\hat{\Lambda}_1(t; \mathbf{z_0}) = \int_0^t \exp\{\mathbf{z_0}(s)\hat{\beta}\}d\hat{\Lambda}_{10}(s)$$

and the predicted cumulative incidence function as

$$\hat{F}_1(t; \mathbf{z_0}) = 1 - \exp\{-\hat{\Lambda}_1(t; \mathbf{z_0})\}.$$

1.3 Number Needed to Treat for Time To Event Data with Competing Risks

The number of patients whose treatment, on average, increases favourable patient outcomes by exactly one, is usually referred to as the "number needed to treat," or NNT. Introduced by Laupacis et al. (Laupacis et al., 1988) in 1988 for controlled clinical trials where the patient-level outcome is binary (event or no event, with an event assumed everywhere in further text to be a 'bad' outcome, such as death or a myocardial infarction), the NNT has become popular as a convenient statistic to communicate treatment benefit. If the groups in a two-group parallel-arm trial are labeled as Treatment (*Trt*) and Control (*Ctl*), with respective probabilities π^{Trt}, π^{Ctl} for an such an event, then NNT is the group size with one fewer

expected event under Treatment than Control, defined formally by $NNT(\pi^{Ctl} - \pi^{Trt}) = 1$ or, equivalently,

$$NNT = \frac{1}{\pi^{Ctl} - \pi^{Trt}}.$$

The denominator is known to epidemiologists as the risk difference (RD), absolute risk (AR), or absolute risk reduction (ARR) (Last, 1988; Cook and Sackett, 1995); succinctly, we write $NNT = ARR^{-1}$. The more efficient a treatment, the greater the ARR and the smaller the NNT. This simplicity of interpretation facilitates use of NNT in communicating the practical impact of a treatment effect on a scale more readily accessible to health care providers and patients than either absolute or relative risk. The clinical trial-based definition of NNT may be extended to communicate the potential benefits of a preventive behavior or public policy change, by replacing control and treatment groups in the clinical trial setting with higher and lower levels of a modifiable exposure in a field trial or observational study. Although caution is clearly required when interpreting NNT from non-experimental data, in such contexts ARR^{-1} may be interpreted as the "number needed to prevent" if the exposure is assumed to be causal.

To estimate the NNT, one needs respective estimates $\hat{\pi}^{Trt}, \hat{\pi}^{Ctl}$ of π^{Trt}, π^{Ctl}, from which a point estimate $\widehat{ARR} = (\hat{\pi}^{Ctl} - \hat{\pi}^{Trt})$ of ARR immediately follows, and a $100(1 - \alpha)\%$ confidence interval (ARR_L, ARR_U) may be obtained using a number of methods (Connor and Imrey, 2005). From these we obtain corresponding point estimate $\widehat{NNT} = \widehat{ARR}^{-1}$ and $100(1 - \alpha)\%$ confidence interval $(NNT_L, NNT_U) = (ARR_U^{-1}, ARR_L^{-1})$ for the NNT.

Some technical challenges in reporting and interpreting the NNT and its confidence interval arise when the difference in the event rates between the groups is not statistically significant. In this case the confidence interval for the absolute risk reduction contains zero and its lower limit is negative. Therefore, the confidence interval for the NNT will contain infinity and have a form $(-\infty; -a] \cup [b; \infty)$ for some $a > 0, b > 0$. The interpretation of the negative part of this confidence interval may be the number of patients who need to be treated in order to observe one event more in the treatment group compared to the control group, which is sometimes referred to as "the number needed to harm" (or NNH), which naturally arises when evaluating side effects of therapies (Altman, 1998). McQuay and Moore in (McQuay and Moore, 1997) suggest that in such cases only the point estimate without the confidence interval be reported. However, we will report such confidence intervals, because they can convey information useful for decision-making in clinical settings.

The concepts above may be generalized from Bernoulli outcomes by considering the NNT as a function NNT(t) of observation time t. When events occur with low constant hazards (equivalently, incidence density rates) $\lambda^{Trt}, \lambda^{Ctl}$ relative to the period of observation, the

6

person-time denominated constant $NNT^* = (\lambda^{Ctl} - \lambda^{Trt})^{-1}$ can be used as an underlying and simpler summary of treatment impact than the full NNT function. Under these circumstances, the survival functions are declining exponentials, with their complements π^{Ctl}, π^{Trt} thus respectively well-approximated for sufficiently low t by $t\lambda^{Ctl}, t\lambda^{Trt}$; this yields the simple approximation $NNT(t) \sim NNT^*/t$. The hazards may be estimated in each group as the ratio of total events to total person-time monitored. Confidence intervals for NNT^* may be formed based on the Poisson distribution if patients and time are both assumed homogeneous, and on the negative binomial or other model to take patient heterogeneity into account.

However, when hazard rates vary during the intervals in which patients are observed, the ARR computed from the average incidence densities and the NNT taken as its inverse do not accurately describe the trajectory of treatment or exposure impact over time, and may be seriously misleading. For such time-to-event studies, which thus require survival analysis, Altman and Andersen (Altman and Andersen, 1999) have defined the NNT in the clinical trial context as a function of time:

$$NNT(t) = \frac{1}{S^{Trt}(t) - S^{Ctl}(t)}$$

where $S^{Trt}(t)$, $S^{Ctl}(t)$ are the probabilities to survive without an event up to time t in the treatment and control groups respectively. Often only a few given points in time would be of practical interest. Altman and Andersen (Altman and Andersen, 1999) suggested methods to obtain estimates and confidence intervals for the NNT(t) from published results of clinical trials with time to event data when either non-parametric or semi-parametric analysis had been performed. Obviously, if raw data are available, NNT(t) can be estimated easily using the same methods.

There is little guidance in the literature on defining and computing NNT from time to event data with competing risks, even though such a definition would be very useful. Koller et al. (Koller et al., 2012) found problematic inattention to competing risk issues in 35 (70%) of 50 follow-up studies in patient populations susceptible to multiple risks reported in high-impact medical journals from late 2007 to late 2010. The tamoxifen clinical trial E1178 described in detail in (Cummings et al., 1993) and (Fine and Gray, 1999) exemplifies studies for which a definition of NNT in a competing risk setting would be useful. The goal of the trial was to establish efficacy of tamoxifen in reducing the probability of relapse when administered as a post-operative treatment to breast cancer patients. Besides relapse and relapse-related deaths, patients in the trial experienced deaths without relapse, such

as death from myocardial infarction, complications of diabetes, other types of cancer etc., which makes tools developed for a competing risks setup most appropriate for analysis of the trial data. However, if one wished to quantify the benefits of tamoxifen demonstrated by this trial using NNT(t), one would face the problem that the NNT(t) defined for standard time to event settings may not be appropriate and that no definition of NNT(t) for data with competing risks exists. Chapter 2 of this dissertation suggests such a definition.

1.4 Endpoints for HIV Trials Measuring Viral Suppression and Multi-State Models

A well-defined outcome is fundamental to the analysis of time to event data. However, in some settings a clear definition of the event of interest is a challenge. An example of such settings are prospective studies, including recent clinical trials evaluating the difference between treatments or exposures which are intended to suppress the level of HIV-1 RNA viral load (henceforth viral load) in people infected with HIV.

Infection with HIV is monitored by the number of copies of viral load present in circulating plasma (Mellors et al., 1996). The level and change in viral load is an important indicator of HIV disease progression. HIV research relies heavily on viral load levels for evaluating the comparative efficacy and effectiveness of competing therapy regimens, and estimating the prognosis of HIV-infected individuals (Egger et al., 2002; Cole et al., 2007; Riddler et al., 2008). The relative performance of HIV treatments depends on the combination of how quickly and to what extent the treatment suppresses viral load, and how well a treatment maintains a suppressed viral load.

There is a history, when comparing HIV treatments, of constructing a single composite event which combines all of the above aspects. For example, Gulick (Gulick et al., 2004; Gulick et al., 2006) defined virologic failure as the first of two consecutive viral load assessments greater than 200 copies/ml at or after week 16 from randomization. Using this definition, patients who never suppress virus are defined as failures at 16 weeks. Moreover, the dynamics of suppression and rebound before 16 weeks are hidden. A more recent example of such a definition can be found in the paper by Riddler et al (Riddler et al., 2008) describing the ACTG A5142 trial. The primary endpoint of the trial was chosen to be a virologic failure defined as "rebound or lack of suppression by m weeks". While such endpoints facilitate the application of standard methodology for right censored time to event data in an intent to treat analysis, there are practical concerns which arise from the event definition. First, the interpretation of the composite event is complicated and may be difficult to communicate to clinicians. Second, when including "lack of suppression by m weeks" in the definition, we

arbitrarily (by the choice of m) redefine the time of event, which can have a notable impact on results, as evidenced in the simulation studies further in this document, in Chapter 3, Section 3.3.

To avoid the above mentioned problems, we suggest the use of a different endpoint and different analysis methods based on non-parametric methods for multi-state models developed by Pepe (Pepe, 1991). The methods have been developed for the situation when each patient can experience not one but several distinct events of types $j, j = 1, ..., J$. The event of each type can be experienced only once. The data from each patient has a form of (T_{ji}), where T_{ji} is the time of event j for the $i-th$ patient. For each of the event types we can define estimators of individual cumulative incidence functions $\hat{F}_j(t)$. For a smooth function g of several such estimators, Pepe derives the asymptotic properties of the resulting quantity and proposes a two-sample weighted test statistic for it. Using our illustration from the Section 1.2, if the event of type 1 is the suppression of the viral load, and the event of type 2 is the rebound of the viral load, with $F_1(t), F_2(t)$ being their cumulative incidence functions, then we can define the probability that a patient's viral load is suppressed at a given time t as $Pr_{suppr}(t) = F_1(t) - F_2(t) = g(F_1(t), F_2(t))$, with function g being $g(a, b) = a - b$. We can obtain the standard error $\widehat{SE}(\widehat{Pr}_{suppr}(t))$, and, further, we can construct a test statistic, if we have two groups of patients and wish to compare the probability of a patient's viral load being suppressed between these two groups.

1.5   On the Choice of Time Scales in Competing Risks Predictions

It is well recognized that in the presence of competing risks, standard survival analysis methods for estimation of cause-specific failure probabilities may not be valid, with the Kaplan-Meier estimator and predicted failure probabilities from the proportional hazards regression model corresponding instead to pseudo-survival functions derived from the cause-specific hazard functions (Prentice et al., 1978). To estimate the cumulative incidence of a particular event type requires synthesizing the cause-specific hazard for the event of interest with those for the competing events. Non-parametric and semi-parametric approaches to estimation via the cause specific hazards have been studied (Aalen and Johansen, 1978; Lin, 1997; Cheng et al., 1998; Shen and Cheng, 1999; Scheike and Zhang, 2003), with a definitive treatment of the theoretical issues provided by martingale arguments (Andersen et al., 1993).

Predicted cumulative incidence functions for event of type j can be computed as described

in Cheng et al (Cheng et al., 1998) :

$$\hat{F}_j(t|z_0)) = \int_0^t \hat{S}(u|z_0) d\hat{\Lambda}_j(u|z_0)$$

where

$$\hat{S}(u|z_0) = exp\{-\sum_{j=1}^J \hat{\Lambda}_j(u|z_0)\},$$

$$\hat{\Lambda}_j(u|z_0) = \hat{\Lambda}_{0j}(u) exp(\hat{\beta}_j z_0),$$

$$\hat{\Lambda}_{0j}(u) = \sum_{i=1}^n \delta_{ji} I(X_i \le u) \{ \sum_{k=1}^n I(X_i \le X_k) exp(\beta_j Z_k) \}^{-1},$$

where $\delta_{ji} = \delta_i I(\epsilon_i = j)$, and $\hat{\beta}_j$ are regression parameter estimates from the Cox model.

In the regression analyses of the cause-specific hazard functions, it seems natural to model death from disease on the time since diagnosis time scale and death from other causes, which may occur both before and after disease diagnosis, on the age time scale. The choice of a time scale has been discussed in literature for modelling data with only one cause of failure. Korn et al (Korn et al., 1997), using the NHANES Followup Study as an example, argued that for all-cause mortality a proper choice of time scale should be age rather than time on study. They gave two conditions under which using time on study as the time scale would still yield correct or close to correct estimates of regression parameters, even if the true model was proportional on the age scale. The first condition was if the true baseline hazard function was exponential, and the second was that age at enrolment to the study and covariates of interest were independent. The authors mention that when the true time scale is age and the baseline hazard function on the age time scale follows a Gompertz distribution (Gompertz, 1825), then the proportional hazards model holds on both time scales. Thiebaut and Benichou (Thiebaut and Benichou, 2004) confirmed these conclusions in simulations and generated an example when both of the above conditions are violated and a choice of a wrong time scale lead to substantial bias. The topic was further studied via simulations by Pencina et al (Pencina et al., 2007) who also noted the importance of accounting for left truncation when using age as a time scale in a situation when patients were observed only after the time of their enrolment to the study.

1.6 Competing Risks Data with Missing Cause of Failure

As it was described above in Section 1.2, an essential component of competing risks data is the cause of each observed failure, denoted by ϵ_i. However, in practice the cause of failure is often unknown, either at all or at the moment of data analysis. For example, in an ongoing cohort study with follow-up, the fact and date of death may be known from follow-up calls or from obituaries, but the properly adjudicated cause of death for some patients may become available with a significant lag time and thus be missing at the time of data analysis. Alternatively, if an autopsy was not performed, it may be not possible to classify a death as a cancer-related or a non-cancer death. There have been suggested numerous methods to analyse competing risks data with possible missing cause of failure.

In 1982 Dinse (Dinse, 1982) provided a classification of incomplete competing risks data, which along with the standard fully observed and fully censored observations allowed two kinds of partially complete observations, with either observed time of failure and missing type of failure, or with an observed type but censored time of failure. To analyse such data, he suggested a non-parametric maximum likelihood estimator (NPMLE) obtained by an EM algorithm, which reduced to a closed-form estimator in case the data contained no observations with observed failure type and censored failure time. The estimated quantities were the overall survival function and the time-dependent probabilities $\pi_j(t_k)$ of having an failure of a type j, given the fact of the failure in the time interval $[t_k, t_{k+1})$ (the time in his analysis was discrete and thus there were more than one event in each interval). The proposed estimator for $\pi_j(t_k)$ was a proportion of failures of type j among all failures with known type that had occurred during the time interval and did not involve any smoothing which made the estimates "extremely erratic", though, interestingly, smoothing was used for display purposes in figures. The missingnes of the cause of failure was assumed to be independent of the type of failure. In a subsequent paper (Dinse, 1986), Dinse suggested another EM algorithm, for the case when time of event is always observed but the probability of the type of failure being observed depends on the type of failure.

Racine-Poon and Hoel (Racine-Poon and Hoel, 1984) considered non-parametric estimation of net survival functions $S_j(t) = \exp(-\int_0^t \lambda_j(u)du)$ for data with exactly two mutually exclusive causes of failure where the cause of failure was determined only with some degree of certainty. Unlike in the standard competing risks setup, their data consisted of pairs (T_i, P_i) where T_i is the time of failure as usual, and P_i is the probability that the failure of subject i occurred due to cause 1. Such probability was provided by an expert for each subject. The hazard function for cause 1 is estimated as $\hat{\lambda}_1(t) = \hat{\pi}_1(t)\hat{\lambda}(t)$ where $\hat{\lambda}(t)$ is the usual

estimator of the overall hazard function, and $\hat{\pi}_1(T_i)$ was taken to be equal to P_i.

Miyakawa (Miyakawa, 1984) studied a fully parametric exponential model and a non-parametric model, both under the MCAR assumption for the missing type of failure. For the non-parametric case, he suggested yet another version of an EM algorithm. At each estimation step, the probabilities of failure of type 1 were expressed as $\hat{\lambda}_1(t)/[\hat{\lambda}_1(t) + \hat{\lambda}_2(t)]$ where $\hat{\lambda}_j(t), j = 1, 2$, were obtained as a simplest local constant smoothers from the Kaplan-Meyer-type estimates of the net survival functions.

Lo (Lo, 1991) regarded a competing cause of failure as censoring, with the censoring indicator possibly missing, and suggested two NPMLE estimators for the case when the probability of the censoring indicator to be observed was the same for all subjects and constant over time. The first one used all observations, and the contribution to the likelihood from the observations with the missing censoring indicator was weighted by the proportion of uncensored events $\pi(t)$. The estimates $\hat{\pi}(t)$ could be viewed as being smoothed over the tail of the distribution of t. The second estimator introduced a simplest inverse probability weighting scheme, where only the fully observed data was used in the Kaplan-Meyer estimator, but contributions of all complete case observations were weighted by the inverse probability of the censoring indicator being observed (which was assumed constant).

The topic has been extensively explored since. Goetghebeur and Ryan (Goetghebeur and Ryan, 1990) proposed a modification of the logrank test for competing risks with missing cause of failure, which was further developed by Dewanji (Dewanji, 1992). Semi-parametric estimators were developed by Gijbels, Lin, and Ying (Gijbels et al., 1993), Goetghebeur and Ryan in the 1995 paper (Goetghebeur and Ryan, 1995), Gao and Tsiatis (Gao and Tsiatis, 2005). Lu and Tsiatis (Lu and Tsiatis, 2001) suggested using multiple imputations for the missing event type, based on a parametric model for the conditional probability of the event of interest given that an event of any type has occurred. The problem of time-to-event data without competing risks but with possibly missing censoring indicator was further studied by McKeague and Subramanian (McKeague and Subramanian, 1998). Yet another version of the problem was addressed by Craiu and Duchesne (Craiu and Duchesne, 2004) and Craiu and Reiser (Craiu and Reiser, 2006). They considered competing risks data with masked event types, that is for some subjects the event type could be determined only up to some subset of all possible event types.

Most recently, for a somewhat similar problem of recurrent events with missing event type, Lin et al (Lin et al., 2013) have developed a non-parametric estimator which uses local polynomial regression to estimate the conditional probability of the event of interest. This local estimate is then used to weight the contribution to the mean event rate from the

events with missing type. A similar approach can be applied to the competing risks setup for estimating the cause-specific hazards.

Chapter 2. Number Needed to Treat for Time To Event Data with Competing Risks

2.1 Introduction

The intent of this chapter is to propose methods to define the number needed to treat for time to event data in presence of competing risks and to estimate it from the raw data using both non-parametric and semi-parametric approaches. This is done in Section 2.2. The methods are illustrated on the breast cancer data in Section 2.3 with some practical remarks concluding in Section 2.4. The work has been published (Gouskova et al., 2014b).

2.2 Methods

We will use the notation introduced in the Chapter 1 above. For purposes of this chapter it is sufficient to consider events of only two types, 1 and 2, with type 1 the event of primary interest and type 2 the competing risk. If more than one competing risk is present, then we can combine them all into one type of event without loss of generality. The non-parametric and semi-parametric methods discussed below are not sensitive to the grouping of the competing events.

Let us denote the cumulative incidence function for an event of type j for $j = 1, 2$ as $F_j(t; \mathbf{Z}) = Pr(T \leq t, \epsilon = j \mid \mathbf{Z})$. In the sequel, we will let superscripts Trt and Ctl denote the membership in the treatment and control groups respectively.

We define the NNT with respect to the event of type 1 as follows:

$$NNT_1(t; \mathbf{Z}) = \frac{1}{ARR_1(t; \mathbf{Z})}$$

where

$$ARR_1(t; \mathbf{Z}) = F_1^{Ctl}(t; \mathbf{Z}) - F_1^{Trt}(t; \mathbf{Z})$$

is the absolute risk reduction from the subdistribution.

To obtain a point estimate of $NNT_1(t; \mathbf{Z})$ we need estimates of $F_1^{Ctl}(t; \mathbf{Z})$ and $F_1^{Trt}(t; \mathbf{Z})$, which can be obtained by using various methods. We will discuss the details of using non-parametric and semi-parametric methods in the corresponding sections below.

The confidence intervals for the NNT can be obtained by transforming the confidence intervals for the ARR. As we have discussed earlier, when the confidence interval for the ARR includes zero, the confidence interval for the NNT should include infinity and consist of two parts. We will report such confidence intervals, because they can convey informa-

tion useful for decision-making in clinical settings. Estimating the confidence interval for the NNT is therefore a two-stage process. First, we estimate the $ARR_1(t; \mathbf{Z})$, obtain its confidence interval and determine if the ARR is significantly greater than zero. Second, if the ARR is significantly greater than zero, then the confidence interval for the NNT is $[1/ARR_U(t); 1/ARR_L(t)]$, where $0 < 1/ARR_U(t) < 1/ARR_L(t)$. If the ARR is not significantly greater than zero, then we report two confidence intervals, $[1/ARR_U(t), \infty)$ for the NNT and $[-1/ARR_L(t), \infty)$ for the NNH, where $ARR_U(t) > 0$ and $ARR_L(t) < 0$.

Under the non-parametric approach, the cumulative incidence function is estimated separately for the treatment and the control groups. Since we assume that all patients are independent, the estimates of the cumulative incidence functions in the treatment and the control groups are independent and the covariance between them is equal to zero. Hence the variance of the $\widehat{ARR_1}(t)$ is

$$Var(\widehat{ARR_1}(t)) = Var(\hat{F}_1^{Ctl}(t)) + Var(\hat{F}_1^{Trt}(t))$$

The 95% confidence interval for $\widehat{ARR_1}(t)$ is given by $[ARR_L(t); ARR_U(t)]$ where

$$ARR_L(t) = \widehat{ARR_1}(t) - 1.96 \cdot \sqrt{(\widehat{Var}(\widehat{ARR_1}(t)))} \text{ and } ARR_U(t) = \widehat{ARR_1}(t) + 1.96 \cdot \sqrt{\widehat{Var}(\widehat{ARR_1}(t))}.$$

The point estimate and the confidence interval for the NNT can be computed now as described above.

When it is desirable to look at the effect of covariates on the treatment effect, a semi-parametric approach can be used. Any method allowing estimation of the cumulative incidence function given covariate values can be used for the purpose of the definition of the NNT.

For example, we can use the Fine and Gray model (Fine and Gray, 1999). When one of the covariates, say z_{0p}, is the treatment group assignment, 1 for the active treatment and 0 - for control, we can compute the cumulative incidence functions for the treatment and control groups as

$$\hat{F}_1^{Trt}(t; \mathbf{z_0'}) = \hat{F}_1(t; \mathbf{z_0'}, z_{0p} = 1)$$

$$\hat{F}_1^{Ctl}(t; \mathbf{z_0'}) = \hat{F}_1(t; \mathbf{z_0'}, z_{0p} = 0)$$

where $\mathbf{z_0'} = (z_{01}, ..., z_{0,p-1})^T$ is the vector of covariates other than treatment group. Having those, we can compute the point estimate for the $ARR(t; \mathbf{z_0'})$ and $NNT(t; \mathbf{z_0'})$.

Computing the confidence intervals for the $ARR(t; \mathbf{z_0'})$ is more challenging for the semi-

parametric than for the non-parametric approach, because in this case we use all of the observations to estimate of both $F_1^{Trt}(t; \mathbf{z_0'})$ and $F_1^{Ctl}(t; \mathbf{z_0'})$. Hence, the estimates of the cumulative incidence functions $\hat{F}_1^{Trt}(t; \mathbf{z_0'})$ and $\hat{F}_1^{Ctl}(t; \mathbf{z_0'})$ in the treatment and the control groups are no longer independent. Therefore to compute the variance of the $\mathrm{ARR}(t; \mathbf{z_0'})$ we need the covariance between $\hat{F}_1^{Trt}(t; \mathbf{z_0'})$ and $\hat{F}_1^{Ctl}(t; \mathbf{z_0'})$, for which no simple closed form is known. This challenge is the same as for computing the confidence intervals for the number needed to treat for classic time to event data without competing risks, described by Altman and Andersen in (Altman and Andersen, 1999).

A practical solution is to obtain the variance of the $\mathrm{ARR}(t; \mathbf{z_0'})$ via bootstrapping. To do this, at each repetition of the bootstrapping process, we draw two bootstrap samples independently, one from the treatment group and the other from the control group, combine them into a single bootstrap sample, re-fit the model to the combined bootstrap sample and compute the ARR estimate, and after repeating the procedure the desired number of times compute the estimator of the variance of the $\widehat{ARR}(t; \mathbf{z_0'})$ from all of the bootstrap samples.

A more general direct regression model for the cumulative incidence function was suggested by Fine (Fine, 2001). This includes alternatives to the proportional subdistribution hazards model. Klein et al. (Klein et al., 2007) proposed fitting such models using a pseudo-value approach. In particular, if we are interested in fitting such a model at a specific time point t_0 of interest, such as 5 years after the treatment, we let $\hat{F}_1(t_0) = \sum_{t_i \leq t_0} \hat{S}(t_i-)\frac{d_{i1}}{y_i}$ be the non-parametric estimate of the cumulative incidence function from the whole sample as defined in Section 3.1 above, and $\hat{F}_1^{(j)}(t_0)$ be the same estimate computed with the $j-th$ observation removed from the sample. Defining the pseudo-observations for the cumulative incidence function as $\hat{\theta}_j = n\hat{F}_1(t_0) - (n-1)\hat{F}_1^{(j)}(t_0)$, one may estimate the generalized linear model $E(\hat{\theta}_j | \mathbf{Z_j})$ proposed by Fine (Fine, 2001) using a generalized estimating equations approach. As an example, assuming a linear model for the cumulative incidence function with an identity link and employing a working independence covariance matrix, the estimate of the ARR reduces to $\widehat{ARR}(t_0; \mathbf{z_0'}) = -\hat{\beta}_p$. The variance of $\hat{\beta}_p$ and hence confidence intervals for the ARR can be obtained as described in (Klein et al., 2007). The point estimate and the confidence interval for the NNT as a function of time can now be computed as described just prior to Section 3.1.

2.3 Tamoxifen Trial Example

To illustrate our methods, we use the breast cancer dataset from the tamoxifen trial E1178 briefly introduced earlier. In this trial, tamoxifen or placebo was administered daily for 2 years to patients who underwent mastectomy for potentially curable breast cancer within

the 10-week period before entry into the trial. The data used in this example consist of the observations on 167 eligible patients of the trial, 82 of whom were randomized to placebo and 85 to tamoxifen treatment. The median observation period for these patients was 5.06 years (range from 0.14 to 15.95 years). Of the patients in the placebo group, 59 experienced a recurrence of breast cancer and 19 died without relapse from other causes. In the active treatment group, 42 had a relapse and 23 died without recurrence. The data available for each patient also includes age at time of randomization (ranging from 65 to 84, with median age 71), tumor size (from 3 to 170 mm, median 25), and the number of positive nodes (from 1 to 34, median 3). The last relapse occurred in the control group at 13.16 years, and the last non-relapse death - in the tamoxifen group at 15.7 years, as can be seen in the non-parametric plots in Figure 2.1.

We computed the NNT based on these data using non-parametric and semi-parametric approaches. The choice of a semi-parametric model and model diagnostics were discussed in detail in (Fine and Gray, 1999). Here we will only mention briefly that the proportional hazards model $\lambda_1(t; Z) = \lambda_{10}(t) \exp(\beta_0 Z)$ does not fit the data well, and a semi-parametric model allowing the hazard ratio to be quadratic in time, $\lambda_1(t; Z) = \lambda_{10}(t) \exp(\beta_0 Z + \beta_1 Z t + \beta_2 Z t^2)$, is a more appropriate choice.

Figure 2.1 shows different estimates of the probability of relapse and non-relapse death by treatment group: the estimates of cumulative incidence functions obtained by using the three models, non-parametric, semi-parametric with proportional subdistributional hazards, and semi-parametric with the quadratic in time hazard ratio, along with the complements of the Kaplan-Meier estimator for comparison. As one can see, the estimates of the cumulative incidence functions for the non-relapse death are very close in both groups and don't differ much across the three analysis methods. The estimates are slightly higher in the tamoxifen group. The estimated cumulative incidence functions for relapse differ noticeably between the two treatment groups, with higher estimates in the control group.

Figure 2.2 shows the plot of the non-parametric estimates of the absolute risk reduction and the NNT, both accompanied by pointwise 95% confidence regions. The ARR was not significantly different than zero around 6 months and between 5.5 and 7.5 years (as seen on the plot, the lower confidence limit is negative), therefore the pointwise confidence regions for the NNT around 6 months and between 5.5 and 7.5 years consist of two parts and include infinity. We plotted the negative part of confidence regions for the NNT in a separate panel, as the number needed to harm (NNH).

Figure 2.3 contrasts the non-parametric estimate of the NNT function from Figure 2.2 with its counterparts from the semi-parametric proportional subdistributional hazards model
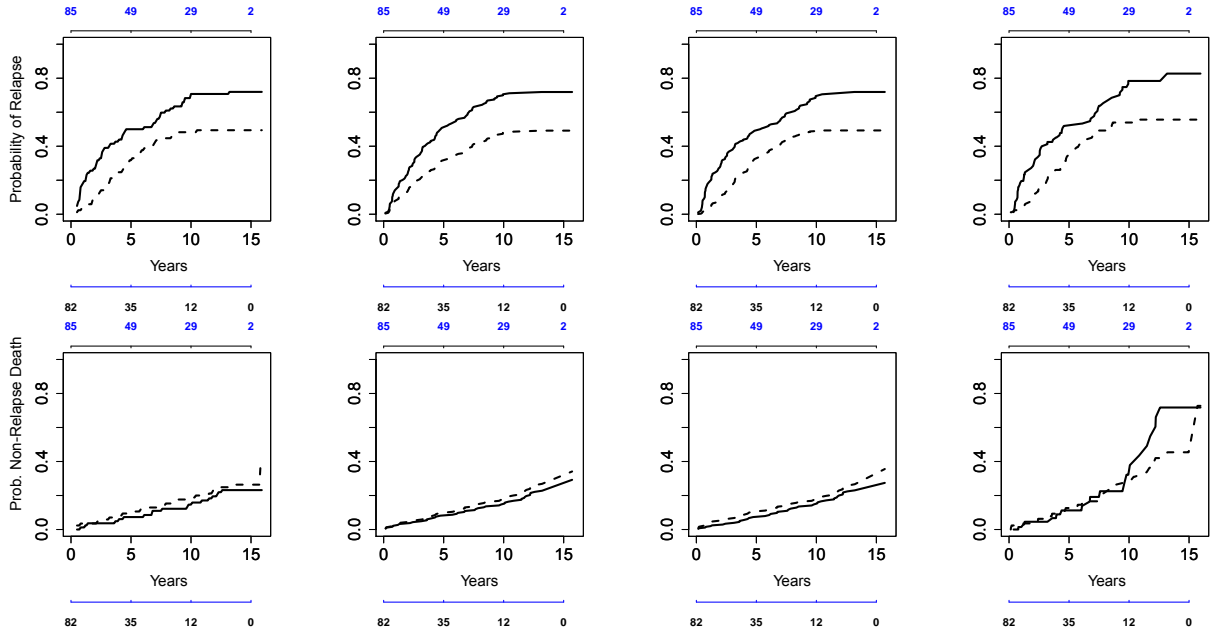
Figure 2.1: Probability of relapse (top row) and non-relapse death (bottom row). Column 1: cumulative incidence, non-parametric model; Column 2: cumulative incidence, semi-parametric proportional hazards; Column 3: cumulative incidence, semi-parametric quadratic in time; Column 4: complement of the Kaplan-Meier estimator. (*Solid line - control group, dashed line - tamoxifen group. Number at risk for tamoxifen group - above each panels, for control group - below each panel.*)
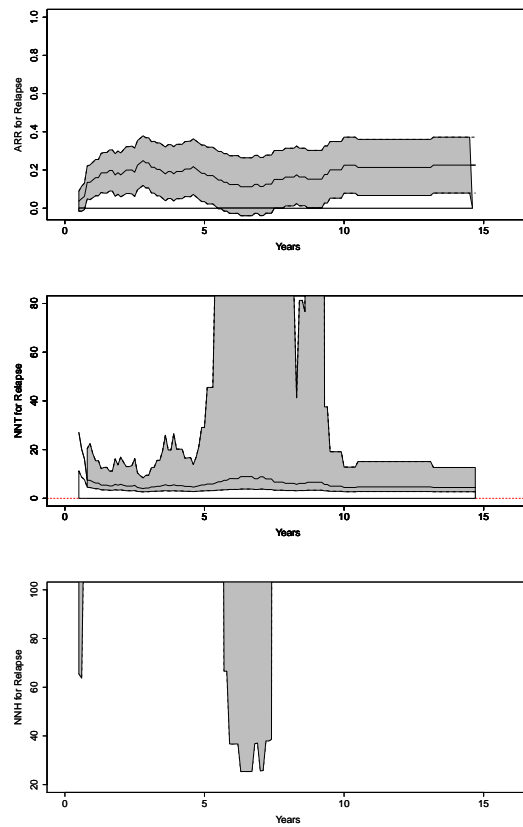
Figure 2.2: Non-parametric estimates of the ARR (top), NNT (middle), and NNH (bottom) with 95% pointwise confidence regions.

in Section 3.2, and the time-quadratic generalization considered in (Fine and Gray, 1999) and noted above. The non-parametric NNT function and its counterpart from the semi-parametric model with quadratic in time hazard ratio behave very similarly, especially at the early times between 0.5 and 4 years and in the tail after 8 years. The NNT function from the semi-parametric proportional hazards subdistributional model without covariate $\times$ time interaction terms differs from these two models very noticeably early, overestimating the NNT prior to 4 years, and does not reflect the non-parametric function's local maximum around 6-7 years. Beyond 10 years, however, all three estimates are very similar. For example, at 12 years the non-parametric estimate is 4.69 with 95% confidence interval [2.78; 15.09], the estimate from proportional hazards model is 4.41 with 95% confidence interval [2.73; 11.45], and the estimate from quadratic in time model is 4.70 with 95% confidence interval [2.76; 15.72]. The more optimistic point estimate and upper confidence interval limit from the proportional hazards model appear to reflect oversmoothing by the proportional hazards model in the 5-10 year period, visible in Figure 2.1.
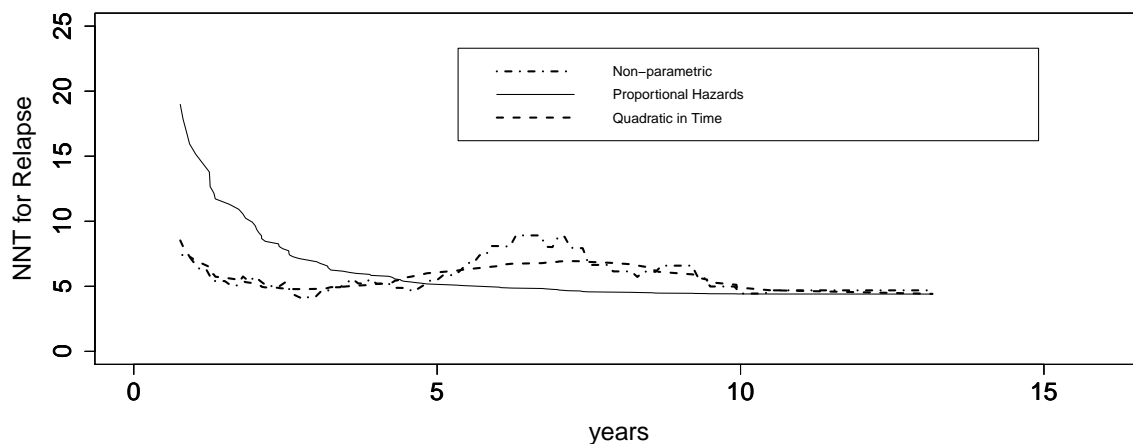
Figure 2.3: Estimates of the NNT for non-parametric, semi-parametric proportional hazards, and semi-parametric quadratic in time models.

In order to use the NNT estimates obtained from the results of this trial, we would first choose a scientifically relevant time point or a few time points that would illustrate the treatment effect best. For tamoxifen, the long-term benefit is perhaps of greatest interest.

Hence, a single time point at 12 years after the beginning of the treatment might be an appropriate choice, after which the rate of relapse is rather low. At 12 years after mastectomy, the patients who received tamoxifen as a post-operative treatment experienced substantially fewer breast cancer recurrences compared to those who were on placebo. The reduction in the probability of relapse corresponds to one event of relapse in approximately every 5 patients treated, with a 95% confidence interval approximately 3 to 16.

It may also be useful to estimate the NNT for a specific subgroup of patients, e.g., based on age and disease severity. To do this, we included age, tumor size, and the logarithm of number of positive nodes, in addition to the treatment group, as time-independent covariates. Age had no significant effect on the distribution of time to relapse. Figure 2.4 compares the predicted NNT computed for an 80-year-old patient with the tumor size 20 mm and 1 positive node with the NNT for an 80-year-old patient with the tumor size 80 mm and 10 positive nodes. The non-parametric NNT computed from the data of all the patients is also given as a reference. The estimated NNT is lower for a patient at a more severe stage of disease, indicating greater treatment effect. This makes sense because one expects the relapse probability for any given time from mastectomy to be lower, in both tamoxifen and control groups, among patients initially treated at earlier vs. later stages of breast cancer. This lower baseline risk at earlier disease stages limits the potential absolute tamoxifen benefit, and the absolute impact of a constant relative benefit will be less pronounced than among patients with diagnosis and surgery later in the disease course. The difference between subgroups is greater in the short run and attenuates by 12 years.

We also performed an analysis similar to that described above, but with death prior to relapse as the event of interest. Other than age, no covariates, including treatment group, had any statistically significant effect on the distribution of the time to non-relapse death. There was no significant difference in the probability of the non-relapse death between the tamoxifen and placebo groups at any time point, regardless of which method was used to compute the estimates. To ascertain the sensitivity of the analysis of these data to treating death without relapse as a competing risk rather than as an independent censoring mechanism, we computed the NNT(t) using the complement of the Kaplan-Meier estimators instead of the cumulative incidence functions as estimates of the probabilities of relapse. For these data, results changed little.

2.4 Practical Remarks

The NNT is a very useful tool for quantifying the efficacy of a treatment in a manner which facilitates communication of risk to non-statisticians. With binary outcomes and time to
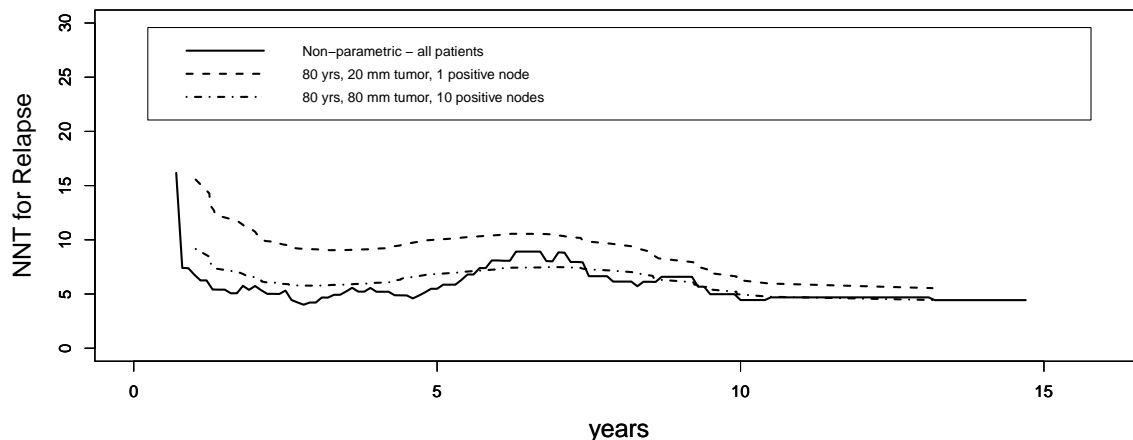
Figure 2.4: Estimates of the NNT for subgroups with specific values of covariates.

event data without competing risks, there is generally a single event of interest and correspondingly a single NNT may be used to summarize the effect of treatment. While it is conceptually straightforward to extend NNT to time to event data with competing risks, multiple NNTs may be needed to capture the overall effect of treatment on the different events. That is, the NNT for a particular event of interest, by itself, may not be a suitable tool for determining if the treatment in question is beneficial, particularly if the effect of treatment on other events is harmful. In the breast cancer example where treatment influenced only the cumulative incidence for the event of interest, the NNT for relapse may be sufficient. However, in other applications, the treatment in question may have a desirable effect on the event of interest but undesirable effects on events of other types. The benefits of the treatment should always be weighted against its negative effects, which requires an examination of the NNTs and, potentially, NNHs for different event types. The NNT for the event of interest, being derived only from the subdistribution of the event of interest, does not provide direct information about the effect of the treatment on the events of other types.

Chapter 3. Combining Times to Suppression and Rebound in HIV Studies

3.1 Introduction

The goal of this chapter is the development of an endpoint which is tailored to the objectives of HIV-1 and similar studies and provides an intuitive summary of treatment differences.

As mentioned earlier in Chapter 1, Section 1.4, the endpoints currently used to assess the suppression of the viral load in HIV studies may suffer from some problems. To avoid these problems, we suggest using a different endpoint and different analysis methods based on multi-state models (Pepe, 1991). These methods explicitly acknowledge the fact that we have two distinct events, viral load suppression and rebound, with corresponding survival functions $S^S(t)$ and $S^R(t)$ respectively. Our proposed endpoint is based on the probability of being in suppression $G(t)$, which is simply $S^R(t) - S^S(t)$. We suggest an intuitive summary of treatment efficacy based on a weighted integral of this difference over a specified time interval of interest, say one year. With equal weights over time, this measure reduces to the restricted mean time suppressed over the time interval. One may tailor the weights to emphasize the timepoints of scientific interest, enabling a rigorous exploration of either early or late suppression dynamics. This endpoint is well-defined and has a clear and simple interpretation which may permit comparisons across trials and populations. The proposed analysis accounts for the fact that a proportion of patients will never suppress their viral load and allows investigators to simultaneously assess differences in both time to viral suppression and time to viral rebound, emphasizing those timepoints relevant to treatment evaluation.

The proposed endpoint is described in Section 3.2. A simulation study assessing performance of the proposed endpoint in comparison to endpoints based on composite events is discussed in Section 3.3. The practical utility of the analysis is illustrated in a reanalysis of ACTG A5142 in Section 3.4. A discussion concludes in Section 3.5. The work has been accepted for publication (Gouskova et al., 2014a).

3.2 Methods

For patient i, let R_i be the treatment regimen assignment at time of randomization, with the focus being an intent to treat analysis of treatment efficacy. The potential time at which patient i has their viral load suppressed is denoted by T_i^S and the potential time at which patient i has their viral load rebound is denoted by T_i^R. Let C_i denote the potential censoring time for patient i, with the binary indicators δ_i^S and δ_i^R equal to 1 when T_i^S and T_i^R are smaller than C_i, respectively, and 0 otherwise. Furthermore, define $X_i^S = min(T_i^S, C_i)$

and $X_i^R = min(T_i^R, C_i)$. In general, $\delta_i^S \geq \delta_i^R$, because the time to rebound of viral load may only occur subsequent to viral load suppression. For patient i, the observed data consists of $(X_i^S, \delta_i^S, X_i^R, \delta_i^R, R_i)$. The main difficulty in conducting a time-to-event intent to treat analysis using this data structure is that there is not an obvious single "time to event" on which to base the analysis.

Suppose for simplicity that there are two treatment groups, $r = 1$ and 2, and let S_r^S and S_r^R denote the survival functions for T_i^S and T_i^R, respectively, in group $r = 1, 2$. The endpoint we propose for viral suppression studies is the probability of being suppressed at time t, $G_r(t) = S_r^R(t) - S_r^S(t), r = 1, 2$. This endpoint is defined without conditioning on information observed post randomization and may be analyzed using intent to treat methods. However, because the event probability is the difference of two survival functions and is not itself a survival function for a single time to event, the Kaplan-Meier estimator and logrank test are not applicable. Inferential methods for multi-state data must be used in the development of non-parametric estimators and tests for treatment differences.

Following (Pepe, 1991), we employ the Kaplan-Meier estimates $\hat{S}_r^S(t), \hat{S}_r^R(t), r = 1, 2$, of survival functions for time to viral suppression and time to viral rebound respectively. Note that time to viral rebound defined as above is measured from randomization. The survival function for time to viral rebound will be the marginal survival function, not conditional on being suppressed. We can estimate the probability for a patient to be in the state of suppression, within each treatment group separately, as

$$\hat{G}_r(t) = \hat{S}_r^R(t) - \hat{S}_r^S(t) \text{ for } r = 1, 2.$$

The variance estimator for $\hat{G}_r(t)$, $r = 1, 2$, is given by

$$\widehat{Var}(\hat{G}_r(t)) = \frac{1}{n_r^2} \sum_{i:R_i=r} [\hat{X}_{G_r}^i(t)]^2$$

where n_r is the number of subjects in group r,

$$\hat{X}_{G_r}^i(t) = n_r \hat{S}_r^S(t) \{ \int_0^t \frac{1}{Y_S(u)} dN_S^i - \int_0^t \frac{Y_S^i(u)}{(Y_S(u))^2} dN_S(u) \} -$$

$$- n_r \hat{S}_r^R(t) \{ \int_0^t \frac{1}{Y_R(u)} dN_R^i - \int_0^t \frac{Y_R^i(u)}{(Y_R(u))^2} dN_R(u) \},$$

$N_S^i(u), N_R^i(u)$ are the counting processes for the events of suppression and rebound re-

spectively for a patient i, $Y_S^i(u)$ and $Y_R^i(u)$ are the at risk processes for suppression and rebound respectively for a patient i, and

$$Y_\epsilon(u) = \sum_{i:R_i=r} Y_\epsilon^i(u) \text{ and } N_\epsilon(u) = \sum_{i:R_i=r} N_\epsilon^i(u) \text{ for } \epsilon \in \{S, R\}.$$

The probability of being in suppression $G_r(t)$ for group $r = 1, 2$ varies over time, similarly to a survival function, albeit not a monotonically decreasing function of t. As with standard time to event analyses, simple summary measures are needed for quantifying differences among treatment regimens. One should recognize that $G_r(t)$ does not have a corresponding hazard function and treatment differences cannot be summarized using hazard ratios, as they might in separate analyses of S_r^R and S_r^S. We suggest summarizing using the weighted restricted mean time a patient from group r will spend in suppression in the time interval $[0, t_0]$, which is $\int_0^{t_0} \hat{W}(u) G_r(u) du$, where $\hat{W}(u)$ is an estimate of some appropriately chosen weight function $W(u)$ discussed below.

The analysis may be tailored to capture the information of greatest importance with a careful choice of the weight function. When $W(u) \equiv 1$, the weighted integral estimates the restricted mean time spent in viral suppression. For those interested in short term outcomes, larger weights may be applied at early time points, and vice versa for long term outcomes. For example, for those interested primarily in long term maintenance, zero weights may be employed at time points before some predetermined cut-off for suppression, eg 24 weeks. On the other hand, for those interested in population health where individuals with circulating virus present a transmission risk, non-zero weights at early time points would be an important consideration.

Following (Pepe, 1991), for the purpose of hypothesis testing one may compute a simple Z type test statistic as the difference of the weighted averages in the two treatment arms. The test statistic is:

$$WG = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \int_0^{t_0} \hat{W}(u)\{\hat{G}_1(u) - \hat{G}_2(u)\} du.$$

Under the null hypothesis, the test statistic is asymptotically normal with zero mean and its asymptotic variance can be estimated by

$$\widehat{Var}(WG) = \frac{n_1 n_2}{n_1 + n_2}(\hat{V}_1 + \hat{V}_2)$$

25

where

$$\widehat{V}_r = \frac{1}{n_r^2} \sum_{j:R_j=r} \left[\int_0^{t_0} \hat{W}(u)\hat{X}_{G_r}^j du \right]^2, r = 1, 2.$$

Wald type confidence intervals for the weighted average time in suppression may be calculated using the asymptotic normality of the estimator $\int_0^{t_0} \hat{W}(u)\hat{G}_r(u)du$ and its variance estimator $\hat{V}_r, r = 1, 2$.

The choice of the weight function may also be directed towards improving the power of the test statistic to detect treatment differences in the probability of suppression over time. As suggested by Pepe and Fleming (1989), one may downweight at time points where $\hat{G}_1 - \hat{G}_2$ is highly variable using the weight function:

$$\hat{W}_{se}(u) = 1/\hat{SE}[\hat{G}_1(u) - \hat{G}_2(u)],$$

where $\widehat{SE}[\hat{G}_1(u) - \hat{G}_2(u)] = \sqrt{\widehat{Var}(\hat{G}_1(t)) + \widehat{Var}(\hat{G}_2(t))}$. This may also be accomplished using some function of the censoring distributions in the two groups (Pepe and Fleming, 1989), with weight:

$$\hat{W}_{cens}(u) = [\hat{S}_1^C(t) \times \hat{S}_2^C(t)]/[p_1\hat{S}_1^C(t) + p_2\hat{S}_2^C(t)],$$

where $\hat{S}_r^C(t)$ is the Kaplan-Meier estimator of the survival function of C_i, $S_r^C(t)$, in group $r = 1, 2$ and p_r is the proportion of patients allocated to group $r = 1, 2$. The unity weight assigns equal weight to all time points, while the second and third weights tend to assign higher weight to earlier time points, where the estimation is typically less variable, potentially resulting in increased power. In applications with focused scientific objectives, the choice of the weight should be driven by those objectives and not by unguided power considerations.

If we have several strata $j = 1, ..., J$ and wish to conduct a stratified analysis, we can compute the above WG statistics separately within each stratum j and then let

$$SWG = \frac{\sum_{j=1}^J \omega_j WG_j}{\sqrt{\sum_{j=1}^J \omega_j^2 \hat{Var}(WG_j)}}$$

where WG_j and $\hat{Var}(WG_j)$ are the test statistic and its estimated variance within stratum $j = 1, ...J$. The scalar ω_j determines the relative weight given to stratum $j = 1, ...J$. Under the null hypothesis of no difference between the groups, the SWG statistic is asympttically

normal $N(0,1)$.

To perform power and sample size calculations for studies using the proposed endpoint, one can use standard formulas for continuous normally distributed outcomes. The standard deviation of the test statistic necessary for such computations can be obtained by re-analysis of previously available similar data or via simulations. For example, for the test statistic based on the unity weight, for a trial with two arms of equal size and assuming equal variances in both arms, we can take the desired effect size Δ to be a clinically relevant difference in average time spent in suppression between the treatment and control arms (for example, 4 weeks, if weeks is the chosen time scale). If the data from an earlier similar trial is available, we can compute the WG_{prior} statistic for the prior trial data and estimate its standard error. Due to the scaling of WG by $\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$, the standard error $\hat{SE}(WG_{prior})$ is an estimate of the true standard deviation of the time spent in suppression. Hence we can use Δ and $\hat{SE}(WG_{prior})$ as the effect size and the standard deviation in the standard sample size formulas. The results of a small simulation study verifying this approach are provided at the end of Section 3.3.

3.3 Simulation Results

We conducted a simulation study to compare performance of the proposed endpoint with the virologic failure endpoint used in the A5142 trial and TLOVR. For each simulated patient we generated a treatment group assignment and then, conditionally on the treatment assignment, times to suppression (T_i^S), rebound (T_i^R), and censoring (C_i). The time was on the weeks scale, and the length of the observation period was chosen to be 80 weeks. We first generated time from randomization to suppression, then time from suppression to rebound, and then computed the time from randomization to rebound as the sum of the two above times.

We employ 3 different simulation scenarios shown in Figure 3.1. In scenario 1 the treatment group was the same as control in terms of suppression and had much later rebound, thus maintaining suppression much longer than the control group. Under scenario 2, the treatment group had faster suppression but also faster rebound. On average, in scenario 2, the treatment group was suppressed longer. In scenario 3, the treatment group suppressed later than in the control group, but maintained suppression longer. Thus, under scenario 3, the treatment group had reduced probability of suppression in the beginning of the observation period which reversed at later times.

We used the Weibull distribution for all time variables in the simulations, due to its flexible shape, with the CDF function $F(t) = 1 - exp\{-(\frac{t}{\beta})^{\alpha}\}$ and the distribution parameters
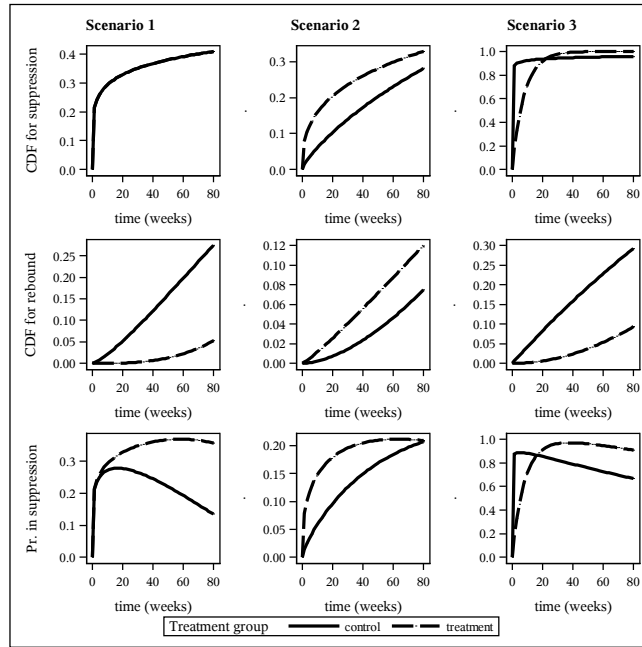
Figure 3.5: Simulations scenarios: CDF for time to suppression, CDF for time to rebound, and probability to be in suppression, by treatment group.

α and β as follows. Scenario 1: the treatment group - $\alpha^S = 0.2, \beta^S = 4000, \alpha^R_{cond} = 4, \beta^R_{cond} = 120$, the control group - $\alpha^S = 0.2, \beta^S = 4000, \alpha^R_{cond} = 1.35, \beta^R_{cond} = 64$. Scenario 2: the treatment group - $\alpha^S = 0.4, \beta^S = 800, \alpha^R_{cond} = 1, \beta^R_{cond} = 120$, the control group - $\alpha^S = 0.8, \beta^S = 320, \alpha^R_{cond} = 1, \beta^R_{cond} = 120$. Scenario 3: the treatment group - $\alpha^S = 1, \beta^S = 8, \alpha^R_{cond} = 2, \beta^R_{cond} = 240$, the control group - $\alpha^S = 0.1, \beta^S = 0.0008, \alpha^R_{cond} = 1, \beta^R_{cond} = 200$. The censoring distribution was the same in both treatment groups and across all scenarios, with $\alpha^{cens} = 1.5, \beta^{cens} = 400$. Treatment assigment was generated as a Bernoulli random variable with success probability 0.5. We assessed several sample sizes between 250 and 2000 patients. All simulations were conducted using 1000 samples. For the proposed method, we defined observed data as $X_i^S = min(T_i^S, C_i)$, $X_i^R = min(T_i^R, C_i)$, $\delta_i^S = I(X_i^S = T_i^S)$ and $\delta_i^R = I(X_i^R = T_i^R)$. We computed the test statistic WG, with each of the three weight functions described in Section 3.2.

To define virologic failure as in the A5142 trial or for the TLOVR-like endpoint, we first chose a cut-off point γ_0, non-suppression prior to which should be considered a failure. Then, given the cut-off, we defined the observed data for the composite event in A5142 as:

$$X_i^{comp} = \begin{cases} min(T_i^R, C_i), & 0 < T_i^S \le \gamma_0 \\ min(\gamma_0, C_i), & \gamma_0 < T_i^S, \end{cases}$$

and

$$\delta_i^{comp} = \begin{cases} I(X_i^{comp} = T_i^R), & 0 < T_i^S \le \gamma_0 \\ I(X_i^{comp} = \gamma_0), & \gamma_0 < T_i^S. \end{cases}$$

Similarly, the data for the TLOVR-like event were defined as:

$$X_i^{TLOVR} = \begin{cases} min(T_i^R, C_i), & 0 < T_i^S \le \gamma_0 \\ 0, & \gamma_0 < T_i^S, \end{cases}$$

and

$$\delta_i^{TLOVR} = \begin{cases} I(X_i^{TLOVR} = T_i^R), & 0 < T_i^S \le \gamma_0 \\ 1, & \gamma_0 < T_i^S. \end{cases}$$

Using the composite endpoint from the A5142 protocol and TLOVR independently, we separately performed two-sided logrank tests and then determined the direction of the difference by fitting the Cox model using treatment group as the sole covariate, to mimic the

intent to treat analysis from (Riddler et al., 2008). We looked at a range of possible cut-off points in the definition of composite events for the A5142 and TLOVR endpoints.

The observed type I error rate was close to the nominal level for all three methods, ranging from 0.041 to 0.057 for the proposed endpoint and from 0.040 to 0.060 for A5142 and TLOVR endpoints (not shown in tables). The results for power are summarized in Table 3.1. For the proposed method, the power to reject the null hypothesis was consistent for all scenarios, for all choices of the weight function, and increased with sample size. However, for the A5142 composite endpoint and for TLOVR, the power varied from being higher than that for the proposed method to being almost zero, depending on the scenario and the choice of the cut-off point γ_0. For scenario 1, the power for both composite endpoints was much higher than for the proposed method. For scenario 2, the power for A5142 and TLOVR endpoints was sometimes worse than for the proposed method, depending on the chosen value of the cut-off. The results for scenario 3 are the most interesting. If we look at which treatment arm was selected under scenario 3, for some values of γ_0, the A5142 and TLOVR analyses always incorrectly selected the control arm. For a large sample size (2000 patients), the null hypothesis was rejected in favor of the wrong treatment group 81% of the time using the A5142 endpoint and 92% of the time using TLOVR. Such a reversal of results happened because both composite endpoints from A5142 and TLOVR re-defined the time of event. For some values of the cut-off γ_0 (prior to 12 weeks in the scenario 3), the failures in the treatment group were forced to happen earlier than in the control group.

Table 3.1: Simulation results: Power to reject the null hypothesis, and the preferred treatment arm, by value of the cut-off time point for the A5142 and TLOVR endpoints, and by weight fucntion for the proposed method.

							Method							
Scenario 1		A5142 trial					TLOVR					Proposed		
		Cut-off (weeks)					Cut-off (weeks)					Weight		
Sample size		8	16	24	32	40	8	16	24	32	40	unity	1/se	cens
125	power	0.58	0.70	0.79	0.86	0.90	0.53	0.57	0.62	0.62	0.64	0.39	0.39	0.36
250	power	0.89	0.95	0.98	0.99	1.00	0.85	0.91	0.92	0.93	0.92	0.67	0.67	0.64
500	power	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.92	0.92	0.89
1000	power	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99

							Method							
Scenario 2		A5142 trial					TLOVR					Proposed		
		Cut-off (weeks)					Cut-off (weeks)					Weight		
Sample size		8	16	24	32	40	8	16	24	32	40	unity	1/se	cens
125	power	0.30	0.19	0.14	0.11	0.06	0.35	0.26	0.21	0.19	0.15	0.31	0.35	0.32
250	power	0.50	0.36	0.22	0.15	0.10	0.57	0.48	0.33	0.30	0.23	0.57	0.63	0.59
500	power	0.79	0.61	0.43	0.26	0.15	0.85	0.76	0.68	0.55	0.43	0.85	0.89	0.86
1000	power	0.98	0.89	0.68	0.45	0.26	0.99	0.97	0.91	0.82	0.69	0.99	0.99	0.99

							Method							
Scenario 3		A5142 trial					TLOVR					Proposed		
		Cut-off (weeks)					Cut-off (weeks)					Weight		
Sample size		8	10	12	14	16	8	10	12	14	16	unity	1/se	cens
250	choose treatment	0.00	0.05	0.25	0.58	0.80	0.00	0.03	0.18	0.48	0.74	0.84	0.87	0.81
	choose control	0.15	0.01	0.00	0.00	0.00	0.21	0.02	0.00	0.00	0.00	0.00	0.00	0.00
2000	choose treatment	0.00	0.10	0.94	1.00	1.00	0.00	0.03	0.81	1.00	1.00	1.00	1.00	1.00
	choose control	0.81	0.10	0.00	0.00	0.00	0.92	0.02	0.00	0.00	0.00	0.00	0.00	0.00

Table 3.2: Simulation results: Predicted vs. observed power.

Sample size	Weight					
	Unity		1/SE		Censoring	
	Observed	Predicted	Observed	Predicted	Observed	Predicted
125	0.48	0.50	0.52	0.56	0.44	0.47
250	0.84	0.82	0.87	0.86	0.81	0.78
500	0.99	0.98	0.99	0.99	0.98	0.98

We also conducted a small simulation study to test the sample size computations for the proposed endpoint. We generated 1000 samples from the known distributions under scenario 3 described above, assuming a known effect size. Based on each simulated sample, we estimated standard deviations for our test statistics and computed predicted power based on the observed standard deviations and hypothesized effect size (using SAS procedure POWER). Then we compared the average predicted power with the power observed in 1000 simulations. The results summarized in Table 3.2 generally exhibit good agreement between the observed and predicted powers.

3.4 Re-analysis of the A5142 Trial

As an example, we re-analysed the ACTG A5142 trial using the virologic failure endpoint from A5142 and the proposed method. The A5142 trial included 753 patients whose baseline viral load was at least 2000 copies/ml. Patients were randomized to one of the three treatment arms, efavirenz plus two NRTIs (efavirenz group), lopinavir–ritonavir plus two NRTIs (lopinavir–ritonavir group), or lopinavir-ritonavir plus efavirenz (NRTI-sparing arm). The median follow-up was 112 weeks, with the longest follow-up time being 157 weeks.

The definition of a virologic failure for A5142 ((Riddler et al., 2008), p.2097) was lack of confirmed viral load suppression below 200 copies/ml or by log_{10} by 8 weeks; or lack of confirmed viral suppression below 200 copies/ml by 32 weeks; or confirmed viral rebound. The definition of viral rebound also varied depending on when the rebound occurred. Early rebound (prior to 32 weeks) was defined as a viral load increase to over 1000 copies/ml for patients whose viral load had suppressed below 200 copies/ml; or viral load increase by log_{10} from the nadir value for patients whose viral load had never suppressed below 200 copies/ml. Late rebound (after 32 weeks) was defined as a viral load $\geq$ 200 copies/ml. 227 patients experienced virologic failure by the A5142 definition. The Kaplan-Meier estimators for virologic failure are shown on the top panel of Figure 3.2.
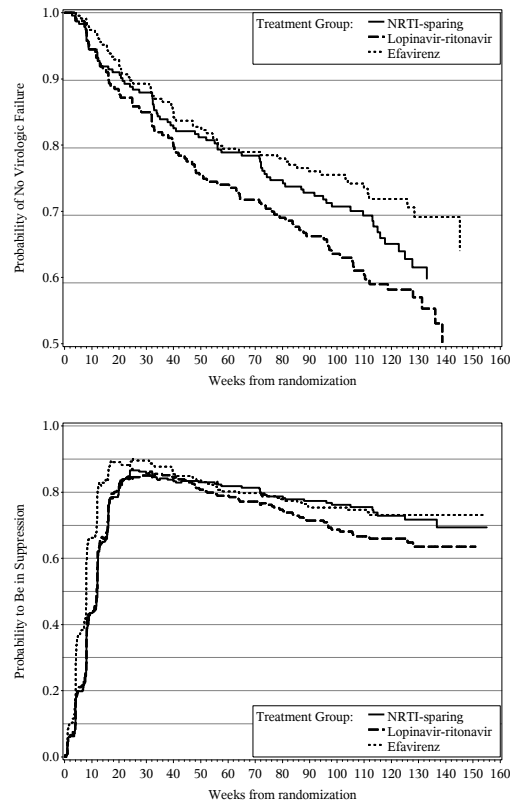
Figure 3.6: Top panel: Survival functions for virologic failure as defined in A5142 trial, by treatment group. Bottom panel: Probability to be in suppression, by treatment group.

Table 3.3: *P-values comparing between treatment groups in the A5142 trial, original analysis vs. proposed method. Endpoint 1: single threshold of 200 copies/ml in the definitions of suppression and rebound. Endpoint 2: different definitions for early and late suppression and rebound. P-values adjusted for multiple comparisons using the Bonferroni correction.*

	Method						
	Riddler et al	Proposed, Endpoint 1			Proposed, Endpoint 2		
		Weight			Weight		
Comparison		unity	1/se	cens	unity	1/se	cens
EFAV vs LOP/RIT	0.006	0.037	0.023	0.019	0.197	0.185	0.178
EFAV vs NRTI	0.490	0.994	1.000	0.521	1.000	1.000	0.807
NRTI vs LOP/RIT	0.130	0.315	0.292	0.400	0.785	0.738	1.000

For the proposed approach, we defined two separate events, viral suppression and viral rebound. We defined viral suppression as viral load being reduced to < 200 copies/ml for two consecutive measurements 4 weeks or less apart. We defined viral rebound as viral load being ≥ 200 copies/ml at two consecutive measurements 4 or less weeks apart. We had 667 patients in all treatment groups whose viral load was suppressed, and 129 patients who experienced viral rebound. A plot of the estimated probability of being suppressed, over time from randomization, by treatment group, is displayed on the bottom panel of Figure 3.2.

Next, we computed the test statistic WG, using the 3 different weight functions introduced in Section 3.2. We integrated over the first 143 weeks of follow-up, capturing almost all information in the dataset. For comparison, logrank tests were also calculated based on the A5142 composite endpoint. Over the 143 week period after randomization, the analysis showed that a patient from the efavirenz group was in a state of viral suppression for 12 weeks longer on average than a patient from the lopinavir-ritonavir group (95% CI=(3,21), p-value after Bonferroni correction $p = 0.032$) and for 3 weeks longer than a patient from the NRTI-sparing group (95% CI=(-3,13), p-value $p = 0.783$). Moreover, a patient from the NRTI-sparing group was in the state of suppression for 5 weeks longer than a patient from the lopinavir-ritonavir group (95% CI=(-2,16), p-value $p = 0.365$). The results of testing the null hypothesis of no difference between the 3 treatment groups using WG are reported in Table 3.3 as endpoint 1. Inferences are similar to those obtained from the original analysis.

Recognizing that there might have been clinical considerations for defining a separate early viral suppression and viral rebound, we performed additional analyses mimicking the definitions of viral suppression and viral rebound from the A5142 trial as closely as possible.

We defined early viral suppression and viral rebound prior to 32 weeks as was done in the A5142 trial. Under this definition, the number of patients in all treatment groups who experienced viral suppression was 691, and who experienced viral rebound was 183. The results of this supplementary analysis are also summarized in Table 3.3 as endpoint 2 and are not statistically significant at 0.05 level, though the direction of the differences remained the same.

We also performed sensitivity analysis to assess how much the results of the A5142 trial depended on the choice of cut-off time of 8 weeks for early rebound and 32 weeks for late rebound. Judging by the plot of the probability to be in suppression by treatment group, we did not expect inference to change when we varied the cut-off times for early and late viral rebound. This is because the best treatment group was uniformly better than the second best treatment group both in terms of viral suppression and viral rebound, with the same ordering holding for the second and third best treatment groups. We re-defined virologic failure using cutoffs ranging from 5 to 15 weeks for early rebound and from 25 to 40 weeks for late rebound. The results confirmed our expectations: the p-values for comparison of the efavirenz and lopinavir-ritonavir groups remained significant and ranged from 0.0107 to 0.0306 (after a Bonferroni correction), all other comparisons were still not statistically significant, and all the differences between the groups were in the same direction.

In summary, certain advantages of the proposed endpoint can be clearly seen in Figure 3.2, where the time-specific treatment differences are cleanly summarized via the probability of being in suppression. The efavrienz group suppresses most rapidly and with higher probability and the suppression is maintained as effectively as in the NRTI-sparing arm. The NRTI-sparing arm has comparable early suppression to that in the lopinavir group, but with superior long term maintenance. Such information is not as readily gleaned from the plot of the survival curves for the A5142 composite endpoint.

3.5   Discussion

DeGruttola (DeGruttola et al., 1998) were the first to discuss the use of HIV-1 RNA viral load as an outcome measure in HIV trials, both as a repeatedly-assessed continuous biomarker and as an indicator of treatment (virologic) failure. (Gilbert et al., 2000) expand on the discussion of virologic failure and consider several competing definitions. (Ribaudo et al., 2006) discussed design issues in HIV trials, concentrating the discussion of endpoints on further refinements in virologic failure. To the best of our knowledge, no one has previously suggested the combined endpoint we propose here.

In this chapter we implemented a novel approach to defining a time to event endpoint in

HIV research that combines time to viral suppression and time to viral rebound into a single measure, the probability of being suppressed over time. As demonstrated in the A5142 data analysis, this quantity precisely captures the interplay of suppression and rebound, yielding a simple graphical representation of early and late suppression dynamics which may be preferable to that for the existing composite endpoints. The integrated probability of suppression can easily be adapted by choice of the weight function to target specific time periods of interest. Employing unity weight provides a particularly attractive summary which may be interpreted as the average number of weeks suppressed over the time period of interest. If there is scientific justification to disregard a portion of the follow-up period, the weights function can be set to zero for those times points.

The probability of suppression endpoint may also be useful in observational studies, albeit with the necessary caveats about confounding. Further work is needed to appropriately adjust for confounding factors in the analysis. Future research is planned into regression modelling of the probability of suppression and the associated weighted average times in suppression. However, the application of the proposed endpoint to observational studies is beyond the scope of the current manuscript which deals with intent to treat analyses in randomized HIV trials with HIV RNA measurements obtained on a specific predefined schedule.

Chapter 4. On the Choice of Time Scales in Competing Risks Predictions

4.1 Introduction

The purpose of this chapter is to introduce a semi-parametric estimator for cumulative incidence functions which uses different time scales for different event types.

As an example, we consider the colon cancer population in the Surveillance, Epidemiology, and End Results (SEER) database. The current colon cancer risk estimates provided by the National Cancer Institute utilize proportional hazards models for the cause-specific hazards for colon cancer and for other causes which adjust for patient specific risk factors. In chronic disease registries like the SEER database, subjects are typically enrolled in the database at the time of disease diagnosis. A primary objective may be to understand the impact of risk factors on subsequent mortality both from the disease and from other causes. Proportional hazards model on the time since diagnosis scale is usually a reasonable assumption when the event is death from cancer. However, for death from other causes it may be not justified, and assuming the model proportional on the calendar age scale may be more appropriate. In this case the estimator from Cheng et al may produce biased estimates and prediction based on it may be incorrect.

In Section 4.2 we propose semi-parametric methods which allow to analyze each event time using its own time scale. The simulation results are provided in Section 4.3. The brief summary of the data analysis is provided in Section 4.4. Asymptotic properties of the estimator are provided in the Appendix.

4.2 Methods

Suppose we are interested in predicting the probability that a patient diagnosed with cancer at some age D_0 will die from cancer by some time t after the diagnosis. Suppose we have two competing causes of death and hence events of two types, event 1 - death from cancer, and event 2 - death from other causes. Suppose the data that we observe for each patient $i = 1, ..., n$ is $(D_i, X_i, \delta_{1i}, \delta_{2i}, U_i, Z_i(D_i))$, where D_i is age at which a patient is diagnosed with cancer, $X_i = min(T_i, C_i)$, T_i - time of event on time since diagnosis scale, C_i - censoring time on time since diagnosis scale, δ_{ik} for $k = 1, 2$ - event indicators for event of type k, and $U_i, Z_i(D_i)$ - covariate vectors. Covariates U_i are time-independent, such as gender, race etc, while covariates $Z_i(D_i)$ are time-dependent and measured at the time of diagnosis D_i, such as tumor size, comorbidity index etc. Since our interest is prediction which is supposed to be made at the moment of diagnosis and we don't want our prediction to depend on future

values of covariates which they may take after the diagnosis, we fix the time-dependent covariates at the moment of diagnosis and we assume that the models for events of both types depend only on the covariate values measured at the time of diagnosis. Suppose we only have data on patients who have been diagnosed prior to death or censoring. Then time of event on the age scale is $T_i^{age} = D_i + T_i$, and observed time is $X_i^{age} = D_i + X_i$. We want to predict the probability that a patient diagnosed at age D_0 will experience an event of type 1 by time t after diagnosis, given covariate values $U_0, Z_0(D_0)$. Note that we can chose on which time scale we want to work, time since diagnosis scale or calendar age scale. On calendar age scale the observed data is $(D_i, X_i^{age}, \delta_{1i}, \delta_{2i}, U_i, Z_i(D_i))$ and the data is left-truncated by D_i - age at diagnosis.

If the proportional hazards model holds for events of both types on the time since diagnosis scale, it makes sense to chose this scale. The cause-specific hazards for events of both types on the time since diagnosis scale can be estimated using the Cox model, treating events of other type as censoring. The model for cause-specific hazards will be the following:

$$\lambda_1(t|D_i, U_i, Z_i(D_i)) = \lambda_{01}(t)exp(\beta_{1D}D_i + \beta_{1U}U_i + \beta_{1Z}Z_i(D_i))$$

$$\lambda_2(t|D_i, U_i, Z_i(D_i)) = \lambda_{02}(t)exp(\beta_{2D}D_i + \beta_{2U}U_i + \beta_{2Z}Z_i(D_i))$$

where $t > 0$ is time since diagnosis. Age at diagnosis D_i is typically included as a covariate in the model for both types of event. Predicted cumulative incidence functions for events $k = 1, 2$ can be computed as described in Cheng et al :

$$\hat{F}_k(t|D_0, U_0, Z_0(D_0)) = \int_0^t \hat{S}(u|D_0, U_0, Z_0(D_0))d\hat{\Lambda}_k(u|D_0, U_0, Z_0(D_0))$$

where

$$\hat{S}(u|D_0, U_0, Z_0(D_0)) = exp\{-\sum_{k=1}^2 \hat{\Lambda}_k(u|D_0, U_0, Z_0(D_0))\},$$

$$\hat{\Lambda}_k(u|D_0, U_0, Z_0(D_0)) = \hat{\Lambda}_{0k}(u)exp(\hat{\beta}_{kD}D_0 + \hat{\beta}_{kU}U_0 + \hat{\beta}_{kZ}Z_0(D_0)),$$

$$\hat{\Lambda}_{0k}(u) =$$

$$\sum_{i=1}^{n} \delta_{ki} I(X_i \le u) \Big\{ \sum_{j=1}^{n} I(X_i \le X_j) exp(\beta_{kD} D_j + \beta_{kU} U_j + \beta_{kZ} Z_j(D_j)) \Big\}^{-1}.$$

However, as mentioned above, it is possible that the models for events of different types are not proportional on the same time scale. To deal with the case when the model for event 1 is proportional on the time since diagnosis scale and the model for the event 2 is proportional on the age scale but not on the time since diagnosis scale, we suggest to estimate cause-specific hazards on two different time scales.

Let the true model for the cause-specific hazards for event 1 be on the time since diagnosis scale as folows:

$$\lambda_1^{time}(t|D_i, U_i, Z_i(D_i)) \quad = \quad \lambda_{01}(t) exp(\beta_{1D} D_i \quad + \quad \beta_{1U} U_i \quad + \quad \beta_{1Z} Z_i(D_i)),$$

where $t > 0$ is time since diagnosis. We can express the hazard function for event 1 on the age scale as:

$$\lambda_1^{age}(a|D_i, U_i, Z_i(D_i)) =$$

$$= \begin{cases} h_1(a, U_i, \{Z_i(s), 0 < s < D_i\}), & 0 < a < D_i \\ \lambda_{01}(a - D_i) exp(\beta_{1D} D_i + \beta_{1U} U_i + \beta_{1Z} Z_i(D_i)), & a \ge D_i. \end{cases}$$

where $a > 0$ is age.

Similarly, let the true model for the cause-specific hazard for event 2 on the age scale be:

$$\lambda_2^{age}(a|D_i, U_i, Z_i(D_i)) =$$

$$= \begin{cases} h_2(a, U_i, \{Z_i(s), 0 < s < D_i\}), & 0 < a < D_i \\ \lambda_{02}(a) exp(\beta_{2D} D_i + \beta_{2U} U_i + \beta_{2Z} Z_i(D_i)), & a \ge D_i \end{cases}$$

Then on the time since diagnosis scale:

$$\lambda_2^{time}(t|D_i, U_i, Z_i(D_i)) \quad = \quad \lambda_{02}(D_i \quad + \quad t) exp(\beta_{2D} D_i \quad + \quad \beta_{2U} U_i \quad + \quad \beta_{2Z} Z_i(D_i)).$$

In the expressions above, $h_1(.)$ and $h_2(.)$ are some unspecified non-negative functions which may depend on the time-independent covariates U_i and on the history of time-

dependent covariates up to the time of diagnosis, $\{Z_i(s), 0 < s < D_i\}$. The functions $h_1(.)$ and $h_2(.)$ are not estimable from our data without making additional assumptions about the joint distribution of D and T^{age}, because we do not have survival data from the patients prior to them being diagnosed. Therefore all the inference we can make will be conditional on $D < T^{age}$. Since our ultimate interest is making prediction for a patient who has been diagnosed and is still alive, such conditional inference will be enough. Also, in order to be able to make a valid prediction from the moment of diagnosis on, we assume that after the time of diagnosis the models depend only on the values of covariates measured at the time of diagnosis and do not depend on future values of time-dependent covariates. Note that technically in general case $\lambda_2^{age}(a|D_i, U_i, Z_i(D_i))$ may be not proportional on the age scale on the whole age axis from 0 to ∞ since we do not specify the model prior to the point D_i. However, conditional on $D_i < T_i^{age}$, the hazards for event 2 are proportional on the age scale, and conditional inference is all that we are attempting to make.

The probabilities which we are interested to predict can be expressed as:

$$
F_1(t|D_0, U_0, Z_0(D_0)) = Pr(T \leq t, \delta_1 = 1 | D < T^{age}, D_0, U_0, Z_0(D_0))
$$
$$
= \int_0^t exp[-\int_0^u \{\lambda_1^{time}(s|D_0, U_0, Z_0(D_0)) + \lambda_2^{age}(D_0 + s|D_0, U_0, Z_0(D_0))\}ds] \times
$$
$$
\times \lambda_1^{time}(u|D_0, U_0, Z_0(D_0))du
$$

and

$$
F_2(t|D_0, U_0, Z_0(D_0)) = Pr(D_0 < T^{age} \leq D_0 + t, \delta_2 = 1 | D < T^{age}, D_0, U_0, Z_0(D_0))
$$
$$
= \int_{D_0}^{D_0+t} exp[-\int_{D_0}^{D_0+u} \{\lambda_1^{time}(s - D_0|D_0, U_0, Z_0(D_0)) + \lambda_2^{age}(s|D_0, U_0, Z_0(D_0))\}ds] \times
$$
$$
\times \lambda_2^{age}(u|D_0, U_0, Z_0(D_0))du
$$

We can estimate cause-specific hazards for events of types 1 and 2 on time since diagnosis and calendar age scales respectively by fitting two Cox models, one on the time since diagnosis scale and the other - on the calendar age scale, treating events of other types as censoring and accounting for the left truncation on the age scale. The estimates for the cause-specific

hazard functions will be:

$$\hat{\lambda}_1^{time}(X_i|D_0, U_0, Z_0(D_0)) = \delta_{1i}exp(\beta_{1U}U_0 + \beta_{1Z}Z_0(D_0) + \beta_{1D}D_0)\times$$

$$\{\sum_{j=1}^{n} I(X_i \leq X_j)exp(\beta_{1U}U_j + \beta_{1Z}Z_j(D_j) + \beta_{1D}D_j)\}^{-1}.$$

and

$$\hat{\lambda}_2^{age}(X_i^{age}|D_0, U_0, Z_0(D_0)) = \delta_{2i}exp(\beta_{2U}U_0 + \beta_{2Z}Z_0(D_0) + \beta_{2D}D_j)\times$$

$$\{\sum_{j=1}^{n} I(D_j \leq X_i^{age} \leq X_j^{age})exp(\beta_{2U}U_j + \beta_{2Z}Z_j(D_j) + \beta_{2D}D_j)\}^{-1}.$$

Then using the above we can estimate the probabilities of interest similar to Cheng et al. as:

$$\hat{F}_1(t|D_0, U_0, Z_0(D_0)) = \sum_{X_i \leq t} exp[- \sum_{X_j \leq X_i} \hat{\lambda}_1^{time}(X_j|D_0, U_0, Z_0(D_0))$$

$$- \sum_{D_0 \leq X_j^{age} \leq D_0 + X_i} \hat{\lambda}_2^{age}(X_j^{age}|D_0, U_0, Z_0(D_0))]\hat{\lambda}_1^{time}(X_i|D_0, U_0, Z_0(D_0))$$

and

$$\hat{F}_2(t|D_0, U_0, Z_0(D_0))$$

$$= \sum_{D_0 \leq X_i^{age} \leq D_0 + t} exp[- \sum_{0 < X_j \leq X_i^{age} - D_0} \hat{\lambda}_1^{time}(X_j|D_0, U_0, Z_0(D_0))$$

$$- \sum_{D_0 \leq X_j^{age} \leq X_i^{age}} \hat{\lambda}_2^{age}(X_j^{age}|D_0, U_0, Z_0(D_0))]\hat{\lambda}_2^{age}(X_i^{age}|D_0, U_0, Z_0(D_0)).$$

The proposed estimators are asymptotically normal and have closed form variance estimator. The formulas and the proofs are provided in the Appendix.

4.3   Simulations

To evaluate the performance of the proposed estimator we conducted the simulations experiment. We generated the data described in the Section 4.2. The model for event of type 1 was proportional on the time since diagnosis scale. The model for event of type 2 was proportional on the calendar age scale but not proportional on the time since diagnosis scale.

The proposed estimator on two time scales was compared with the estimator from Cheng et al which used the time since diagnosis scale for the events of both types, as described in the Section 4.2 above.

The data were generated as follows. We used four covariates: age at diagnosis D, a time-independent categorical covariate U, which can be, for example, gender, and two time-dependent continuous covariates measured at the time of diagnosis, say, tumor size $Z1(D)$ to be included in the model for death from cancer (event 1), and comorbidity index $Z2(D)$, to be included in the model for death from other causes (event 2).

First, all the covariates were generated. Age at diagnosis D was generated as a normal random variable, with mean μ_D and variance σ_D^2. Gender U was generated as a Bernoulli random variable with the success probability π. Tumor size $Z1(D)$ was generated as a normal random variable with the mean μ_1 and variance σ_1^2, and then rounded to an integer. Comorbidity index $Z2(D)$ was generated as a normal random variable with the mean increasing with age, $\mu_2 = D/10$ and variance σ_2^2, and then rounded to an integer.

The cause-specific hazards on the age scale for events of type 1 and 2 were chosen as follows:

$$\lambda_1^{age}(a|D,U,Z1(D)) \quad = \quad \begin{cases} 0, & 0 < a < D \\ \alpha_1 exp(\beta_{1U}U + \beta_{1Z1}Z1(D) + \beta_{1D}D), & a \geq D. \end{cases}$$

and

$$\lambda_2^{age}(a|D,Z2(D)) =$$
$$= \begin{cases} \alpha_{21}I(a < \gamma_2) + \alpha_{22}I(a \geq \gamma_2), & 0 < a < D \\ \{\alpha_{21}I(a < \gamma_2) + \alpha_{22}I(a \geq \gamma_2)\}exp(\beta_{2Z2}Z2(D) + \beta_{2D}D), & a \geq D \end{cases}$$

After all the covariates were generated, the above hazard functions could be evaluated and the overall survival function on the age scale could be computed as

$$S^{age}(a|D,U,Z1(D),Z2(D)) =$$
$$exp(-\int_0^a [\lambda_1^{age}(s|D,U,Z1(D)) + \lambda_2^{age}(s|D,Z2(D))]ds).$$

Then we generated time of event T^{age} on the calendar age scale as a random variable having the survival function $S^{age}(a|D,U,Z1(D),Z2(D))$ and event indicators as binary random

variables with probabilities of success:

$$Pr(\delta_1 = 1) = \frac{\lambda_1^{age}(T^{age}|D, U, Z1(D))}{\lambda_1^{age}(T^{age}|D, U, Z1(D)) + \lambda_2^{age}(T^{age}|D, Z2(D))}$$

and

$$Pr(\delta_2 = 1) = 1 - Pr(\delta_1 = 1).$$

If D was less than T^{age} then the observation was kept in the sample and T was defined as $T^{age} - D$. Time of censoring C was generated as a uniform random variable $Uniform(\gamma_C)$, where γ_C is chosen to achieve the desired proportion of censored observations, and then X was set to be $min(T, C)$. The procedure was repeated until we reached the desired number of observations with $D < T^{age}$.

All the simulations were performed with 1000 iterations. We looked at 2 sample sizes, 200 and 600, and two levels of censoring, about 10% and about 35%.

We report two scenarios here, Scenario 1 with the model for event 2 non-proportional on the time since diagnosis scale, to demonstrate the advantages of using the proposed estimator on two time scales vs. Cheng et al, and Scenario 2 with the model for event 2 proportional on the time since diagnosis scale, to assess the loss of efficiency compared to Cheng et al when estimating both hazards on the time since diagnosis scale is appropriate.

Common parameter values for both scenarios were: $\alpha_1 = 0.03, \beta_{1U} = 0.1, \beta_{1D} = 0.01, \beta_{1Z1} = 0.01, \gamma_2 = 40, \beta_{2D} = 0.01, \beta_{2Z2} = 0.01, \pi = 0.5, \mu_D = 50, \sigma_D^2 = 25, \mu_1 = 5, \sigma_2^2 = 2, \mu_2 = D/10, \sigma_2^2 = 2$. Values of covariates for which we made all the predictions were: $U = 1, D = 30, Z1 = 5, Z2 = 3$. For Scenario 1, the values of α_{21} and α_{22} were not equal, $\alpha_{21} = 0.05, \alpha_{22} = 0.5$. For Scenario 2 they were equal to make the model for event 2 proportional on the time since diagnosis scale, $\alpha_{21} = \alpha_{22} = 0.1$. The 10% and 35% levels of censoring were achieved by setting γ_C to be 70 and 25 respectively for Scenario 1, and 60 and 30 for Scenario 2.

For both scenarios, the empirical variance of the estimates was close to the mean estimated variance. Under Scenario 1, when the model for event 2 was not proportional on the time since diagnosis scale, the proposed estimator on two time scales had small bias for events of both type which improved with the increase of sample size, while Cheng et al. had a very large systematic bias for event of type 2, and hence a significant bias for event of type 1. The coverage probability for the Cheng et al estimator under Scenario 1 was very poor, while for the proposed estimator on two time scales it was close to nominal.

Another important observation was that under Scenario 1, with the increase of the sample size the coverage probability for Cheng et al. estimator declined dramatically, due to the

fact that the systematic bias stayed the same and the variance of the estimates decreased with the increase of sample size. Therefore the use of the Cheng et al when the model for event of type 2 is not proportional on the time since diagnosis scale may lead to a very undesirable situation when with the increase of the sample size the quality of prediction becomes unacceptable instead of improving.

Under Scenario 2, when the model for cause 2 was proportional on the time since diagnosis scale and hence the use of Cheng et al. was appropriate, we observed some loss of efficiency by the estimator on two time scales due to left truncation on the age scale. The variance of the proposed estimator on two time scales for event 1 was about 1.5 times larger than the variance of the estimator from Cheng et al., and for event 2 it was about 3-5 times larger. Such increase of the variance can be explained by the fact that when we estimate the hazard functions on the age scale for event 2, due to left truncation we have fewer subjects at risk for each event of type 2 than when we have when we work on the time since diagnosis scale. Therefore the variability of the estimates increases compared to Cheng et al, and this is the price one has to pay for eliminating the bias. The coverage probabilities for the proposed estimator on two scales were slightly lower than nominal for sample size 200 and improved to almost nominal level when sample size increased to 600.

The results are summarized in tables 4.1 and 4.2 below.

Table 4.4: Scenario 1. Model for event 2 non-proportional on the time since diagnosis scale.

				Two Time Scales				Cheng et al.			
Censored	Event	Time	True	Bias	Emp.Var	$\widehat{Var}$	CP	Bias	Emp.Var	$\widehat{Var}$	CP
				Sample size = 200							
10%		5	0.178	−0.0024	0.00190	0.00194	0.95	−0.0237	0.00147	0.00148	0.89
		10	0.278	−0.0019	0.00371	0.00372	0.95	−0.0619	0.00229	0.00229	0.75
	2	5	0.264	−0.0073	0.00403	0.00420	0.96	0.1668	0.00174	0.00217	0.03
		10	0.411	−0.0119	0.00437	0.00428	0.94	0.2168	0.00205	0.00253	0.01
Censored	Event	Time	True	Bias	Emp.Var	$\widehat{Var}$	CP	Bias	Emp.Var	$\widehat{Var}$	CP
				Sample size = 200							
35%	1	5	0.178	−0.0018	0.00231	0.00226	0.94	−0.0228	0.00176	0.00171	0.89
		10	0.278	−0.0021	0.00474	0.00450	0.95	−0.0599	0.00288	0.00274	0.78
	2	5	0.264	−0.0086	0.00492	0.00557	0.96	0.1607	0.00228	0.00239	0.08
		10	0.411	−0.0122	0.00510	0.00553	0.95	0.2080	0.00268	0.00304	0.04
Censored	Event	Time	True	Bias	Emp.Var	$\widehat{Var}$	CP	Bias	Emp.Var	$\widehat{Var}$	CP
				Sample size = 600							
10%	1	5	0.178	0.0003	0.00068	0.00064	0.94	−0.0216	0.00051	0.00049	0.82
		10	0.278	0.0009	0.00123	0.00124	0.95	−0.0598	0.00077	0.00076	0.45
	2	5	0.264	−0.0045	0.00147	0.00143	0.94	0.1689	0.00062	0.00072	0.00
		10	0.411	−0.0054	0.00152	0.00143	0.94	0.2207	0.00078	0.00083	0.00
Censored	Event	Time	True	Bias	Emp.Var	$\widehat{Var}$	CP	Bias	Emp.Var	$\widehat{Var}$	CP
				Sample size = 600							
35%	1	5	0.178	−0.0003	0.00070	0.00075	0.95	−0.0218	0.00054	0.00056	0.86
		10	0.278	0.0001	0.00141	0.00150	0.96	−0.0586	0.00088	0.00090	0.52
	2	5	0.264	−0.0058	0.00183	0.00189	0.95	0.1632	0.00068	0.00079	0.00
		10	0.411	−0.0073	0.00185	0.00185	0.94	0.2126	0.00090	0.00099	0.00

Table 4.5: Scenario 2. Model for event 2 proportional on the time since diagnosis scale.

Censored	Event	Time	True	Two Time Scales				Cheng et al.			
				Bias	Emp.Var	$\widehat{Var}$	CP	Bias	Emp.Var	$\widehat{Var}$	CP
				Sample size = 200							
10%	1	5	0.153	0.0004	0.00230	0.00219	0.93	−0.0008	0.00200	0.00190	0.94
		10	0.213	0.0018	0.00410	0.00384	0.93	−0.0018	0.00310	0.00297	0.95
	2	5	0.453	−0.0278	0.01203	0.01157	0.94	−0.0013	0.00278	0.00271	0.95
		10	0.631	−0.0416	0.00758	0.00836	0.92	−0.0049	0.00343	0.00324	0.94
Censored	Event	Time	True	Bias	Emp.Var	$\widehat{Var}$	CP	Bias	Emp.Var	$\widehat{Var}$	CP
				Sample size = 200							
35%	1	5	0.153	−0.0027	0.00250	0.00235	0.92	−0.0045	0.00204	0.00202	0.94
		10	0.213	−0.0030	0.00445	0.00416	0.94	−0.0076	0.00323	0.00322	0.95
	2	5	0.453	−0.0351	0.01377	0.01319	0.93	−0.0024	0.00297	0.00294	0.94
		10	0.631	−0.0485	0.00907	0.00962	0.91	−0.0033	0.00364	0.00359	0.94
Censored	Event	Time	True	Bias	Emp.Var	$\widehat{Var}$	CP	Bias	Emp.Var	$\widehat{Var}$	CP
				Sample size = 600							
10%	1	5	0.153	−0.0007	0.00072	0.00070	0.95	−0.0013	0.00059	0.00061	0.95
		10	0.213	0.0000	0.00128	0.00124	0.95	−0.0015	0.00092	0.00097	0.96
	2	5	0.453	−0.0097	0.00415	0.00401	0.93	−0.0003	0.00079	0.00089	0.97
		10	0.631	−0.0135	0.00250	0.00271	0.95	−0.0013	0.00095	0.00105	0.96
Censored	Event	Time	True	Bias	Emp.Var	$\widehat{Var}$	CP	Bias	Emp.Var	$\widehat{Var}$	CP
				Sample size = 600							
35%	1	5	0.153	−0.0014	0.00075	0.00077	0.95	−0.0013	0.00065	0.00066	0.94
		10	0.213	−0.0020	0.00133	0.00138	0.95	−0.0022	0.00103	0.00106	0.95
	2	5	0.453	−0.0065	0.00451	0.00478	0.96	0.0009	0.00090	0.00095	0.96
		10	0.631	−0.0122	0.00285	0.00315	0.96	−0.0004	0.00107	0.00116	0.96

Table 4.6: Covariate Distribution in the analysis data set based on SEER Colon Cancer Data.

Characteristic		Mean(SD) or %	Characteristic		%
Age at diagnosis		77 (6.8)	Site:	proximal	62.3
				distal	37.7
Race/ethnicity	Non-Hispanic white	81.3			
	Hispanic	5.1	Substage	III A	9.7
	Non-Hispanic black	6.8		III B	60.6
	AI/AN	0.3		III C	29.7
	Asian/PI	6.6	Grade	I/II	69.6
				III/IV	30.4
Gender	Male	43.1			
	Female	56.9	Comorbidity scores	0	60.1
				$0 < -1$	30.5
Marital status	Married	48.2		$1 < -2.2$	9.0
	Single	51.8		≥ 2.2	0.4

4.4 Real Data Example

We illustrated our inference procedure using stage III colon cancer data obtained from the SEER program of National Cancer Institute. We analyzed data from patients diagnosed with colon cancer between years 1994 and 2005, with age at diagnosis 66 years and older, restricted to patients with stage III, surgery performed and from the 13 SEER registries (except Alaska and the state of California). For our analysis we used data on the following covariates: tumor site (distal vs proximal), substage, grade, marital status, race/ethnicity, gender, age at diagnosis, comorbidity score, and year of diagnosis. The comorbidity scores were computed from Medicare (part B) data for the year preceding the event of diagnosis. This limited the analysis only to those patients who had Medicare data, hence the restriction on age at diagnosis being 66 and older. Of the 14,657 patients, 5,685 patients (38.8%) died from colon cancer, 3,123 patients (21.3%) died from other causes, and 5,849 patients (39.9%) were censored. The maximum follow-up was 10 years. The descriptive statistics for the sample are shown in Table 4.3.

We analyzed the data two ways, using time since diagnosis time scale for both event types (standard analysis), and using time since diagnosis scale for cancer death and age scale for death of other causes (proposed method). Under both approaches, we fitted a proportional hazards model for cause-specific hazards. Only covariates significant on the 0.05 level were kept in the final models. For the death of cancer, the covariates were age

at diagnosis, race, gender, marital status, year of diagnosis, and all three of cancer-related covariates, tumor site, grade, and substage. For the death of other causes, the final set of covariates on both time scales was age at diagnosis, gender, race, marital status, year of diagnosis, comborbisity score, interaction between comobridity score and age at diagnosis, and, somewhat surprisingly, tumor grade. We obtained regression coefficients from models on both time scales, and then computed predicted cumulative incidence functions based on both methods. The regression coefficients are reported in Table 4.4, and plots of cumulative incidence functions for two specific combinations of covariate values are shown in Figure 4.1.

Table 4.7: Regression coefficients for death of cancer and of other causes obtained by the standard and the proposed methods.

| | Death of colon cancer | | | Death of other causes | | | | | |
| | Time since diagnosis scale (both methods) | | | Time since diagnosis scale (standard method) | | | Age scale (proposed method) | | |
Characteristic	$\hat{\beta}$	SE	P-value	$\hat{\beta}$	SE	P-value	$\hat{\beta}$	SE	P-value
Site, distal vs proximal	-0.084	0.029	0.003						
Substage, III B vs III A	0.833	0.068	< .001						
Substage, III C vs III A	1.416	0.069	< .001						
Grade, III/IV vs I/II	0.261	0.029	< .001	0.12	0.039	0.002	0.119	0.04	0.003
Single vs married	0.129	0.03	< .001	0.302	0.041	< .001	0.294	0.041	< .001
Race (reference - non-Hispanic white)									
Hispanic	0.114	0.06	0.058	-0.041	0.089	0.647	-0.026	0.089	0.773
Non-Hispanic black	0.232	0.051	< .001	0.219	0.068	0.001	0.236	0.068	0.001
AI/AN	0.447	0.251	0.074	0.501	0.318	0.116	0.447	0.319	0.161
API	-0.207	0.059	< .001	-0.38	0.088	< .001	-0.376	0.088	< .001
Female vs male	-0.108	0.029	< .001	-0.358	0.04	< .001	-0.355	0.04	< .001
Age at diagnosis	0.036	0.002	< .001	0.097	0.004	< .001	-0.018	0.009	0.035
Comorbidity				4.118	0.368	< .001	3.765	0.358	< .001
Year of diagnosis	-0.024	0.004	< .001	-0.023	0.006	< .001	-0.029	0.006	< .001
Age at diagnosis x comorbidity				-0.04	0.005	< .001	-0.035	0.005	< .001

The results from both methods are very similar. We only have one set of coefficients for the cancer death, because both approaches required fitting the same model on the same time since diagnosis scale for death of cancer. For death from other causes we have two sets of regression coefficients, one from the model on the time since diagnosis scale, and the other from the model on the age scale. The point estimates and the standard errors are almost identical. The only coefficient notably different between the two models is for age at diagnosis. This could be expected, however, because of the different interpretation of the effect of age in these two models. In the model on the time since diagnosis scale, age at diagnosis captures all the age effect, while in the model on the age scale, age effect is mostly accounted for by the baseline hazard.

Predicted cumulative incidence functions obtained by both methods turned out to be very similar either. The top panel of Figure 4.1 shows plots of predicted probabilities of death from cancer and from other causes over time for the first 10 years after cancer diagnosis for a non-hispanic white man, married, diagnosed in 2005 with cancer with substage IIIB, grade III/IV, proximal site, at age 76 years, which is close to the mean age at diagnosis in this data set, and with a relatively low comorbidity score of 0.4. 10 years after the diagnosis, this patient has about 40% probability to die of cancer, and about 35% probability to die of other causes. There is virtually no difference between predictions obtained by the standard methods and the proposed method.

The plots in the bottom panel of the Figure 4.1 are for the patient with the same covariate values, but diagnosed at a younger age, 66 years, and with the high comorbidity score of 1.4. Even though this patient is much younger, he has a much higher probability to die of other causes 10 years after the diagnosis, about 55%, due to the high comorbidity score. The plots on the bottom panel demonstrate the most extreme difference between two different prediction methods which we observed in this data. The predictions differ for this particular set of covariates most likely due to the fact that the covariate values are very extreme: 66 years is the earliest possible age in the data set, essentially the left boundary, and 1.4 is a very high comorbidity score, about the 90th percentile in this data.

The fact that the two methods produced similar results suggests that for our data the model on the time since diagnosis scale for death from other causes might have been specified correctly, and the distribution of time to death from other causes might be Gompertz or very close to Gompertz. To verify this, we fitted a fully parametric Gompertz model for death from other causes on the age scale scale. We then plotted the baseline hazard estimate from the parametric model and compared it with the Breslow estimator for the baseline hazard from the semi-parametric model (Figure 4.2). The estimates are close. This is quite
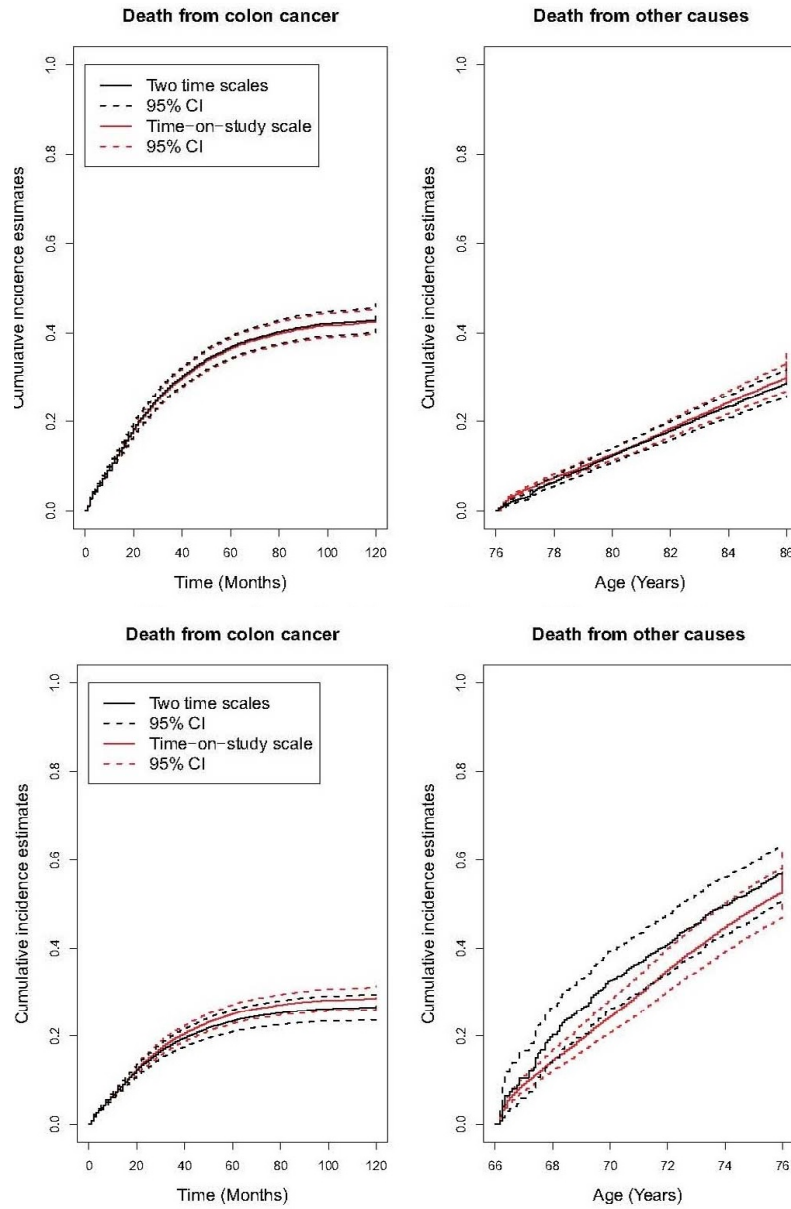
Figure 4.7: Predicted cumulative incidence functions for death of cancer (left column) and of other causes (right column), with 95% pointwise confidence intervals. Top panel shows prediction for a patient diagnosed at 76 years with a comorbidity score 0.4; bottom panel shows prediction for a patient diagnosed at 66 years with a comorbidity score 1.4.

a natural result, if we recall that the Gompertz distribution was originally introduced to describe all-cause mortality in adult human population (Gompertz, 1825).
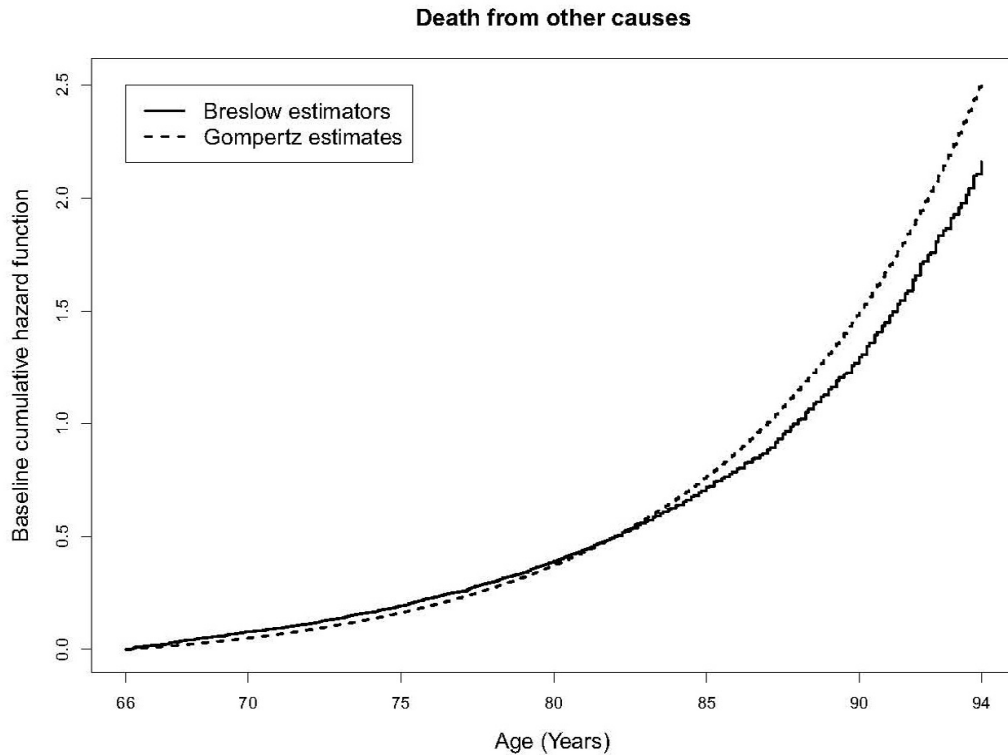

Figure 4.8: Baseline hazard from the model for death from other causes on the age scale. Breslow estimator from the semi-parametric model vs. Gompertz hazard from a fully parametric model.

4.5 Discussion

In theory, there may be situations where cause-specific hazard models for the two causes are both correctly specified on the age scale. In such scenarios, one may modify the single time scale methodology in Cheng et al. to obtain valid estimates of the cumulative incidence functions fitting both cause-specific models using the left truncation approach on the age scale. If the models for death from cancer and death from other causes are correctly specified on both time scales, then the two time scales approach is also valid, as well as fitting both models on the time since diagnosis scale. Which approach is preferred in this setting? One issue is ease of interpretation. For cancer mortality arising in disease registries, it is easier to interpret results on the time since diagnosis scale than on the age scale. In other applications,

it is possible that the age scale may be more easily interpreted. Another issue is the efficiency of the predictions. Simulation results (not shown) demonstrate that predictions based on fitting the two cause-specific models on the age scale using left truncated methodology for both events are generally less efficient than those based on the two time scale approach when both approaches are valid (sometimes much less efficient). In a similar vein, the two time scales approach may be less efficient than single time scale approach on the time since diagnosis scale when both approaches are valid. In general, fitting models on the age scale using left truncated methodology may will yield less efficient predictions than those based on fitting models on the time since diagnosis scale where left truncation is not needed. Such loss of efficiency may be quite large, particularly compared to the loss of efficiency in estimation of the hazard ratio regression parameters, which tends to be modest. This occurs because of the large loss in efficiency in estimating the baseline hazard function on the age scale versus on the time since diagnosis scale. If the predictions are the primary focus, then one would want to use methods which are most efficient amongst those methods which are valid.

4.6 Appendix

Let $W_i = (A_i, Z_{1i}, Z_{2i}(A))$, $w_0 = (a_0, z_{01}, z_{02}(a_0))^T$, $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13})^T$ and $\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23})^T$. Let $\lambda_1^{time}(t; w_0) = \lambda_{01}(t)exp(\beta_1^T w_0)$ and $\lambda_2^{age}(a; w_0) = \lambda_{02}(a)exp(\beta_2^T w_0)$, where $a = a_0 + t$.

Define the counting process notation $N_{1i}(t) = I(X_i \leq t, \delta_{1i} = 1)$, $N_{2i}(t) = I(A \leq X_i^{age} \leq t, \delta_{2i} = 1)$, $Y_{1i}(u) = I(X_i \geq u)$, and $Y_{2i}(u) = I(X_i^{age} \geq u \geq A)$. Let

$$M_{ki}(t) = N_{ki}(t) - \int_0^t Y_{ki}(u)exp(\beta_k^T W_i)d\Lambda_{0k}(u), k = 1, 2.$$

By the arguments similar to Andersen and Gill (1982) and Lin et al. (1994), for $t \in [0, \tau_1]$, $\tau_1 < \inf\{t : Pr(X_1 \geq t) = 0\}$, $n^{1/2}\{\hat{\Lambda}_1^{time}(t; w_0) - \Lambda_1^{time}(t; w_0)\}$ is asymptotically equivalent to

$$G_1(t; w_0) = n^{-1/2} \sum_{i=1}^n \left[\int_0^t \frac{exp(\beta_1^T w_0)}{S_1^{(0)}(\beta_1, u)} dM_{1i}(u) \right.$$
$$\left. + h_1^T(t; w_0)\Omega_1^{-1} \int_0^\infty \{W_i - \bar{W}_1(\beta_1, u)\}dM_{1i}(u) \right] \quad (4.2)$$

and similarly, for $a \in [a_0, \tau_2]$, $\tau_2 < \inf\{t : Pr(X_1^{age} \geq a \geq A_1) = 0\}$, $n^{1/2}[\{\hat{\Lambda}_2^{age}(a; w_0) - \hat{\Lambda}_2^{age}(a_0; w_0)\} - \{\Lambda_2^{age}(t; w_0) - \Lambda_2^{age}(a_0; w_0)\}]$ is asymptotically equivalent to

$$G_2(a; w_0) = n^{-1/2} \sum_{i=1}^{n} \Big[\int_{a_0}^{a} \frac{exp(\beta_2^T w_0)}{S_2^{(0)}(\beta_2, u)} dM_{2i}(u)$$

$$+ (h_2(a; w_0) - h_2(a_0; w_0)) \Omega_2^{-1} \int_{0}^{\infty} \{W_i - \bar{W}_2(\beta_2, u)\} dM_{2i}(u) \Big], \quad (4.3)$$

where $S_k^{(r)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} Y_{ki}(t) exp(\beta^T W_i) W_i^{\otimes r}$, $s_k^{(r)}(\beta, t) = E(S_k^r(\beta, t))$,

$$\bar{W}_k(\beta, t) = \frac{S_k^{(1)}(\beta, t)}{S_k^{(0)}(\beta, t)}, \bar{w}_k(\beta, t) = \frac{s_k^{(1)}(\beta, t)}{s_k^{(0)}(\beta, t)},$$

$$\Omega_k = \int_{0}^{\infty} \{ \frac{s_k^{(2)}(\beta_k, u)}{s_k^{(0)}(\beta_k, u)} - \bar{w}_k(\beta_k, u)^{\otimes 2} s_k^{(0)}(\beta_k, u) d\Lambda_{0k}(u) \},$$

$$h_k(t; z_0) = \int_{0}^{t} \{w_0 - \bar{w}_k(\beta_k, u)\} exp(\beta_k^T w_0) d\Lambda_{0k}(u),$$

for $r = 0, 1, 2$ and following the convention that for a column vector v, $v^{\otimes 0} = 1, v^{\otimes 1} = v$, and $v^{\otimes 2} = vv^T$.

Using the consistency of $\hat{\Lambda}_1^{time}(u; w_0)$, Taylor expansion and integration by parts as de-

scribed in Cheng et al (1998), (4.1) and (4.2), $U_1(t; w_0) = n^{1/2}\{\hat{F}_1(t; w_0) - F_1(t; w_0)\}$ is

$$U_1(t; w_0) = n^{1/2}\Big\{ \int_0^t \hat{S}_1(u; w_0) d\hat{\Lambda}_1^{time}(u; w_0) - \int_0^t S_1(u; w_0) d\Lambda_1^{time}(u; w_0)\Big\}$$

$$= n^{1/2} \int_0^t S_1(u; w_0) d\{\hat{\Lambda}_1^{time}(u; w_0) - \Lambda_1^{time}(u; w_0)\}$$

$$+ n^{1/2} \int_0^t \{\hat{S}_1(u; w_0) - S_1(u; w_0)\} d\hat{\Lambda}_1^{time}(u; w_0)$$

$$\approx \int_0^t S_1(u; w_0) dG_1(u; w_0) - \int_0^t S_1(u; w_0)\{G_1(u; w_0) + G_2(a_0 + u; w_0)\} d\Lambda_1^{time}(u; w_0)$$

$$= \int_0^t S_1(u; w_0) dG_1(u; w_0) - \int_0^t \{G_1(u; w_0) + G_2(a_0 + u; w_0)\} dF_1(u; w_0)$$

$$\int_0^t S_1(u; w_0) dG_1(u; w_0) - \int_0^t \{F_1(t; w_0) - F_1(u; w_0)\}\{dG_1(u; w_0) + dG_2(a; w_0)\}$$

$$= n^{-1/2} \sum_{i=1}^n I_{1i}(t; w_0), \quad (4.4)$$

where

$$I_{1i}(t; w_0) = \int_0^t \Big[S_1(u; w_0) - \{F_1(t; w_0) - F_1(u; w_0)\}\Big] \frac{exp(\beta_1^T w_0)}{S_1^{(0)}(\beta_1, u)} dM_{1i}(u)$$

$$+ \{\phi_1(t; w_0) - \psi_{11}(t; w_0)\}^T \Omega_1^{-1} \int_0^\infty \{W_i - \bar{W}_1(\beta_1, u)\} dM_{1i}(u)$$

$$- \int_0^t \{F_1(t; w_0) - F_1(u; w_0)\} \frac{exp(\beta_2^T w_0)}{S_2^{(0)}(\beta_2, a_0 + u)} dM_{2i}(a_0 + u)$$

$$- \psi_{12}^T(t; w_0) \Omega_2^{-1} \int_0^\infty \{W_i - \bar{W}_2(\beta_2, u)\} dM_{2i}(u),$$

$$S_1(u; w_0) = exp\Big[- \Lambda_1^{time}(u; w_0) - \{\Lambda_2^{age}(a_0 + u; w_0) - \Lambda_2^{age}(a_0; w_0)\}\Big],$$

$$\phi_1(t; w_0) = \int_0^t S_1(u; w_0)\{w_0 - \bar{w}_1(\beta_1, u)\}exp(\beta_1^T w_0)d\Lambda_{01}(u),$$

$$\psi_{11}(t; w_0) = \int_0^t \{F_1(t; w_0) - F(u; w_0)\}\{w_0 - \bar{w}_1(\beta_1, u)\}exp(\beta_1^T w_0)d\Lambda_{01}(u),$$

$$\text{and } \psi_{12}(t; w_0) = \int_0^t \{F_1(t; w_0) - F_1(u; w_0)\}\{w_0 - \bar{w}_2(\beta_2, a_0 + u)\}exp(\beta_2^T w_0)d\Lambda_{02}(a_0 + u).$$

$M_{ki}(t), k = 1, 2$ in (4.3) have different filtrations defined based on the chosen time scales and the martingale theory holds for each of them but with respect to different filtrations. Even though N_{1i} and N_{2i} cannot jump at the same time, $M_{ki}(t), k = 1, 2$ with different filtrations are not orthogonal martingales. $I_{1i}(t; w_0)$ in (4.3) contains $S_1^{(r)}$ and $S_2^{(r)}, r = 0, 1, 2$ which involve data on all n subjects. Hence, $I_{1i}(t; w_0)$ are not i.i.d. In this case, we cannot use the standard martingale theory (Andersen et al, 1993). However, using empirical processes arguments, we can show that

$$n^{-1/2} \sum_{i=1}^n I_{1i}(t; w_0) = n^{-1/2} \sum_{i=1}^n \tilde{I}_{1i}(t; w_0) + o_p(1),$$

where $\tilde{I}_{1i}(t; w_0)$ is $I_{1i}(t; w_0)$ with $S_1^{(r)}$ and $S_2^{(r)}$ replaced by respective theoretical quantities, $s_1^{(r)}$ and $s_2^{(r)}, r = (0, 1, 2)$. Thus, by the central limit theorem,

$$n^{-1/2} \sum_{i=1}^n \tilde{I}_{1i}(t; w_0) \to N(0, E\{\tilde{I}_{1i}(t; w_0)^2\}).$$

The variance $V_1(t; w_0) = E\{\tilde{I}_{1i}(t; w_0)^2\}$ is estimated by

$$\hat{V}_1(t; w_0) = n^{-1} \sum_{i=1}^n \hat{\tilde{I}}_{1i}(t; w_0)^2,$$

where $\hat{\tilde{I}}_{1i}(t; w_0)$ is $\tilde{I}_{1i}(t; w_0)$ with all theoretical quantities replaced by their empirical counterparts and $M_{ki}(t)$ replaced by $\hat{M}_{ki}(t) = N_{ki}(t) - \int_0^t Y_{ki}(u)exp(\hat{\beta}_k^T W_i)d\Lambda_{0k}(u)$ with $\hat{\Lambda}_{0k}(t)$ being the Breslow estimator for Λ_{0k}.

Using similar arguments, $U_2(a; w_0) = n^{1/2}\{\hat{F}_2(a; w_0) - F_2(a; w_0)\}$ can be expressed as

$$U_2(a; w_0) = n^{1/2}\left\{ \int_{a_0}^{a} \hat{S}_2(s; w_0)d\hat{\Lambda}_2^{age}(s; w_0) - \int_{a_0}^{a} S_2(s; w_0)d\Lambda_2^{age}(s; w_0) \right\}$$

$$= n^{1/2} \int_{a_0}^{a} S_2(s; w_0)d\{\hat{\Lambda}_2^{age}(s; w_0) - \Lambda_2^{age}(s; w_0)\}$$

$$+ n^{1/2} \int_{a_0}^{a} \{\hat{S}_2(s; w_0) - S_2(s; w_0)\}d\hat{\Lambda}_2^{age}(s; w_0)$$

$$\approx \int_{a_0}^{a} S_2(a; w_0)dG_3(s; w_0) - \int_{a_0}^{a} S_2(s; w_0)\{G_1(s - a_0; w_0) + G_3(s; w_0)\}d\Lambda_2^{age}(s; w_0)$$

$$= \int_{a_0}^{a} S_2(s; w_0)dG_3(s; w_0) - \int_{a_0}^{a} \{G_1(s - a_0; w_0) + G_3(s; w_0)\}dF_2(s; w_0)$$

$$\int_{a_0}^{a} S_2(s; w_0)dG_2(s; w_0) - \int_{a_0}^{a} \{F_2(a; w_0) - F_2(a; w_0)\}\{dG_1(s - a_0; w_0) + dG_3(s; w_0)\}$$

$$= n^{-1/2} \sum_{i=1}^{n} I_{2i}(a; w_0),$$

where

$$I_{2i}(t; w_0) = \int_{a_0}^{a} \left[S_2(s; w_0) - \{F_2(a; w_0) - F_2(s; w_0)\} \right] \frac{exp(\beta_2^T w_0)}{S_2^{(0)}(\beta_2, u)} dM_{2i}(u)$$

$$+ \{\phi_2(a; w_0) - \psi_{22}(a; w_0)\}^T \Omega_2^{-1} \int_0^{\infty} \{W_i - \bar{W}_2(\beta_2, u)\}dM_{2i}(u)$$

$$- \int_{a_0}^{a} \{F_2(a; w_0) - F_2(s; w_0)\} \frac{exp(\beta_1^T w_0)}{S_1^{(0)}(\beta_1, s - a_0)} dM_{1i}(s - a_0)$$

$$- \psi_{21}^T(a; w_0)\Omega_1^{-1} \int_0^{\infty} \{W_i - \bar{W}_1(\beta_1, u)\}dM_{1i}(u),$$

$$G_3(a; w_0) = n^{-1/2} \sum_{i=1}^{n} \Big[\int_{a_0}^{a} \frac{exp(\beta_2^T w_0)}{S_2^{(0)}(\beta_2, u)} dM_{2i}(u)$$

$$+ h_2(a; w_0) \Omega_2^{-1} \int_{0}^{\infty} \{W_i - \bar{W}_2(\beta_2, u)\} dM_{2i}(u) \Big],$$

$$S_2(s; w_0) = exp\Big[- \Lambda_1^{time}(s - a_0; w_0) - \Lambda_2^{age}(s; w_0) \Big],$$

$$\phi_2(a; w_0) = \int_{a_0}^{a} S_2(s; w_0)\{w_0 - \bar{w}_2(\beta_2, s)\} exp(\beta_2^T w_0) d\Lambda_{02}(s),$$

$$\psi_{22}(t; w_0) = \int_{a_0}^{a} \{F_2(a; w_0) - F_2(s; w_0)\}\{w_0 - (w)_2(\bar{\beta}_2, s)\} exp(\beta_1^T w_0) d\Lambda_{01}(s),$$

and $\psi_{21}(a; w_0) = \int_{a_0}^{a} \{F_2(a; w_0) - F_2(s; w_0)\}\{w_0 - \bar{w}_1(\beta_1, s - a_0)\} exp(\beta_1^T w_0) d\Lambda_{01}(s - a_0).$

Using empirical processes arguments, we can show that

$$n^{-1/2} \sum_{i=1}^{n} I_{2i}(a; w_0) = n^{-1/2} \sum_{i=1}^{n} \tilde{I}_{2k}(a; w_0) + o_p(1),$$

where $\tilde{I}_{2i}(a; w_0)$ is $I_{2i}(a; w_0)$ with $S_1^{(r)}$ and $S_2^{(r)}$ replaced by respective theoretical quantities, $s_1^{(r)}$ and $s_2^{(r)}$, $r = (0, 1, 2)$. Thus, by the central limit theorem,

$$n^{-1/2} \sum_{i=1}^{n} \tilde{I}_{2i}(a; w_0) \to N(0, E\{\tilde{I}_{2i}(a; w_0)^2\}).$$

The variance $V_2(t; w_0) = E\{\tilde{I}_{2i}(a; w_0)^2\}$ is estimated by

$$\hat{V}_2(a; w_0) = n^{-1} \sum_{i=1}^{n} \hat{\tilde{I}}_{2i}(a; w_0)^2,$$

where $\hat{\tilde{I}}_{2i}(a; w_0)$ is $\tilde{I}_{2i}(a; w_0)$ with all theoretical quantities replaced by their empirical counterparts and $M_{ki}(t)$ replaced by $\hat{M}_{ki}(t)$.

Chapter 5. Non-Parametric Estimation of Cumulative Incidence Functions for Competing Risks data with Missing Cause of Failure

5.1 Introduction

When studying time-to-event data with more than one possible cause of failure, specifically competing risks set-up, it is essential to be able to determine the cause of each observed failure in order to classify the events. However, in practice sometimes such classification is not possible, either at all or at the moment of data analysis, which results in possibly missing event type. The missingness mechanism can be classified in the usual fashion as missing completely at random (MCAR), missing at random (MAR), or missing not at random. In the non-parametric framework for competing risks setup, the MAR assumption corresponds to the missingness pattern of the event type possibly depending on the event time, but not depending on the event type, which constitutes the unobserved data in these settings. The MCAR assumption requires that the missingness pattern is also independent of event time (Laan and McKeague, 1998). As described earlier in Chapter 1, Section 1.6, treating observations with missing cause of failure as censored leads to underestimation of the hazard functions, even if the event type is missing completely at random.

Numerous methods to account for events with missing cause of failure have been suggested. In 1982 Dinse (Dinse, 1982) provided a classification of incomplete competing risks data, which along with the standard fully observed and fully censored observations allowed two kinds of partially complete observations, with either observed time of failure and missing type of failure, or with an observed type but censored time of failure. To analyse such data, he suggested a non-parametric maximum likelihood estimator (NPMLE) obtained by an EM algorithm, which reduced to a closed-form estimator in case the data contained no observations with observed failure type and censored failure time. Later he further developed his methods (Dinse, 1986) and considered cases of missingness patterns equivaluent both to MCAR and MAR. Miyakawa (Miyakawa, 1984) studied a fully parametric exponential model and a non-parametric model, both under the MCAR assumption for the missing type of failure.

A closely related problem of univariate time-to-event data without competing risks but with possibly missing at random censoring indicator was further studied by McKeague and Subramanian (McKeague and Subramanian, 1998), van der Laan and McKeague (Laan and McKeague, 1998), and Subramanian and Bean (Subramanian, 2006; Subramanian and Bean, 2008). Cook and Kosorok (Cook and Kosorok, 2004) considered a similar problem of time-to-event data with incomplete event adjudication.

One of the most cited in practice methods appears to be the methods by Lu and Tsiatis (Lu and Tsiatis, 2001), who suggested to combine semi-parametric inference with multiple imputations of the missing event type, based on a parametric model for the conditional probability of the event of interest given that an event of any type has occurred. The multiple imputations method was further developed by Lee et al for semi-parametric settings (Lee et al., 2011) and in non-parametric settings (Lee et al., 2013).

Effraimidis and Dalh (Effraimidis and Dahl, 2013) recently developed a non-parametric estimator for cumulative incidence functions with the inverse probability weighting scheme which utilizes estimators of the probability of the event type being observed smoothed over covariate values.

This chapter was motivated by the paper by Lin et al (Lin et al., 2013), who for the recurrent events set-up have proposed a non-parametric rate proportion estimator for mean event rates, which utilizes local polynomial regression over time to estimate probabilities of events of specific types and then plugs these event probability estimates into the expression for mean event rates. The advantage of this method is that it doesn't require any model specification for missingness pattern or event probabilities and works under the missing at random (MAR) assumptions. The idea is related to presmoothed Nelson-Aalen, Kaplan-Meier, and Aalen-Johansen estimators, described recently in literature (Cao et al., 2003; Jacome and Cao, 2007; El-Nouty and Lancar, 2005). Using such smoothing techniques for analysis of univariate survival data with censoring indicators missing at random have been then proposed by Subramanian and Bean, who developed estimators for the integrated hazard, the survival function, and the hazard rate function (Subramanian, 2006; Subramanian and Bean, 2008). Wang, Dinse and Liu (Wang et al., 2012) suggested to use the same idea for competing risks data with missing event type, however, they didn't go further than estimating the hazard rates $\lambda_j(t)$. The goal of this chapter is to fully extend the approach to the competing risks set-up and to propose estimators for cumulative cause-specific hazards and cumulative incidence functions. Deriving the estimators themselves for cumulative quantities is trivial and can be easily done by plugging in the known estimators for hazard rates into the expression for the cause-specific cumulative hazard. However, deriving the finite sample variance and the asypmtotic properties of the resulting estimators for cumulative hazards and cumulative incidence functions is a much more technically challenging and interesting part. Unlike the case of the hazard rates, the estimators of the cumulative quantities can be shown to be $\sqrt{n}$-consistent under some conditions, even though they utilize the local regression estimates which are known to have a slower than $\sqrt{n}$ convergence rate.

Despite the multidude of papers on the topic, it turned out to be hard to find a direct

comparator to the method which we are proposing in this paper for the competing risks setup. Most methods that are applied in practice either are semi-parametric and concerned with parameter estimates (Lu and Tsiatis, 2001; Gijbels et al., 2007; Cook and Kosorok, 2004), or estimate different quantities such as 1-KM (Racine-Poon and Hoel, 1984), or assume the probability of the event type being observed constant over time (Effraimidis and Dahl, 2013). The second (alternative) estimator for survival data with missing censoring indicator in van der Laan and McKeague's paper (Laan and McKeague, 1998) seems very comparable in the sense that it yeilds a non-parametric estimate of a subdistribution function. However, this estimator treats censoring as a competing risk and assumes no further independent censoring and hence can't be used in our settings without modification, doing which is beyond the scope of this chapter.

This chapter is organized as follows. Section 5.2 introduces the proposed estimator and establishes its asymptotic properties. Section 5.3 contains some simulations results, and Section 5.4 - real data analysis example.

5.2 Methods

Let us first introduce some notation. Let's assume without the loss of generality that we have $J = 2$ possible event types, indexed by $j = 1, 2$, and the event of interest is $j = 1$. The data which we observe for the $i - th$ patient, $i = 1, ..., n$, is $(X_i, \delta_i, R_i(X_i), \delta_i \epsilon_i)$, where X_i is an observed time, δ_i is the censoring indicator, $\epsilon_i \in \{1, 2\}$ is the event type which is allowed to be missing, and $R_i(X_i)$ is the indicator of whether the event type was observed or not (with $R_i(X_i) = 1$ if the event type is observed). Note that $R_i(X_i)$ is not observed when X_i is a censoring time. Let $\pi(t) = Pr(R_i(X_i) = 1 | X_i = t, \delta_i = 1)$. We'll use notation $R_i(X_i)$ to emphasize the fact that the probability of $R_i(X_i) = 1$ may depend on the event time X_i. Let $p_1(t) = Pr(\epsilon_i = 1 | X_i = t, \delta_i = 1)$ be the true conditional probability of event type being 1, given an event of any type was observed. Let $N_{ji}(t)$ be counting processes corresponding to events with observed event type $j = 1, 2$, $N_{ji}(t) = I(X_i \leq t, \delta_i = 1, R_i(X_i) = 1, \epsilon_i = j)$. Let $N_{0i}(t)$ be the overall counting process which has jumps when any of the events occurs, $N_{0i}(t) = I(X_i \leq t, \delta_i = 1)$. Note that $N_{ji}(t)$ only jump when an event with the observed event type occurs, while the overall process $N_{0i}(t)$ jumps even if the event type is missing. Let $Y_i(t) = I(X_i \geq t)$, $Y.(t) = \sum_{i=1}^{n} Y_i(t)$, and $y(t) = Pr(Y_1(t)) = 1)$.

Let $\Lambda_j(t)$ denote the true cause-specific cumulative hazard for cause $j = 1, 2$, and let

$\Lambda_0(t) = \sum\limits_{j=1}^{2} \Lambda_j(t)$ be the true overall cumulative hazard. Let's define

$$\hat{\Lambda}_1^{cc}(t) = \sum_{i=1}^{n} \int_0^t Y(u)^{-1} R_i(u) dN_{1i}(u) \text{ - complete case estimator for } \Lambda_1(t),$$

$$\hat{\Lambda}_0(t) = \sum_{i=1}^{n} \int_0^t Y(u)^{-1} dN_{0i}(u) \text{ - Nelson-Aalen estimator for } \Lambda_0(t),$$

$\hat{S}(t)$ - the Kaplan-Meier estimator of the overall survival function.

Note that all the quantities that we need to compute the last two estimators for the overall cumulative hazard and the overall survival function are fully observed and hence there is no need to account for missing event type to obtain them.

The idea of the proposed estimator for the cumulative incidence function is the following. Using local polynomial regression (Fan and Gijbels, 1996), we can obtain a non-parametric estimate $\hat{p}_1(u)$ of the conditional probability $p_1(u)$ of an event of type 1 occuring at time u, given an event of any type has occurred at time u. The estimate $\hat{p}_1(u)$ is obtained only from complete case observations, however, it will be asymptotically unbiased if the event type is missing at random, that is $\pi(u)$ depends on the event time but does not depend on the event type. By the Bayes rule and the MAR assumption we have

$$\begin{aligned}
p_1(u) = Pr(\epsilon_i = 1 | X_i = u, \delta_i = 1) &= \\
&= \frac{Pr(R_i(X_i) = 1 | X_i = u, \delta_i = 1) Pr(\epsilon_i = 1 | X_i = u, \delta_i = 1)}{Pr(R_i(X_i) = 1 | X_i = u, \delta_i = 1)} = \\
&= \frac{Pr(R_i(X_i) = 1 | X_i = u, \delta_i = 1, \epsilon_i = 1) Pr(\epsilon_i = 1 | X_i = u, \delta_i = 1)}{Pr(R_i(X_i) = 1 | X_i = u, \delta_i = 1)} = \\
&= \frac{Pr(R_i(X_i) = 1, \epsilon_i = 1 | X_i = u, \delta_i = 1)}{Pr(R_i(X_i) = 1 | X_i = u, \delta_i = 1)} = \\
&= Pr(\epsilon_i = 1 | X_i = u, \delta_i = 1, R_i(X_i) = 1),
\end{aligned}$$

where the latter can be correctly estimated from the complete case data only.

The local estimate of $p_1(t)$ for t in some neighborhood of t_0 can be obtained by regressing $\theta(t) = g(p_1(t))$ on $(t - t_0)^r$ where $g(.)$ is a link function and $r = 1, ..., R$ with R being the highest degree of the polynomials of time that we use. The idea comes from the fact that for $t \in (t_0 - b, t_0 + b)$ and for $b > 0$ - small, we can approximate $\theta(t)$ using Taylor series

expansion as

$$\theta(t) \approx \sum_{r=0}^{R} \frac{1}{r!} \theta^{(r)}(t_0)(t - t_0)^r = \sum_{r=0}^{R} \beta_r(t_0)(t - t_0)^r,$$

where the vector of regressions coefficients $\boldsymbol{\beta}(t_0)$ can be etimated by minimizing the local loglikelihood

$$l(\boldsymbol{\beta}(t_0)) = \sum_{i=1}^{n} \int_0^\tau K_b(u - t_0) l_i(u, t_0, \boldsymbol{\beta}(t_0))) R_i(u) dN_{0i}(u)$$

where $l_i(u, t_0, \boldsymbol{\beta}(t_0)))$ is the contribution to the loglikelihood from the $i-th$ observation derived in the standard fashion, $K_b(u - t_0) = \frac{1}{b} K(\frac{u-t_0}{b})$ is a kernel function, and $b = b_n$ - bandwidth. The estimate of $p_1(t_0)$ is then $g^{-1}(\hat{\theta}(t_0)) = g^{-1}(\hat{\beta}_0(t_0))$ where $\beta_0(t_0)$ is the local intercept. (See (Fan and Gijbels, 1996) for more detail).

In practice, polynomials of degrees higher than 0 and 1 are rarely used. From here on, we will limit ourselves to the local constant fit (that is, degree 0), known as Nadaraya-Watson estimator, due to the relative ease of proofs of the asymptotic properties of the resuting estimators for the cumulative cause-specific hazard and incidence function. The Nadaraya-Watson estimator has the following explicit form:

$$\hat{p}_1(u) = \frac{\frac{1}{n} \sum_{j=1}^{n} R_j(X_j) K_b(u - X_j) I(\epsilon_j = 1)}{\frac{1}{n} \sum_{j=1}^{n} R_j(X_j) K_b(u - X_j)}.$$

where $K_b(.)$ is the kernel function. Note that the summation is only over observations with non-missing event type, that is those with $R_i(X_i) = 1$.

The proposed estimator for integrated cause-specific hazard is then

$$\hat{\Lambda}_1(t, \hat{p}_1(t)) = \sum_{i=1}^{n} \int_0^t Y(u)^{-1} R_i(u) dN_{1i}(u) + \sum_{i=1}^{n} \int_0^t Y(u)^{-1} \big(1 - R_i(u)\big) \hat{p}_1(u) dN_{0i}(u) \quad (5.5)$$

In other words, events with missing event types contribute to both cause-specific hazards, and the contribution of events with missing event type is distributed between the two cause-specific hazards according to our estimate of the conditional probability of event of type 1 to occur.

This leads to the proposed estimator for cumulative incidence function:

$$\hat{F}_1(t) = \int\limits_0^t \hat{S}(u) d\hat{\Lambda}_1(u, \hat{p}_1(u)) =$$

$$= \sum_{i=1}^n \int\limits_0^t \hat{S}(u) Y(u)^{-1} \big[R_i(u) dN_{1i}(u) + \big(1 - R_i(u)\big) \hat{p}_1(u) dN_{0i}(u) \big].$$

To derive the asymptotic properties of the proposed estimators, we will follow the approach used by Cao et al in (Cao et al., 2003) for presmoothed Nelson-Aalen and Kaplan-Meier estimators. We will refer to Silverman's and Mack and Silverman's results for uniform convergence rates of the Nadaraya-Watson and kernel density estimators, namely, to Lemma 1 and Theorem B in (Mack and Silverman, 1982). We will need the following assumptions:

A1. $(X_i, \delta_i, R_i, \delta_i R_i \epsilon_i)$ are i.i.d.

A2. Censoring time is independent of failure time and failure type.

A3. The event type ϵ_i is missing at random, that is the indicator of the event type being observed $R_i(X_i)$ is independent of the event type ϵ_i.

A4. $t \in [0, \tau]$ where τ is chosen so that $y(\tau) = Pr(Y_1(\tau) = 1) > 0$.

A5. The density function $f_{cc}(t)$ for the distribution of event times with observed event type (complete cases) is bounded and bounded away from zero on $[0, \tau]$. This implies that both density of the event times $f(u)$ and the conditional probability $\pi(t) = Pr(R_i(X_i) = 1 | X_i = t, \delta_i = 1)$ of the event type being observed are bounded away from zero on $[0, \tau]$.

A6. Densities $f_{cc}(u)$, $f(u)$, the density $h(u)$ of $X_1 = min(T_1, C_1)$ and the probability $p_1(u)$ are twice differentiable with bounded second derivatives.

A7. The kernel $K(.)$ is a symmetric function of bounded variation, twice differentiable with the bounded second derivative. It has support on some interval $[-L, L]$ for $0 < L < \infty$. It satisfies $\int\limits_{-L}^L K(s) ds = 1$ and $\int\limits_{-L}^L s K(s) ds = 0$.

A8. The bandwidth sequence $b_n = Cn^{-\alpha} + o_p(n^{-\alpha})$, for some $C > 0$, with $1/4 < \alpha < 1/2$.

Theorem 1. Under the above assumptions, the estimator for cumulative cause-specific hazard $\hat{\Lambda}_1(t, \hat{p}_1(t))$ defined by (5.1) is uniformly consistent on $[0, \tau]$ and asymptotically normal for $t \in [0, \tau]$ with

$$\sqrt{n}\big(\hat{\Lambda}_1(t, \hat{p}_1(t)) - \Lambda_1(t)\big) \xrightarrow{d} N(0, E(G_1^2(t)),$$

where

$$G_1(t) =$$

$$= \int_0^t \big(1 - R_i(u)\big) y(u)^{-1} [p_1(u) dN_{0i}(u) - Y_1(u) d\Lambda_1(u)] +$$

$$+ \int_0^t R_1(u) y(u)^{-1} [dN_{1i}(u) - Y_1(u)\Lambda_1(u)] +$$

$$+ I(X_1 \le t) R_1(X_1) f_{cc}(X_1)^{-1} [I(\epsilon_1 = 1) - p_1(X_1)]\big(1 - \pi(X_1)\big)\lambda_0(X_1).$$

where $f_{cc}(u)$ is the density function for the times of complete case observations.

Theorem 2. Under the same assumptions as Theorem 1, the estimator for cumulative incidence function $\hat{F}_1(t)$ is uniformly consistent on $[0, \tau]$ and asymptotically normal for $t \in [0, \tau]$ with

$$\sqrt{n}\big(\hat{F}_1(t) - F_1(t)\big) \xrightarrow{d} N(0, E(I_1^2(t)),$$

where

$$I_1(t) =$$

$$= \int_0^t \big(1 - R_1(u)\big) \frac{S(u)}{y(u)} [p_1(u) dN_{0i}(u) - Y_1(u) d\Lambda_1(u)] +$$

$$+ \int_0^t R_1(u) \frac{S(u)}{y(u)} [dN_{1i}(u) - Y_1(u) d\Lambda_1(u)] +$$

$$+ \int_0^t R_1(u) \frac{S(u)}{f_{cc}(u)} K_b(u - X_1)[I(\epsilon_1 = 1) - p_1(u)]\big(1 - \pi(u)\big) d\Lambda_0(u) +$$

$$+ \int_0^t y(u)^{-1} \big\{ F_1(t) - F_1(u) \big\} [dN_{0i}(u) - Y_1(u) d\Lambda_0(u)].$$

For practical purposes, the estimates of $f_{cc}(u)$ and $\pi(u)$ can be obtained as kernel estimates in a similar fashion to the way we estimate $p_1(u)$.

The proofs of both Theorems are provided in the Appendix 1.

Note that finite sample variance estimators for the cumulative cause-specific hazard and cumulative incidence function can be obtained from $\frac{1}{n^2}\sum_{j=1}^{n}\tilde{G}_{i,n}(t)$ and $\frac{1}{n^2}\sum_{j=1}^{n}\tilde{I}_{i,n}(t)$ respectively by plugging in the empirical estimates instead of theoretical quantities into

$$\tilde{G}_{i,n}(t) =$$

$$= \int_0^t \big(1 - R_i(u)\big)y(u)^{-1}[p_1(u)dN_{0i}(u) - Y_i(u)d\Lambda_1(u)]+$$

$$+ \int_0^t R_i(u)y(u)^{-1}[dN_{1i}(u) - Y_i(u)\Lambda_1(u)]+$$

$$+ \int_0^t R_i(u)f_{cc}(u)^{-1}K_b(u - X_i)[I(\epsilon_i = 1) - p_1(u)]\big(1 - \pi(u)\big)d\Lambda_0(u)$$

and

$$\tilde{I}_{i,n}(t) =$$

$$= \int_0^t \big(1 - R_i(u)\big)\frac{S(u)}{y(u)}[p_1(u)dN_{0i}(u) - Y_i(u)d\Lambda_1(u)]+$$

$$+ \int_0^t R_i(u)\frac{S(u)}{y(u)}[dN_{1i}(u) - Y_i(u)d\Lambda_1(u)]+$$

$$+ \int_0^t R_i(u)\frac{S(u)}{f_{cc}(u)}K_b(u - X_i)[I(\epsilon_i = 1) - p_1(u)]\big(1 - \pi(u)\big)d\Lambda_0(u)+$$

$$+ \int_0^t y(u)^{-1}\big\{F_1(t) - F_1(u)\big\}\big[dN_{0i}(u) - Y_i(u)d\Lambda_0(u)\big].$$

The derivation is a part of the proof and is provided in the Appendix 1. These estimators were used in the simulations in tables 5.1-5.4.

5.3   Simulation Results

Table 5.8: Comparison of the proposed estimator of the cumulative incidence function with the complete case estimator and the hypothetical full data estimator. Scenario 1 with time varying conditional probability of event 1.

		Method											
		Full data				Proposed				Complete case			
Time	True	Bias	Emp.SE	$\widehat{SE}$	CP	Bias	Emp.SE	$\widehat{SE}$	CP	Bias	Emp.SE	$\widehat{SE}$	CP
						Sample size = 200							
0.10	0.095	0.0007	0.02120	0.02082	0.95	−0.0004	0.02131	0.02088	0.94	−0.0213	0.01917	0.01856	0.81
1.00	0.507	0.0012	0.03549	0.03774	0.96	−0.0026	0.04023	0.04245	0.97	−0.0669	0.04142	0.04169	0.61
						Sample size = 500							
0.10	0.095	0.0002	0.01268	0.01318	0.96	−0.0002	0.01289	0.01334	0.95	−0.0218	0.01091	0.01176	0.58
1.00	0.507	−0.0006	0.02368	0.02381	0.94	−0.0018	0.02591	0.02702	0.95	−0.0685	0.02552	0.02630	0.25
						Sample size = 1,000							
0.10	0.095	−0.0001	0.00924	0.00931	0.96	−0.0005	0.00948	0.00947	0.95	−0.0220	0.00840	0.00831	0.31
1.00	0.507	−0.0003	0.01684	0.01681	0.94	−0.0013	0.01874	0.01920	0.96	−0.0687	0.01819	0.01857	0.03
						Sample size = 2,000							
0.10	0.095	−0.0002	0.00668	0.00659	0.95	−0.0005	0.00678	0.00671	0.94	−0.0221	0.00593	0.00588	0.05
1.00	0.507	−0.0006	0.01203	0.01188	0.95	−0.0013	0.01330	0.01360	0.96	−0.0688	0.01327	0.01311	0.00
						Sample size = 4,000							
0.10	0.095	−0.0003	0.00464	0.00465	0.96	−0.0005	0.00476	0.00475	0.96	−0.0222	0.00409	0.00416	0.00
1.00	0.507	0.0000	0.00822	0.00840	0.95	0.0000	0.00910	0.00964	0.96	−0.0676	0.00918	0.00928	0.00

Table 5.9: Comparison of the proposed estimator of the cumulative incidence function with the complete case estimator and the hypothetical full data estimator. Scenario 2 with constant conditional probability of event 1.

		Method											
		Full data				Proposed				Complete case			
Time	True	Bias	Emp.SE	$\widehat{SE}$	CP	Bias	Emp.SE	$\widehat{SE}$	CP	Bias	Emp.SE	$\widehat{SE}$	CP
						Sample size = 200							
0.10	0.086	0.0006	0.01971	0.01995	0.96	0.0008	0.02206	0.02323	0.96	−0.0179	0.01834	0.01815	0.84
1.00	0.317	0.0007	0.03486	0.03451	0.94	−0.0019	0.04005	0.04121	0.95	−0.0176	0.04045	0.03969	0.92
						Sample size = 500							
0.10	0.086	−0.0001	0.01204	0.01261	0.97	0.0003	0.01320	0.01498	0.97	−0.0182	0.01120	0.01152	0.71
1.00	0.317	0.0006	0.02065	0.02174	0.96	0.0002	0.02395	0.02590	0.96	−0.0173	0.02425	0.02498	0.90
						Sample size = 1,000							
0.10	0.086	0.0002	0.00869	0.00894	0.97	0.0001	0.00939	0.01073	0.97	−0.0185	0.00790	0.00814	0.41
1.00	0.317	−0.0005	0.01568	0.01533	0.94	−0.0015	0.01822	0.01848	0.97	−0.0199	0.01852	0.01755	0.77
						Sample size = 2,000							
0.10	0.086	0.0002	0.00660	0.00632	0.94	0.0001	0.00724	0.00765	0.96	−0.0186	0.00590	0.00575	0.14
1.00	0.317	0.0003	0.01101	0.01084	0.94	−0.0003	0.01242	0.01315	0.96	−0.0187	0.01232	0.01241	0.68
						Sample size = 4,000							
0.10	0.086	−0.0001	0.00430	0.00447	0.95	−0.0001	0.00480	0.00543	0.97	−0.0186	0.00404	0.00407	0.01
1.00	0.317	−0.0001	0.00783	0.00766	0.95	−0.0005	0.00874	0.00934	0.96	−0.0188	0.00874	0.00877	0.43

Table 5.10: Simulations results comparing different variance estimators for the proposed method, for Scenario 1.

					Variance estimation method							
					Infl. function		Naive		Bias corr.		Local Var corr.	
		Point Estimate										
Sample size	Time	True	Bias	Emp.SE	$\widehat{SE}$	CP	$\widehat{SE}$	CP	$\widehat{SE}$	CP	$\widehat{SE}$	CP
200	0.10	0.095	−0.0004	0.02131	0.02088	0.94	0.02029	0.93	0.02038	0.94	0.02059	0.94
	0.50	0.367	−0.0021	0.03686	0.03802	0.95	0.03308	0.92	0.03313	0.92	0.03608	0.95
	1.00	0.507	−0.0026	0.04023	0.04245	0.97	0.03379	0.91	0.03384	0.91	0.04169	0.96
500	0.10	0.095	−0.0002	0.01289	0.01334	0.95	0.01295	0.94	0.01296	0.94	0.01310	0.95
	0.50	0.367	−0.0012	0.02428	0.02418	0.95	0.02098	0.92	0.02099	0.92	0.02275	0.93
	1.00	0.507	−0.0018	0.02591	0.02702	0.95	0.02142	0.90	0.02142	0.90	0.02630	0.95
1000	0.10	0.095	−0.0005	0.00948	0.00947	0.95	0.00916	0.94	0.00917	0.95	0.00925	0.95
	0.50	0.367	−0.0009	0.01736	0.01714	0.95	0.01485	0.91	0.01486	0.91	0.01606	0.93
	1.00	0.507	−0.0013	0.01874	0.01920	0.96	0.01514	0.88	0.01515	0.88	0.01856	0.94
2000	0.10	0.095	−0.0005	0.00678	0.00671	0.94	0.00649	0.93	0.00649	0.93	0.00654	0.94
	0.50	0.367	−0.0003	0.01186	0.01214	0.95	0.01051	0.91	0.01051	0.91	0.01134	0.93
	1.00	0.507	−0.0013	0.01330	0.01360	0.96	0.01072	0.88	0.01072	0.88	0.01308	0.95
4000	0.10	0.095	−0.0005	0.00476	0.00475	0.96	0.00459	0.95	0.00459	0.95	0.00463	0.95
	0.50	0.367	0.0000	0.00854	0.00859	0.95	0.00744	0.92	0.00744	0.92	0.00802	0.93
	1.00	0.507	0.0000	0.00910	0.00964	0.96	0.00758	0.90	0.00758	0.90	0.00925	0.95

Table 5.11: Simulations results comparing different variance estimators for the proposed method, for Scenario 2.

| | | Point Estimate | | | Variance estimation method | | | | | | | | |
| | | | | | Infl. function | | Naive | | Bias corr. | | Local Var corr. | |
Sample size	Time	True	Bias	Emp.SE	$\widehat{SE}$	CP	$\widehat{SE}$	CP	$\widehat{SE}$	CP	$\widehat{SE}$	CP
200	0.10	0.086	0.0008	0.02206	0.02323	0.96	0.01804	0.90	0.01815	0.90	0.01885	0.92
	0.50	0.259	−0.0010	0.03695	0.03737	0.95	0.02797	0.86	0.02805	0.86	0.03197	0.90
	1.00	0.317	−0.0019	0.04005	0.04121	0.95	0.02963	0.85	0.02973	0.85	0.03611	0.92
500	0.10	0.086	0.0003	0.01320	0.01498	0.97	0.01147	0.93	0.01149	0.93	0.01185	0.93
	0.50	0.259	0.0005	0.02122	0.02402	0.97	0.01777	0.90	0.01779	0.90	0.02008	0.93
	1.00	0.317	0.0002	0.02395	0.02590	0.96	0.01885	0.87	0.01887	0.87	0.02276	0.93
1000	0.10	0.086	0.0001	0.00939	0.01073	0.97	0.00811	0.91	0.00811	0.91	0.00835	0.92
	0.50	0.259	−0.0008	0.01649	0.01712	0.96	0.01255	0.86	0.01255	0.86	0.01413	0.92
	1.00	0.317	−0.0015	0.01822	0.01848	0.97	0.01331	0.85	0.01332	0.85	0.01604	0.92
2000	0.10	0.086	0.0001	0.00724	0.00765	0.96	0.00573	0.88	0.00574	0.87	0.00589	0.89
	0.50	0.259	0.0003	0.01140	0.01218	0.96	0.00889	0.87	0.00889	0.87	0.00997	0.90
	1.00	0.317	−0.0003	0.01242	0.01315	0.96	0.00944	0.86	0.00944	0.86	0.01132	0.92
4000	0.10	0.086	−0.0001	0.00480	0.00543	0.97	0.00405	0.89	0.00406	0.89	0.00416	0.91
	0.50	0.259	−0.0004	0.00795	0.00864	0.97	0.00628	0.87	0.00628	0.87	0.00704	0.91
	1.00	0.317	−0.0005	0.00874	0.00934	0.96	0.00667	0.88	0.00668	0.88	0.00800	0.93

To assess the performance of the proposed estimator $\hat{F}_1(t)$ for the cumulative incidence function, we conducted a simulation study, the results of which are shown in Tables 5.1-5.4. We compared the proposed estimator with a hypothetical case of full data, when the event types for all observations are known (unobservable in reality), and with the complete case estimator. We generated competing risks data with 2 causes of failure. Event of type 1 was the event of interest. Censoring was generated as a uniform random variable and was about 10% in both scenarios. The indicator $R_i(X_i)$ of the event type being observed was generated as a Bernoulli random variable with a time-dependent probability of sucess $\frac{exp(1-0.1*t+0.3Z_i)}{1+exp(1-0.1*t+0.3Z_i)}$ where $Z_i \sim Bernoulli(0.5)$, which corresponded to the MAR assumption. Overall, about 25% of events had missing event type. We used the proposed method with the local constant fit (polynomials of degree zero) and Epanechnikov kernel. In the results shown below, we forced the bandwidth sequence go to zero at the rate of $n^{-0.3}$. We also did smaller simulation studies with bandwidth convergence rate ranging from -0.25 to -0.5, the results were very similar (not shown). We computed estimates at several time points, at an early time, in the middle, and in the tail of the distribution. We assessed the bias of the point estimate of $F_1(t)$, and its empirical standard error, asymptotic standard error, and coverage probability (using the log-log transformation).

We generated data from a range of scenarios, of which we here report two. Under scenario 1, the conditional probability of event 1 varied over time. The cause-specific hazards were $\lambda_1(t) = 1$ for the event 1, and $\lambda_2(t) = 2t$ for the event 2. Under scenario 2 both cause-specific hazards were constant over time, $\lambda_1(t) = 1$ and $\lambda_2(t) = 2$, with the constant conditional probability of event 1.

The simulation results show that the bias of the proposed estimator of $F_1(t)$ is very small, on the same scale as the bias observed in the hypothetical case of having the full data. The performance of the variance estimator was also very good. It slightly overestimated the empirical variance, which some cases resulted in coverage probabilily of about 96-97%.

Since finding the correct variance estimator was the main difficulty of this research and since we considered several potential variance estimators along the way, we did a more detailed comparison of all the variance estimators which we considered. The description of the rejected candidate variance estimators is provided in the Appendix 2. The results are shown in Tables 5.3-5.4. We used the same scenarios 1 and 2 as in Tables 5.1 and 5.2. We compared the variance estimator obtained from the influence function with the naive variance estimator, the estimator with the bias correction, with the local variance correction, and also the bootstrap variance estimator (not shown in tables). The bootstrap estimator as well as the naive variance estimator both significantly underestimated the variance. The estimate

with the bias correction was very slightly larger but didn't correct the underestimation too well. The estimator with the local variance correction was much closer to the truth. However, all of the alternative estimators noticeably underestimated the true variance and lead to coverage probabilities less than nominal. This confirmed that the variability of the local estimates could not be ignored and deriving the influence function and a variance estimator based on it were necessary.

5.4 Real Data Example

To illustrate the proposed method with a real data example, we used data from the paper which motivated this research (Lin et al., 2013), also previously described by Lai et al (Lai et al., 2004). This is patient data from the United States Cyctic Fibrosis Foundation Patient Registry which contains information on *P.aeruginosa* infections for 6,823 patients, collected between 1997 and 2007. The follow-up period ranged from 4 to 3,626 days, with the average follow-up length of 997 days per patient. The recorded information for each patient contained the type of infection: non-mucoid, mucoid, or both. In a number of cases, the type of infection was not known, even though the fact of infection was determined, which resulted in events with missing events category.

The original data contains information on recurrent infections and has multiple observations per patient. We analyzed time to the first infection for each patient, with the infection type (non-mucoid, mucoid, or both) being the event type. Out of 6,823 patients, a non-mucoidinfection occured first for 3,106 (45.5%) patients, mucoid infection for 263 (3.9%) patients, infections of both types were present for 193 (2.9%) patients, and for 563 (8.3%) patients the infection type was missing. Large proportion of patients were censored prior to having any infection (2,668 or 39.1%). There were also 30 patients in the data set (0.4%) who were deceased prior to having any infections. Even though technically death without an infection constitutes a competing risk in these settings and ideally should be analyzed as such, for our analysis we treated death same as censoring and did not view it as the 4th type of event. Due to a very small number of such patients, we felt that combining death prior to infection with censoring will have no effect on the analysis results for outcomes of interest, which are in this case the infection types.

To better understand the data, we plotted, as functions of time, non-parametric estimates of the probability of the infection type to be observed and the probabilities of developing an infection of a specific type, given the fact of infection. The estimates were obtained using local polynomial regression, alongside our main analyses which will be described below. The plots are shown on the Figure 5.1. All probabilities varied somewhat over time, which made
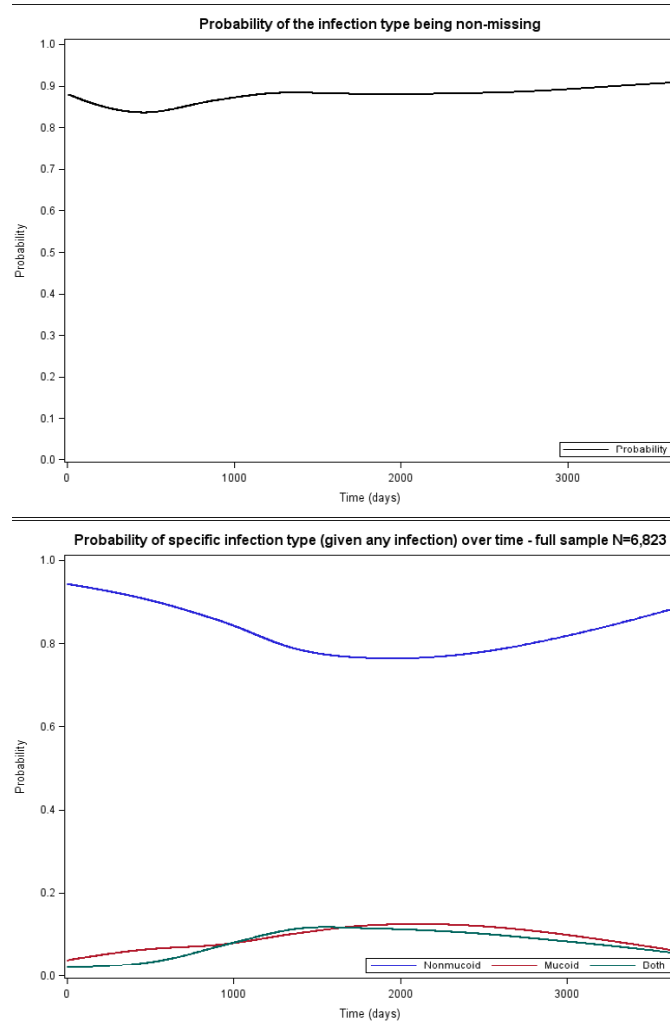
Figure 5.9: Top panel: Probability of the infection type being observed. Bottom panel: Conditional probabilities of the first infection to have a specific type, given an infection of some type has occured.

the proposed method more preferable than methods which assume either the probability of an event type being observed, or the probabilities of specific event types to be constant over time.

We analyzed the data using the proposed method. We used complete-case data to obtain local polynomial regression estimates of the conditional probabilities of an infection of a specific type to occur given the fact of infection. We fitted local logistic model using first degree polynomials of time as a predictor, with Epanechnikov kernel. The bandwidth was selected using the AIC criteria. Then we used the estimates of these conditional probabilities at the event times with missing event type to evaluate the contribution of the events with missing type to each cause-specific hazard. We computed the cumulative incidence functions for each of the three event types, along with their 95% confidence intervals. For comparison, we performed the complete case analysis treating the events with missing event type as censored. All analysis was done using R packages *locfit* and *cmprsk*, and SAS 9.3 software (SAS Institute, Cary, NC). The results are plotted on the Figure 5.2.

The plots show that after a period of 10 years over 70% of patients will develop a non-mucoid infection. The other two infection types are much more rare: the respective probabilities of developing a mucoid infection is about 6% and an infection of both types - about 5%. As one would expect, the complete case under-estimates the cumulative incidence functions for all event types. The difference is very small and for mucoid infections and infections of both types. For these infection types, both complete case and proposed method estimates clearly lie within each other's confidence limits. For the non-mucoid infection type, the estimate from the proposed method being approximately the upper limit of the 95% confidence interval of the complete case estimate. The results are consistent with the findings for the recurrent events from the motivating Lin et al 2013 paper.

5.5 Discussion

The proposed method is easy to implement in practice, has a plug-in variance estimator can be used in the wide range of settings, and despite a large amount of publications on the topic does not have an immediate comparator, to the author's best knowledge.

The idea of this estimator is related to presmoothed Nelson-Aalen and Kaplan-Meier estimators. As mentioned in the Introduction, the use of smoothing techniques to account for missing event types or censoring indicators in time-to-event data was suggested more than once in literature. However, to the author's best knowledge for the competing risks setup with missing at random cause of failure these suggestions stopped at deriving the estimators of the hazard rate functions, $\lambda_j(t)$ (Wang et al., 2012). Somewhat different assumptions than
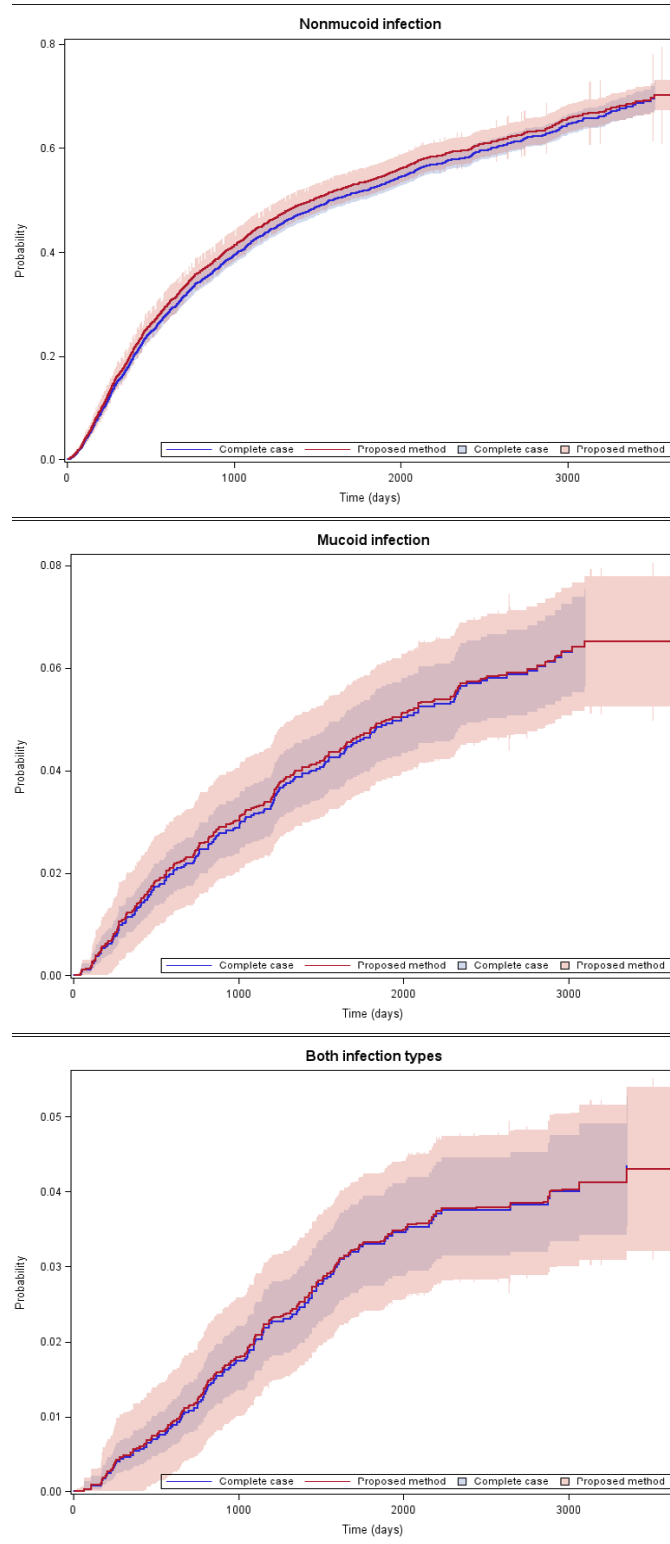
Figure 5.10: Cumulative incidence functions and 95% CI by infection type, obtained by the proposed method (red) compared to complete case analysis (blue). Top panel: non-mucoid infections. Center panel: mucoid infection. Bottom panel: both infection types.

76

used by Wang et al for hazard rates are required to show $\sqrt{n}$-consistency of presmoothed estimators of the integrated hazards and cumulative incidence functions, and the asymptotic representation of such cumulative estimators has a more complicated form.

All of the authors cited earlier in this chapter suggested fully presmoothed estimators, that is, when the estimates of the conditional probability $p_1(u)$ are used at all observed event time points, not only at the time points with missing event type. The fully presmoothed version of the proposed estimator and its variance should be easily obtained using the same technique as in this paper. From the literature on the presmoothed Nelson-Aalen and Kaplan-Meier estimators (Cao et al., 2003; Jacome and Cao, 2007), it may be beneficial in terms of reducing the MSE, especially in the tail. It will be interesting to compare the fully presmoothed version of the proposed estimator with the current one.

Another direction related to that, in which more work can be done, is bandwith selection. There exists extensive literature on the topic which should be applicable to the proposed estimator.

We used the local constant fit estimator to obtain $\hat{p}_1(u)$. Other practical candidates are the local linear estimator and local logistic regression. The former has better performance than the local constant fit in terms of both bias and variance, especially the bias on the boundaries. However, it has a disadvantage of being able to produce out-of-range values for predicted probability, due to having to built-in constraints for the estimate to stay between 0 and 1. Based on the literature, local logistic regression would be the best option, however, the theoretical derivations and proofs may be much more tedious. From the literature (Jacome et al., 2008), all three estimators should be asymptotically equivalent as $b \to 0$. This is yet another direction in which this work can be extended.

The assumptions that we used in this chapter are stronger than those suggested in the motivating paper (Lin et al., 2013), specifically in the part which prescribes the convergence rate for the bandwidth sequence. The proof which is provided in this chapter relies on these assumptions. Note that the required range for α, $1/4 < \alpha < 1/2$ does not include the "optimal" rate of $n^{-1/5}$. It would be interesting to see if the proof can be modified so that the assumptions can be weakened to match those in Lin et al. To the present moment, the author didn't see assumptions much weaker than A1-A8 in the literature.

5.6   Appendix 1

Proof of Theorem 1.

Consistency. Let's re-write

$$\hat{\Lambda}_1(t, \hat{p}_1(t)) - \Lambda_1(t) =$$
$$= \hat{\Lambda}_1(t, \hat{p}_1(t)) - \Lambda_1(t) + \hat{\Lambda}_1(t, p_1(t)) - \hat{\Lambda}_1(t, p_1(t)) + \hat{\Lambda}_1^{full}(t) - \hat{\Lambda}_1^{full}(t)$$

where $\hat{\Lambda}_1(t, p_1(t))$ denotes the result of substituting the true value $p_1(u)$ instead of its esti-mate into the expression (5.1) for the cumulative cause-specific hazard, and $\hat{\Lambda}_1^{full}(t)$ is the Nelson-Aalen estimator of the cumulative hazard for event 1 obtained from the hypothetical unobservable full data. Plugging in explicit expressions for all the estimators and removing the terms that cancel out, we have

$$\hat{\Lambda}_1(t, \hat{p}_1(t)) - \Lambda_1(t) =$$
$$\sum_{i=1}^{n} \int_0^t \left(1 - R_i(u)\right) Y^{-1}(u) [\hat{p}_1(u) - p_1(u)] dN_{0i}(u) +$$
$$\sum_{i=1}^{n} \int_0^t \left(1 - R_i(u)\right) Y^{-1}(u) [p_1(u) - I(\epsilon_i = 1)] dN_{0i}(u) +$$
$$+ \left[\sum_{i=1}^{n} \int_0^t Y^{-1}(u) I(\epsilon_i = 1) dN_{0i}(u) - \Lambda_1(t) \right]. \quad (5.6)$$

In the first term of (5.2), the part $[\hat{p}_1(u) - p_1(u)]$ converges to zero uniformly in u under our assumptions, by the Theorem B from Mack and Silverman (Mack and Silverman, 1982), which combined with consistency of the Nelson-Aalen estimator on $[0, \tau]$ implies convergence to zero of the first term. The third term also converges to zero due to the uniform consistency of the Nelson-Aalen estimator.

To show consistency of the remaining second term, let's consider counting processes $N_{1i}^{miss}(u) = I(X_i \le u, \delta_i = 1, \epsilon_i = 1, \text{ and } R_i(X_i) = 0)$ (this process is unobservable) and $N_{0i}^{miss}(u) = I(X_i \le u, \delta_i = 1, \text{ and } R_i(X_i) = 0)$. Note that

$$dN_{1i}^{miss}(u) = \left(1 - R_i(u)\right) I(\epsilon_i = 1) dN_{1i}(u)$$

and

$$dN_{0i}^{miss}(u) = \left(1 - R_i(u)\right) dN_{0i}(u).$$

Let's denote by $\Lambda_1^{miss}(u)$ and $\Lambda_0^{miss}(u)$ the true cumulative hazards corresponding to these

78

processes. Then under the MAR assumption and by the definition of $p_1(u)$ and $\pi(u)$ it is easy to show that

$$d\Lambda_1^{miss}(u) = \big(1 - \pi(u)\big)p_1(u)d\Lambda_0(u) = p_1(u)d\Lambda_0^{miss}(u).$$

Now the second term in (5.2) can be re-written as

$$\sum_{i=1}^{n} \int_0^t \big(1 - R_i(u)\big)Y^{-1}(u)[p_1(u) - I(\epsilon_i = 1)]dN_{0i}(u) =$$

$$= \sum_{i=1}^{n} \int_0^t Y^{-1}(u)[p_1(u)dN_{0i}^{miss}(u) - dN_{1i}^{miss}(u)] =$$

$$= \sum_{i=1}^{n} \int_0^t Y^{-1}(u)[p_1(u)dN_{0i}^{miss}(u) - Y_i(u)d\Lambda_1^{miss}(u) + Y_i(u)d\Lambda_1^{miss}(u) - dN_{1i}^{miss}(u)] =$$

$$= \sum_{i=1}^{n} \int_0^t Y^{-1}(u)[p_1(u)dN_{0i}^{miss}(u) - Y_i(u)p_1(u)d\Lambda_0^{miss}(u) + Y_i(u)d\Lambda_1^{miss}(u) - dN_{1i}^{miss}(u)] =$$

$$= \sum_{i=1}^{n} \int_0^t Y^{-1}(u)p_1(u)[dN_{0i}^{miss}(u) - Y_i(u)d\Lambda_0^{miss}(u)] +$$

$$+ \sum_{i=1}^{n} \int_0^t Y^{-1}(u)[Y_i(u)d\Lambda_1^{miss}(u) - dN_{1i}^{miss}(u)] =$$

$$= \int_0^t p_1(u)[d\hat{\Lambda}_0^{miss}(u) - d\Lambda_0^{miss}(u)] -$$

$$- [\hat{\Lambda}_1^{miss}(t) - \Lambda_1^{miss}(t)] =$$

$$= p_1(t)[\hat{\Lambda}_0^{miss}(t) - \Lambda_0^{miss}(t)] - \int_0^t p_1'(u)[\hat{\Lambda}_0^{miss}(u) - \Lambda_0^{miss}(u)]du -$$

$$- [\hat{\Lambda}_1^{miss}(t) - \Lambda_1^{miss}(t)],$$

$$(5.7)$$

where $\hat{\Lambda}_1^{miss}(u)$ and $\hat{\Lambda}_0^{miss}(u)$ denote the Nelson-Aalen estimators for the corresponding hazards. Note that $p_1''(u)$ exsists and is bounded on $[0, \tau]$ by the assumptions. Hence $p_1'(u)$ is uniformly bounded on $[0, \tau]$ which combined with the uniform consistency of the Nelson-

Aalen estimator implies that the whole expression (5.3) converges to zero for $t \in [0, \tau]$. This concludes the consistency proof. Note that this proof also implies uniform consistency of $\hat{\Lambda}_1(t, \hat{p}_1(t))$ on $[0, \tau]$, since each term in (5.2) and (5.3) is uniformly consistent.

Asymptotic representation.

$$
\hat{\Lambda}_1(t, \hat{p}_1(t)) - \Lambda_1(t) = \int_0^t d\hat{\Lambda}_1(u, \hat{p}_1(u)) - \int_0^t d\Lambda_1(u) =
$$

$$
= \sum_{i=1}^n \int_0^t Y(u)^{-1} R_i(u) dN_{1i}(u) - \int_0^t \pi(u) d\Lambda_1(u) +
$$

$$
+ \sum_{i=1}^n \int_0^t Y(u)^{-1} (1 - R_i(u)) \hat{p}_1(u) dN_{0i}(u) - \int_0^t (1 - \pi(u)) d\Lambda_1(t) =
$$

$$
= \sum_{i=1}^n \int_0^t Y(u)^{-1} \big[R_i(u) dN_{1i}(u) - Y_i(u) \pi(u) d\Lambda_1(u) \big] +
$$

$$
+ \sum_{i=1}^n \int_0^t Y(u)^{-1} \big[(1 - R_i(u)) \hat{p}_1(u) dN_{0i}(u) - Y_i(u) (1 - \pi(u)) d\Lambda_1(u) \big] =
$$

$$
= \frac{1}{n} \sum_{i=1}^n \int_0^t y(u)^{-1} \big[R_i(u) dN_{1i}(u) - Y_i(u) \pi(u) d\Lambda_1(u) \big] +
$$

$$
+ \frac{1}{n} \sum_{i=1}^n \int_0^t y(u)^{-1} \big[(1 - R_i(u)) \hat{p}_1(u) dN_{0i}(u) - Y_i(u) (1 - \pi(u)) d\Lambda_1(u) \big] =
$$

$$
+ o_p(n^{-1/2}).
$$

(5.8)

The first term in (5.4) already has the form of the sum of the i.i.d random variables. It is easy to see that they are zero mean, if we notice that $R_i(u) dN_{1i}(u) = dN_{1i}^{obs}(u)$ is the counting process with the corresponding hazard rate $d\Lambda_1^{obs}(u) = \pi(u) d\Lambda_1(u)$. The second term, however, contains $\hat{p}_1(u)$ under the summation, with $\hat{p}_1(u)$ being obtained from complete case observations from the neighborhood of $u = X_i$. Since the convergence rate of $\hat{p}_1(u)$ is known to be slower than $n^{-1/2}$ (Fan and Gijbels, 1996), we cannot use the standard arguments to replace $\hat{p}_1(u)$ with $p_1(u)$, and need some additional work on the second term.

Let's re-write the second term in (5.4) as follows:

$$\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{t}y(u)^{-1}[\hat{p}_1(u)\big(1-R_i(u)\big)dN_{0i}(u)-Y_i(u)\big(1-\pi(u)\big)d\Lambda_1(u)]=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{t}y(u)^{-1}[\hat{p}_1(u)\big(1-R_i(u)\big)dN_{0i}(u)-Y_i(u)\big(1-\pi(u)\big)p_1(u)d\Lambda_0(u)]=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{t}y(u)^{-1}\times\big\{\hat{p}_1(u)\big(1-R_i(u)\big)dN_{0i}(u)-Y_i(u)\big(1-\pi(u)\big)p_1(u)d\Lambda_0(u)+$$

$$+Y_i(u)[\hat{p}_1(u)-p_1(u)]\big(1-\pi(u)\big)d\Lambda_0(u)-Y_i(u)[\hat{p}_1(u)-p_1(u)]\big(1-\pi(u)\big)d\Lambda_0(u)+$$

$$+p_1(u)\big(1-R_i(u)\big)dN_{0i}(u)-p_1(u)\big(1-R_i(u)\big)dN_{0i}(u)\big\}=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{t}y(u)^{-1}p_1(u)[dN_{0i}^{miss}(u)-Y_i(u)d\Lambda_0^{miss}(u)]+$$

$$+\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{t}y(u)^{-1}Y_i(u)[\hat{p}_1(u)-p_1(u)]d\Lambda_0^{miss}(u)+$$

$$\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{t}y(u)^{-1}[\hat{p}_1(u)-p_1(u)]\times[\big(1-R_i(u)\big)dN_{0i}(u)-Y_i(u)d\Lambda_0^{miss}(u)] \quad (5.9)$$

The first term of (5.5) is the sum of the i.i.d zero mean random variables, using same arguments as in (5.3).

For the second term of (5.5), note that

$$\hat{p}_1(u)-p_1(u)=\frac{\frac{1}{n}\sum_{j=1}^{n}R_j(X_j)K_b(u-X_j)[I(\epsilon_j=1)-p_1(u)]}{\frac{1}{n}\sum_{j=1}^{n}R_j(X_j)K_b(u-X_j)}.$$

where $K_b(.)$ is the kernel function. Also note that

$$\frac{1}{n}\sum_{j=1}^{n}R_j(X_j)K_b(u-X_j)=\hat{f}_{cc}(u)$$

is a kernel estimator of the density of the event times of complete case observations.

Using Taylor expansion,

$$
\hat{p}_1(u) - p_1(u) = \frac{\frac{1}{n}\sum_{j=1}^{n} R_j(X_j)K_b(u - X_j)[I(\epsilon_j = 1) - p_1(u)]}{\hat{f}_{cc}(u)} =
$$

$$
\approx \frac{\frac{1}{n}\sum_{j=1}^{n} R_j(X_j)K_b(u - X_j)[I(\epsilon_j = 1) - p_1(u)]}{f_{cc}(u)} -
$$

$$
- \frac{\frac{1}{n}\sum_{j=1}^{n} R_j(X_j)K_b(u - X_j)[I(\epsilon_j = 1) - p_1(u)]}{f_{cc}(u)^2}[\hat{f}_{cc}(u) - f_{cc}(u)] =
$$

$$
= \frac{\frac{1}{n}\sum_{j=1}^{n} R_j(X_j)K_b(u - X_j)[I(\epsilon_j = 1) - p_1(u)]}{f_{cc}(u)} -
$$

$$
- \frac{\frac{1}{n}\sum_{j=1}^{n} R_j(X_j)K_b(u - X_j)[I(\epsilon_j = 1) - p_1(u)]}{\hat{f}_{cc}(u)}\frac{\hat{f}_{cc}(u)}{f_{cc}(u)^2}[\hat{f}_{cc}(u) - f_{cc}(u)] =
$$

$$
= \frac{\frac{1}{n}\sum_{j=1}^{n} R_j(X_j)K_b(u - X_j)[I(\epsilon_j = 1) - p_1(u)]}{f_{cc}(u)} -
$$

$$
- \frac{\hat{f}_{cc}(u)}{f_{cc}(u)^2}[\hat{p}_1(u) - p_1(u)] \times [\hat{f}_{cc}(u) - f_{cc}(u)].
$$

$$(5.10)$$

Note that $f_{cc}(u)$ is bounded away from zero on $[0, \tau]$, hence $f_{cc}^{-2}(u)$ is bounded. By Lemma 1 and Theorem B from Mack and Silverman, both $sup_{u\in[0,\tau]}|\hat{p}_1(u) - p_1(u)|$ and $sup_{u\in[0,\tau]}|\hat{f}_{cc}(u) - E\hat{f}_{cc}(u)|$ are $O_p\big((nb)^{-1/2}(log(1/b)^{1/2})\big)$. From the theory of kernel estimators (Fan and Gijbels, 1996), $sup_{u\in[0,\tau]}|E\hat{f}_{cc}(u) - f_{cc}(u)| = O_p(b^2)$. After some algebra and under our assumptions, this implies that

$$
sup_{u\in[0,\tau]}|\hat{p}_1(u) - p_1(u)| \times sup_{u\in[0,\tau]}|\hat{f}_{cc}(u) - f_{cc}(u)| = o_p(n^{-1/2}).
$$

Note that since we only use the complete case observations in the estimator $\hat{p}_1(u)$, technically all the convergence rates for kernel estimators in (5.6) are with respect to m, the number of complete case observations in the sample. However, since we require the probability of the event type being observed $\pi(u)$ to be bounded away from zero, this means that $\frac{n}{m} < M$ for some constant $M < \infty$, and thus the convergence rates can be equivalently expressed in

terms of n.

Using that and substituting (5.6) into (5.5), the second term of (5.5) can be re-written as

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^t y(u)^{-1}Y_i(u)[\hat{p}_1(u) - p_1(u)]d\Lambda_0^{miss}(u) =$$

$$\int_0^t [\hat{p}_1(u) - p_1(u)]d\Lambda_0^{miss}(u) + o_p(n^{-1/2}) =$$

$$= \int_0^t \frac{\frac{1}{n}\sum_{j=1}^{n} R_j(X_j)K_b(u - X_j)[I(\epsilon_j = 1) - p_1(u)]}{f_{cc}(u)} d\Lambda_0^{miss}(u) + o_p(n^{-1/2}) =$$

$$= \frac{1}{n}\sum_{j=1}^{n} R_j(X_j)\int_0^t f_{cc}^{-1}(u)K_b(u - X_j)[I(\epsilon_j = 1) - p_1(u)]d\Lambda_0^{miss}(u) + o_p(n^{-1/2}) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} R_i(X_i)\int_0^t f_{cc}^{-1}(u)K_b(u - X_i)[I(\epsilon_i = 1) - p_1(u)]\big(1 - \pi(u)\big)d\Lambda_0(u) + o_p(n^{-1/2}).$$

$$(5.11)$$

The third term of (5.5) can be shown to be $o_p(n^{-1/2})$.

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^t y(u)^{-1}[\hat{p}_1(u) - p_1(u)] \times [\big(1 - R_i(u)\big)dN_{0i}(u) - Y_i(u)d\Lambda_0^{miss}(u)] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int_0^t y(u)^{-1}[\hat{p}_1(u) - p_1(u)]\big(1 - R_i(u)\big)dN_{0i}(u) - \int_0^t [\hat{p}_1(u) - p_1(u)]d\Lambda_0^{miss}(u) + o_p(n^{-1/2}) =$$

$$= \frac{1}{n^2}\sum_{i,j=1}^{n}\int_0^t y(u)^{-1}f_{cc}^{-1}(u)R_j(X_j)K_b(u - X_j)[I(\epsilon_j = 1) - p_1(u)]\big(1 - R_i(u)\big)dN_{0i}(u) -$$

$$- \frac{1}{n}\sum_{i=1}^{n}\int_0^t R_i(X_i)f_{cc}^{-1}(u)K_b(u - X_i)[I(\epsilon_i = 1) - p_1(u)]d\Lambda_0^{miss}(u) + o_p(n^{-1/2}),$$

$$(5.12)$$

83

using same arguments as in (5.6).

Define

$$v(X_i, R_i(X_i), \delta_i, X_j, R_j(X_j), \epsilon_j) = \frac{R_j(X_j)K_b(X_i - X_j)(I(\epsilon_j = 1) - p_1(X_i))(1 - R_i(X_i))\delta_i}{f_{cc}(X_i)y(X_i)}.$$

Note that the diagonal elements $v(X_i, R_i(X_i), \delta_i, X_i, R_i(X_i), \epsilon_i) = 0$ because they always involve $R_i(X_i)(1 - R_i(X_i)) \equiv 0$. Therefore,

$$(5.8) = \frac{1}{n^2} \sum_{i,j=1, i \neq j}^{n} v(X_i, R_i(X_i), \delta_i, X_j, R_j(X_j), \epsilon_j)$$

$$- \frac{1}{n} \sum_{i=1}^{n} \int_0^t f_{cc}^{-1}(u) K_b(u - X_i)[I(\epsilon_i = 1) - p_1(u)] d\Lambda_0^{miss}(u) + o_p(n^{-1/2}).$$

To make the kernel symmetric, let

$$w(X_i, R_i(X_i), \delta_i, \epsilon_i, X_j, R_j(X_j), \delta_j, \epsilon_j) =$$
$$= \frac{1}{2}v(X_i, R_i(X_i), \delta_i, X_j, R_j(X_j), \epsilon_j) + \frac{1}{2}v(X_j, R_j(X_j), \delta_j, X_i, R_i(X_i), \epsilon_i),$$

and define

$$U_n = C_k^n \sum_{1 \leq i < j \leq n} w(X_i, R_i(X_i), \delta_i, \epsilon_i, X_j, R_j(X_j), \delta_j, \epsilon_j).$$

Then

$$(5.8) = U_n - \frac{1}{n} \sum_{i=1}^{n} \int_0^t f_{cc}^{-1}(u) K_b(u - X_i)[I(\epsilon_i = 1) - p_1(u)] d\Lambda_0^{miss}(u) + o_p(n^{-1/2}). \quad (5.13)$$

To define the Hajek projections of U_n, let

$$G(X_1, R_1(X_1), \delta_1, \epsilon_1) = E(w(X_1, R_1(X_1), \delta_1, \epsilon_1, X_2, R_2(X_2), \delta_2, \epsilon_2)|X_1, R_1(X_1), \delta_1, \epsilon_1) =$$
$$= \frac{1}{2}E(v(X_1, R_1(X_1), \delta_1, X_2, R_2(X_2), \epsilon_2)|X_1, R_1(X_1), \delta_1) +$$
$$+ \frac{1}{2}E(v(X_2, R_2(X_2), \delta_2, X_1, R_1(X_1), \epsilon_1)|X_1, R_1(X_1), \epsilon_1).$$

We can compute each part:

$$G_1(x_1, r_1, d_1) = E(v(X_1, R_1(X_1), \delta_1, X_2, R_2(X_2), \epsilon_2)|X_1 = x_1, R_1(X_1) = r_1, \delta_1 = d_1) =$$
$$E\left(\frac{R_2(X_2)K_b(x_1 - X_2)(I(\epsilon_2 = 1) - p_1(x_1))(1 - r_1)d_1}{f_{cc}(x_1)y(x_1)}|X_1 = x_1, R_1(X_1) = r_1, \delta_1 = d_1\right) =$$
$$= \frac{r_1 d_1}{f_{cc}(x_1)y(x_1)}\int K_b(x_1 - u)(E(I(\epsilon_2 = 1)|T_2 = u) - p_1(x_1))h(u)du =$$
$$= \frac{r_1 d_1}{f_{cc}(x_1)y(x_1)}\int K_b(x_1 - u)(p_1(u) - p_1(x_1))h(u)du$$

where $h(u)$ is the density of $X_i = min(T_i, C_i)$, and

$$G_2(x_1, r_1, e_1) = E(v(X_2, R_2(X_2), \delta_2, X_1, R_1(X_1), \epsilon_1)|X_1 = x_1, R_1(X_1) = r_1, \epsilon_1 = e_1) =$$
$$E\left(\frac{r_1 K_b(X_2 - x_1)(I(e_1 = 1) - p_1(X_2))(1 - R_2(X_2))\delta_2}{f_{cc}(X_2)y(X_2)}|X_1 = x_1, R_1(X_1) = r_1, \epsilon_1 = e_1\right) =$$
$$= \int r_1 K_b(u-x_1)\frac{(I(\epsilon_1 = 1) - p_1(u))}{y(u)f_{cc}(u)}E(I(R_2(X_2) = 0)|T_2 = u \text{ and } \delta_2 = 1)E(I(\delta_2 = 1)|X_2 = u)h(u)du =$$
$$= \int r_1 K_b(u - x_1)\frac{(I(\epsilon_1 = 1) - p_1(u))}{y(u)f_{cc}(u)}(1 - \pi(u))Pr(\delta_2 = 1|X_2 = u)h(u)du =$$
$$= \int r_1 K_b(u - x_1)\frac{(I(\epsilon_1 = 1) - p_1(u))}{f_{cc}(u)}(1 - \pi(u))d\Lambda_0(u)du.$$

Using asymptotic properties of U-statistics (DasGupta, 2008),

$$U_n = -\theta_n + \frac{1}{n}\sum_{i=1}^{n}G_1(X_i, R_i(X_i), \delta_i) + \frac{1}{n}\sum_{i=1}^{n}G_2(X_i, R_i(X_i), \epsilon_i) + o_p(n^{-1/2}),$$

where

$$\theta_n = E(w(X_i, R_i(X_i), \delta_i, \epsilon_i, X_j, R_j(X_j), \delta_j, \epsilon_j)) =$$
$$= (E(G(X_1, R_1(X_1), \delta_1, \epsilon_1)).$$

Plugging it into (5.9) and noticing that $\frac{1}{n}\sum_{i=1}^{n} G_2(X_i, R_i(X_i), \epsilon_i)$ cancels out,

$$(5.8) = -\theta_n + \frac{1}{n}\sum_{i=1}^{n} G_1(X_i, R_i(X_i), \delta_i) + o_p(n^{-1/2}).$$

Using the properties of the kernel function $K(.)$ and the definition of $K_b(s) = \frac{1}{b}K(\frac{s}{b})$, and doing variable substitution when integrating expressions for $G_1(.)$ and $G_2(.)$, it is easy to show that all the terms in the expression (5.8) are of the order $O_p(b^2)$. Since by assumtion $b = cn^{-\alpha} + o_p(n^{-\alpha})$ with $\alpha > 1/4$, this means that (5.8) is $o_p(n^{-1/2})$.

Combining the above result with (5.4), (5.5) and (5.7), we obtain

$$\hat{\Lambda}_1(t, \hat{p}_1(t)) - \Lambda_1(t) =$$

$$= \sum_{i=1}^{n} \int_0^t y(u)^{-1} p_1(u)[(1 - R_i(u))dN_{0i}(u) - Y_i(u)(1 - \pi(u))d\Lambda_0(u)] +$$

$$+ \sum_{i=1}^{n} \int_0^t y(u)^{-1}[R_i(u)dN_{1i}(u) - Y_i(u)\pi(u)d\Lambda_1(u)] +$$

$$+ \sum_{i=1}^{n} \int_0^t R_i(u) f_{cc}^{-1}(u) K_b(u - X_i)[I(\epsilon_i = 1) - p_1(u)](1 - \pi(u))d\Lambda_0(u) + o_p(n^{-1/2}) =$$

$$= \sum_{i=1}^{n} \int_{0}^{t} y(u)^{-1}[p_1(u)(1 - R_i(u))dN_{0i}(u)] +$$

$$+ \sum_{i=1}^{n} \int_{0}^{t} y(u)^{-1}[R_i(u)dN_{1i}(u)] - \int_{0}^{t} (1 - \pi(u) + \pi(u))d\Lambda_1(u)) +$$

$$+ \sum_{i=1}^{n} \int_{0}^{t} R_i(u)f_{cc}^{-1}(u)K_b(u - X_i)[I(\epsilon_i = 1) - p_1(u)](1 - \pi(u))d\Lambda_0(u) + o_p(n^{-1/2}) =$$

$$= \sum_{i=1}^{n} \int_{0}^{t} y(u)^{-1}(1 - R_i(u))[p_1(u)dN_{0i}(u) - Y_i(u)d\Lambda_1(u)] +$$

$$+ \sum_{i=1}^{n} \int_{0}^{t} y(u)^{-1}R_i(u)[dN_{1i}(u) - Y_i(u)d\Lambda_1(u)] +$$

$$+ \sum_{i=1}^{n} \int_{0}^{t} R_i(u)f_{cc}^{-1}(u)K_b(u - X_i)[I(\epsilon_i = 1) - p_1(u)](1 - \pi(u))d\Lambda_0(u) + o_p(n^{-1/2}) =$$

$$= \frac{1}{n} \sum_{j=1}^{n} \tilde{G}_{i,n}(t) + o_p(n^{-1/2}), \quad (5.14)$$

where

$$\tilde{G}_{i,n}(t) =$$

$$= \int_{0}^{t} (1 - R_i(u))y(u)^{-1}[p_1(u)dN_{0i}(u) - Y_i(u)d\Lambda_1(u)] +$$

$$+ \int_{0}^{t} R_i(u)y(u)^{-1}[dN_{1i}(u) - Y_i(u)d\Lambda_1(u)] +$$

$$+ \int_{0}^{t} R_i(u)f_{cc}^{-1}(u)K_b(u - X_i)[I(\epsilon_i = 1) - p_1(u)](1 - \pi(u))d\Lambda_0(u).$$

Note that $\tilde{G}_{i,n}(t)$ still depends implicitly on n through $b = b_n$ in $K_b(.)$. Therefore we need to make one more step and take the limit of $\tilde{G}_{i,n}(t)$ as $b_n \to 0$ when $n \to \infty$.

It is a known fact that $K_b(u - X_i) \to \delta(u - X_i)$ as $b \to 0$, where $\delta(s)$ - Dirac function.

Also, it is known that

$$\int_{-\infty}^{t} \delta(u - X_i)g(u)du = I(X_i \le t)g(X_i)$$

and

$$\int_{-\infty}^{t} K_b(u - X_i)g(u)du \to I(X_i \le t)g(X_i)$$

as $b \to 0$.

Define

$$F(b, t, X_i, ...) = \int_{0}^{t} K_b(u - X_i)g(u, ...)du.$$

Let $F(0, t, X_i, ...) = I(X_i \le t)g(X_i, ...)$.

If $t < X_i - b$ then $F(b, t, X_i, ...) = 0$ by definition of $K_b(.)$ and $I(X_i \le t) = 0$, so trivially

$$F(b, t, X_i, ...) - F(0, t, X_i, ...) = 0 = O(b^2).$$

If $X_i \in [b, t - b]$ then

$$F(b, t, X_i, ...) = \int_{0}^{t} K_b(u - X_i)g(u, ...)du =$$

$$= \int_{X_i-b}^{X_i+b} K_b(u - X_i)g(u, ...)du =$$

$$= \int_{-1}^{1} K(s)g(X_i + sb, ...)ds$$

Then

$$\frac{\partial}{\partial b}F(b, t, X_i, ...) = \int_{-1}^{1} K(s)\frac{\partial}{\partial b}g(X_i + sb, ...)ds =$$

$$= \int_{-1}^{1} sK(s)\frac{\partial}{\partial u}g(u, ...)|_{u=X_i+sb}ds$$

88

and

$$\frac{\partial^2}{\partial b^2}F(b, t, X_i, ...) = \int_{-1}^{1} s^2 K(s)\frac{\partial^2}{\partial u^2}g(u, ...)|_{u=X_i+sb}ds$$

Note that the first derivative

$$\frac{\partial}{\partial b}F(b, t, X_i, ...)|_{b=0} = \int_{-1}^{1} sK(s)\frac{\partial}{\partial u}g(u, ...)|_{u=X_i}ds =$$

$$= \frac{\partial}{\partial u}g(u, ...)|_{u=X_i}\int_{-1}^{1} sK(s)ds = 0$$

by the choice of kernel.

Therefore, Taylor expansion of $F(b, ...)$ at $b = 0$ to the second term:

$$|F(b, ...) - F(0, ...)| = |\frac{b^2}{2} \times \int_{-1}^{1} s^2 K(s)\frac{\partial^2}{\partial u^2}g(u, ...)|_{u=X_i}ds| \le$$

$$\frac{b^2}{2} \times |\frac{\partial^2}{\partial u^2}g(u, ...)|_{u=X_i}| \int_{-1}^{1} s^2 K(s)ds \le Cb^2$$

if g is such that the second derivative exists and bounded. In our case this is true because the function $g(.)$ is

$$g(u, r, r) = r f_{cc}(u)^{-1}[r - p_1(u)](1 - \pi(u))\lambda_0(u)$$

where all functions of u are twice differentiable with bounded second derivatives, $f_{cc}(u)$ bounded away from zero, $p_1(u), \pi(u)$ - probabilities, and r and e take values 0 or 1. Also note that for a given e, $[e - p_1(u)]$ is either always positive or always negative.

There are somewhat more tedious cases when X_i is closer than b to t or to 0. But recalling that our $g(u, ...)$ is either always positive or always negative, those partial integrals near boundaries will be smaller in absolute value than the integral over the full neighborhood of the width b, and therefore are smaller than Cb^2.

Therefore,
$$F(b, t, X_i, ...) - F(0, t, X_i, ...) = O(b^2) \to 0 \text{ as } b \to 0$$

and the rate of convergence is b^2 which is faster than $n^{-1/2}$ under our assumptions.

Therefore,
$$\tilde{G}_{i,n}(t) = G_i(t) + o_p(n^{-1/2})$$

where

$$G_i(t) = \int_0^t \left(1 - R_i(u)\right) y(u)^{-1} [p_1(u) dN_{0i}(u) - Y_i(u) d\Lambda_1(u)] +$$

$$+ \int_0^t R_i(u) y(u)^{-1} [dN_{1i}(u) - Y_i(u)\Lambda_1(u)] +$$

$$+ I(X_i \leq t) R_i(X_i) f_{cc}(X_i)^{-1} [I(\epsilon_i = 1) - p_1(X_i)] \left(1 - \pi(X_i)\right) \lambda_0(X_i),$$

and the whole expression (5.10) becomes

$$\hat{\Lambda}_1(t, \hat{p}_1(t)) - \Lambda_1(t) = \frac{1}{n} \sum_{j=1}^n G_i(t) + o_p(n^{-1/2}).$$

Normality. $G_i(t)$ are zero-mean i.i.d. random variables, which by CLT implies that for $t \in [0, \tau]$

$$\sqrt{n}\left(\hat{\Lambda}_1(t, \hat{p}_1(t)) - \Lambda_1(t)\right) \xrightarrow{d} N(0, E(G_i^2(t))$$

and suggests the following estimator for the variance of $\hat{\Lambda}_1(t, \hat{p}_1(t))$:

$$\widehat{Var}\left(\hat{\Lambda}_1(t, \hat{p}_1(t))\right) = \frac{1}{n^2} \sum_{i=1}^n \hat{G}_i(t)^2$$

where $\hat{G}_i(t)$ is obtained by substituting empirical counterparts instead of the theoretical quantities in 5.8. Using the same arguments as for the consistency of the point estimate, it can be shown that this variance estimator is consistent. $\square$.

Proof of Theorem 2.

$$\hat{F}_1(t) - F_1(t) = \int_0^t \hat{S}(u)d\hat{\Lambda}_1(u, \hat{p}_1(u)) - \int_0^t S(u)d\Lambda_1(u)$$

$$= \int_0^t S(u)[d\hat{\Lambda}_1(u, \hat{p}_1(u)) - d\Lambda_1(u)] + \int_0^t [\hat{S}(u) - S(u)]d\hat{\Lambda}_1(u, \hat{p}_1(u)) =$$

$$= S(t)[\hat{\Lambda}_1(t, \hat{p}_1(t)) - \Lambda_1(t)] - \int_0^t [\hat{\Lambda}_1(u, \hat{p}_1(u)) - \Lambda_1(u)]dS(u) +$$

$$+ \int_0^t [\hat{S}(u) - S(u)]d\hat{\Lambda}_1(u, \hat{p}_1(u)).$$

Consistency follows from the uniform consistency of Kaplan-Meier estimator and Theorem 1.

Using Taylor expansion and integration by parts,

$$\hat{F}_1(t) - F_1(t) = \int_0^t \hat{S}(u)d\hat{\Lambda}_1(u, \hat{p}_1(u)) - \int_0^t S(u)d\Lambda_1(u)$$

$$\approx \int_0^t S(u)[d\hat{\Lambda}_1(u, \hat{p}_1(u)) - d\Lambda_1(u)] - \int_0^t S(u)[\hat{\Lambda}_0(u) - \Lambda_0(u)]d\Lambda_1(u) =$$

$$= \int_0^t S(u)[d\hat{\Lambda}_1(u, \hat{p}_1(u)) - d\Lambda_1(u)] + \int_0^t [\hat{\Lambda}_0(u) - \Lambda_0(u)]dF_1(u)$$

$$= \int_0^t S(u)[d\hat{\Lambda}_1(u, \hat{p}_1(u)) - d\Lambda_1(u)] - \int_0^t \{F_1(t) - F_1(u)\}[d\hat{\Lambda}_0(u) - d\Lambda_0(u)] \quad (5.15)$$

Now the results can be obtained by re-tracing the steps from the proof of the Theorem 1.

Note that as mentioned in the Methods section earlier, the finite sample variance estimators for the cumulative cause-specific hazard and the cumulative incidence function can be obtained from $\frac{1}{n^2}\sum_{j=1}^n \tilde{G}_{i,n}(t)$ and $\frac{1}{n^2}\sum_{j=1}^n \tilde{I}_{i,n}(t)$ respectively by plugging in the empirical estimates instead of theoretical quantities into the expression for $\tilde{G}_{i,n}(t)$ and $\tilde{I}_{i,n}(t)$. These are the estimators that were used in simulations for this paper.

5.7 Appendix 2

Alternative variance estimators.

We considered several potential variance estimators.

It was hypothecized in the motivating paper (Lin et al., 2013) that when the bandwidth sequence converges at the optimal rate of $n^{-1/5}$, then the second term in (5.5) has the convergence rate $n^{-3/5}$, which is faster than $n^{-1/2}$ and hence this term can theoretically be omitted for the variance estimation and the naive variance estimator which ignores the variability of the local estimates,

$$\widehat{Var}_{NAIVE}(\hat{\Lambda}_1(t, \hat{p}_1(t))) = n^{-1} \sum_{i=1}^{n} (G_{1i}(t) + G_{2i}(t))^2$$

can be used.

The second option we considered was the variance estimator with the bias correction, as suggested by Lin et al. The corresponding estimators of the variance of the cumulative incidence function $\hat{F}_1(t)$ were obtained in a similar fashion to (Lin et al., 2013).

The third alternative version of the variance estimator for $\hat{F}_1$ which we considered was the estimators with the "local variance" correction:

$$\widehat{Var}_{\text{LOC VAR}}(\hat{F}_1(t, \hat{p}_1(u))) = n^{-1} \sum_{i=1}^{n} \psi_i^2(t, \hat{p}_1(u)) + n^{-1} \sum_{i=1}^{n} \nu_i(t, \hat{p}_1(u))$$

where

$$\psi_i(t, \hat{p}_1(u)) = \int_0^t \hat{S}(u) Y^{-1}(s) dM_{1i}(u, \hat{p}_1(u)) + \int_0^t [\hat{F}_1(t) - \hat{F}_1(u)] dM_{0i}(u),$$

$$dM_{1i}(u, \hat{p}_1(u)) = R_i(X_i) dN_{1i}(u) + (1 - R_i(X_i)) \hat{p}_1(u, \hat{p}_1(u)) dN_{0i}(u) - Y_i(u) d\hat{\Lambda}_1(u, \hat{p}_1(u)),$$

$$dM_{0i}(u) = dN_{0i}(u) - Y_i(u) d\hat{\Lambda}_0(u),$$

and

$$\nu_i(t, \hat{p}_1(u)) = I(X_i \le t)(1 - R_i(X_i)) \hat{S}^2(X_i) Y(X_i)^{-2} \widehat{Var}(\hat{p}_1(X_i)) dN_{0i}(X_i)$$

The variance of the local estimate, $\widehat{Var}(\hat{p}_1(X_i))$, is obtained using the theory for local polynomial regression. For the local linear estimator this variance is $\frac{\sigma^2(X_i)}{f(X_i)nh} \int_{-\infty}^{\infty} K^2(u) du$, where $\sigma^2(u)$ is the variance of $\theta(u)$ and $f(u)$ is the density of the CDF of the times of events with observed event type (Fan and Gijbels, 1996; Loader, 1999). For practical purposes, its

estimate is available from software packages, for example, R package *locfit*.

BIBLIOGRAPHY

Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous markov chains based on censored observations. *Scandinavian Journal of Statistics*, 5:141–150.

Altman, D. (1998). Confidence intervals for the number needed to treat. *BMJ*, 317:1309–1312.

Altman, D. and Andersen, P. (1999). Calculating the number needed to treat for the trials where the outcome is time to event. *BMJ*, 319:1492–1495.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.

Andersen, P. K. and Gill, R. D. (1982). Coxs regression model for counting processes: a large sample study. *Annals of Statistics*, 10:11001120.

Bordes, L. and Gneyou, K. E. (2011). Uniform convergence of nonparametric regressions in competing risk models with right censoring. *Statistics and Probability Letters*, 81(11):1654–1663.

Braun, T. M. and Yuan, Z. (2007). Comparing the small sample performance of several variance estimators under competing risks. *Statistics in Medicine*, 26:1170–1180.

Breslow, N. E. (1972). Contribution to the discussion of paper by D.R.Cox, regression models and life tables. *Journal of the Royal Statistical Society, Series B*, 34:216–217.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics*, 30:89–99.

Cao, R., Lopez-de Ullibarri, I., Janssen, P., and Veraverbeke, N. (2003). Presmoothed Kaplan-Meier and Nelson-Aalen estimators. *Journal of Nonparametric Statistics*, 17:31–56.

CDER (2002). *Guidance for Industry: Antiretroviral Drugs Using Plasma HIV RNA Measurements - Clinical Considerations for Accelerated and Traditional Approval.* U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER).

Chatterjee, N., Sinha, S., Diver, W. R., and Feigelson, H. S. (2010). Analysis of cohort studies with multivariate and partially observed disease classification data. *Biometrika*, 97(3):683–698.

Cheng, S. C., Fine, J. P., and Wei, L. J. (1998). Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*, 54:219–228.

Cole, S., Hernn, M., Anastos, K., Jamieson, B., and Robins, J. (2007). Determining the effect of highly active antiretroviral therapy on changes in human immunodeficiency virus type 1 RNA viral load using a marginal structural left-censored mean model. *American Journal of Epidemiology*, 166:219–227.

Connor, J. and Imrey, P. (2005). Proportions, inferences, and comparisons. In *Encyclopedia of Biostatistics*. John Wiley & Sons, Ltd. DOI: 10.1002/0470011815.b2a10047.

Cook, R. and Sackett, D. (1995). The number needed to treat: a clinically useful measure of treatment effect. *BMJ*, 310:452–454.

Cook, T. D. and Kosorok, M. R. (2004). Analysis of time-to-event data with incomplete event adjudication. *Journal of the American Statistical Association*, 99(468):1140–1152.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of Royal Statistical Society, Series A*, 34:187–220.

Craiu, R. V. and Duchesne, T. (2004). Inference based on the em algorithm for the competing risks model with masked causes of failure. *Biometrika*, 91(3):543–558.

Craiu, R. V. and Reiser, B. (2006). Inference for the dependent competing risks model with masked causes of failure. *Lifetime Data Analysis*, 12(1):21–33.

Cummings, F. J., Gray, R., Tormey, D. C., Davis, T. E., Volk, H., Harris, J., Falkson, G., and Bennett, J. (1993). Adjuvant tamoxifen versus placebo in elderly women with node-positive breast cancer: Long-term follow-up and causes of death. *Journal of Clinical Oncology*, 11:29–35.

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*. Springer Texts in Statistics.

DeGruttola, V., Hughes, M., Gilbert, P., and Phillips, A. (1998). Trial design in the era of highly effective antiviral drug combinations for hiv infection. *AIDS*, 12(Suppl A):S149–56.

Dewanji, A. (1992). A note on a test for competing risks with missing failure type. *Biometrika*, 79(4):855–857.

Dinse, G. E. (1982). Nonparametric estimation for partially-complete time and type of failure data. *Biometrics*, 38(2):417–431.

Dinse, G. E. (1986). Nonparametric prevalence and mortality estimators for animal experiments with incomplete cause-of-death data. *Journal of the American Statistical Association*, 81(394):328–336.

Dony, J., Einmahl, U., and Mason, D. M. (2006). Uniform in bandwidth consistency of local polynomial regression function estimators. *Austrian Journal of Statistics*, 35:105–120.

Effraimidis, G. and Dahl, C. M. (2013). Nonparametric estimation of cumulative incidence functions for competing risks data with missing cause of failure. *Discussion Papers on Business and Economics*, 13.

Egger, M., May, M., Chene, G., Phillips, A., Ledergerber, B., Dabis, F., Costagliola, D., D'Arminio, M. A., de Wolf, F., Reiss, P., Lundgren, J., Justice, A., Staszewski, S., Leport, C., Hogg, R., Sabin, C., Gill, M., Salzberger, B., and Sterne, J. (2002). Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. *The Lancet*, 360:119–129.

El-Nouty, C. and Lancar, R. (2005). The presmoothed nelsonaalen estimator in the competing risk model. *Communications in Statistics - Theory and Methods*, 33(1):135 – 151.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall London.

Feuer, E. J., Lee, M., Mariotto, A. B., Cronin, K. A., Scoppa, S., Penson, D., Hachey, M., Cynkin, L., Carter, G., Campbell, D., Percy-Laurry, A., Zao, Z., Schrag, D., and Hankey, B. F. (2012). The cancer survival query system: making survival estimates from seer more timely and relevant for recently diagnosed patients. *Cancer*, 118.

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

Fine, J. P. (2001). Regression modeling of competing crude failure probabilities. *Biostatistics*, 2:85–97.

Fischl, M., Ribaudo, H., Collier, A., Erice, A., Giuliano, M., Dehlinger, M., Eron, J., Saag, M., Hammer, S., Vella, S., and Morse, G. (2003). A randomized trial of 2 different 4-drug antiretroviral regimens versus a 3-drug regimen, in advanced human immunodeficiency virus disease. *Journal of Infectious Diseases*, 188:625–634.

Gao, G. and Tsiatis, A. A. (2005). Semiparametric estimators for the regression coefficients in the linear transformation competing risks model with missing cause of failure. *Biometrika*, 92(4):875–891.

Gijbels, I., Lin, D., and Ying, Z. (1993). *Non-and semi-parametric analysis of failure time data with missing failure indicators.*

Gijbels, I., Lin, D., and Ying, Z. (2007). Non- and semi-parametric analysis of failure time data with missing failure indicators. *Lecture Notes-Monograph Series*, 54:203–223.

Gilbert, P., Ribaudo, H., Greenberg, L., Yu, G., Bosch, R., Tierney, C., and Kuritzkes, D. (2000). Considerations in choosing a primary virological endpoint for durability in aids antiretroviral trials. *AIDS*, 14:1961–1972.

Goetghebeur, E. and Ryan, L. (1990). A modified log rank test for competing risks with missing failure type. *Biometrika*, 77(1):207–211.

Goetghebeur, E. and Ryan, L. (1995). Analysis of competing risks survival data when some failure types are missing. *Biometrika*, 82(4):821–833.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on the new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London A*, 115:513–580.

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

Gouskova, N., Cole, S., Eron, J., and Fine, J. (2014a). Viral suppression in hiv studies: Combining times to suppression and rebound. *Biometrics*, 70(2):441–448. DOI:10.1111/biom.12140.

Gouskova, N., Kundu, S., Imrey, P., and Fine, J. (2014b). Number needed to treat for time to event data with competing risks. *Statistics in Medicine*, 33(3):181–192. DOI:10.1002/sim.5922.

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

Gulick, R., Ribaudo, H., Shikuma, C., Lalama, C., Schackman, B., Meyer, W., Acosta, E., Schouten, J., Squires, K., Pilcher, C., Murphy, R., Koletar, S., Carlson, M., Reichman, R., Bastow, B., Klingman, K., and Kuritzkes, D. (2006). Three- vs four-drug antiretroviral regimens for the initial treatment of HIV-1 infection a randomized controlled trial. *The Journal of the American Medical Association*, 296:769–781.

Gulick, R., Ribaudo, H., Shikuma, C., Lustgarten, S., Squires, K., Meyer, W., Acosta, E., Schackman, B., Pilcher, C., Murphy, R., Maher, W., Witt, M., Reichman, R., Snyder, S., Klingman, K., and Kuritzkes, D. (2004). Triple-nucleoside regimens versus efavirenz-containing regimens for the initial treatment of HIV-1 infection. *New England Journal of Medicine*, 350:1850–1861.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24:726–748.

Jacome, M. and Cao, R. (2007). Almost sure asymptotic representation for the presmoothed distribution and density estimators for censored data. *Journal of Theoretical and Applied Statistics*, 41(6):517–534.

Jacome, M., Gijbels, I., and Cao, R. (2008). Comparison of presmoothing methods in kernel density estimation under censoring. *Computational Statistics*, 23:381–406.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: John Wiley.

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

Kattan, M. W., Heller, G., and Brennan, M. F. (2003). A competing-risks nomogram for sarcoma-specific death following local recurrence. *Statistics in Medicine*, 22(22):3515–3525.

Klein, J. P., Logan, B., Harhoff, M., and Andersen, P. (2007). Analyzing survival curves at a fixed point in time. *Statistics in Medicine*, 26:4505–4519.

Koller, M., Raatz, H., E.W., S., and M., W. (2012). Competing risks and the clinical community: irrelevance or ignorance? *Statistics in Medicine*, 31:1089–97.

Korn, E. L., Graubard, B. I., and Midthune, D. (1997). Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *American Journal of Epidemiology*, 145(1):72–80.

Laan, M. J. v. d. and McKeague, I. W. (1998). Efficient estimation from right-censored data when failure indicators are missing at random. *The Annals of Statistics*, 26(1):164–182.

Lai, H. J., Cheng, Y., Ch, H., Kosorok, M. R., and Ferrel, P. M. (2004). Association between initial disease presentation, lung disease outcomes, and survival in patients with cystic fibrosis. *American Journal of Epidemiology*, 159(6):537–546.

Last, J. (1988). *A Dictionary of Epidemiology, Second Edition*. Oxford University Press: New York.

Laupacis, A., Sackett, D., and Roberts, R. (1988). An assessment of clinically useful measures of the consequences of treatment. *New England Journal of Medicine*, 318(26):1728–1733.

Laurent, B. (1996). Efficient estimation of intergral functions of a density. *The Annals of Statistics*, 24(2):659–681.

Lecoutre, J.-P. and Ould-Said, E. (1995). Convergence of the conditional Kaplan-Meier estimate under strong mixing. *Journal of Statistical Planning and Inference*, 44(3):359–369.

Lee, M., A., C. K., Gail, M. H., and Dignam, J. J. (2011). Multiple imputation methods for inference on cumulative incidence with missing cause of failure. *Biometrical Journal*, 53(6):974–993.

Lee, M., Cronin, K. A., Gail, M. H., and Feuer, E. J. (2012). Predicting the absolute risk of dying from colorectal cancer and from other causes using population based cancer registry data. *Statistics in Medicine*, 31(5):489–500.

Lee, M., Dignam, J. J., and Han, J. (2013). Multiple imputation methods for nonparametric inference on cumulative incidence with missing cause of failure. *Statistics in Medicine*, 33:4605–4626.

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

Lin, D. Y., Fleming, T. R., and Wei, L. J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika*, 81:73–81.

Lin, F.-C., Cai, J., Fine, J. P., and Lai, H. J. (2013). Nonparametric estimation of the mean function for recurrent event data with missing event category. *Biometrika*, 100(3):727–740.

Linton, O., Mammen, E., Nielsen, J. P., and Van Keilegom, I. (2011). Nonparametric regression with filtered data. *Bernoulli*, 17(1):60–87.

Lo, S.-H. (1991). Estimating a survival function with incomplete cause-of-death data. *Journal of Multivariate Analysis*, 39(2):217–235.

Loader, C. (1999). *Local Regression and Likelihood*, volume 47. Springer New York.

Lu, K. and Tsiatis, A. A. (2001). Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics*, 57(4):1191–1197.

Lu, W. and Liang, Y. (2008). Analysis of competing risks data with missing cause of failure under additive hazards model. *Statistica Sinica*, 18(1):219.

Mack, Y. and Silverman, B. (1982). Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 61(3):405–415.

Mariotto, A. B., Wang, Z., Klabunde, C. N., Cho, H., Das, B., and Feuer, E. J. (2013). Life tables adjusted for comorbidity more accurately estimate noncancer survival for recently diagnosed cancer patients. *Journal of Clinical Epidemiology*. DOI:10.1016/j.jclinepi.2013.07.002.

McKeague, I. W. and Subramanian, S. (1998). Product-limit estimators and Cox regression with missing censoring information. *Scandinavian Journal of Statistics*, 25(4):589–601.

McQuay, H. and Moore, R. (1997). Using numerical results from systematic reviews in clinical practice. *Annals of Internal Medicine*, 126:712–720.

Mellors, J., Rinaldo, C., Gupta, P., White, R., Todd, J., and Kingsley, L. (1996). Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science*, 272:1167–1170.

Miyakawa, M. (1984). Analysis of incomplete data in competing risks model. *Reliability, IEEE Transactions on*, R-33(4):293–296.

Pencina, M. J., Larson, M. G., and DAgostino, R. B. (2007). Choice of time scale and its effect on significance of predictors in longitudinal studies. *Statistics in Medicine*, 26:1343–1359.

Pepe, M. S. (1991). Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association*, 86:770–778.

Pepe, M. S. and Fleming, T. (1989). Weighted Kaplan-Meier statistics - a class of distance tests for censored survival data. *Biometrics*, 45:497–507.

Pepe, M. S. and Mori, M. (1993). Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data. *Statistics in Medicine*, 12:737–751.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 34:541–554.

Racine-Poon, A. H. and Hoel, D. G. (1984). Nonparametric estimation of the survival function when cause of death is uncertain. *Biometrics*, 40(4):1151–1158.

Ribaudo, H. J., Kuritzkes, D. R., Schackman, B. R., Acosta, E. P., Shikuma, C. M., and Gulick, R. M. (2006). Design issues in initial HIV-treatment trials: focus on ACTG A5095. *Antiviral therapy*, 11:751–760.

Riddler, S., Haubrich, R., DiRienzo, A., L., P., W.G., P., K.L., K., K.W., G., T., G., J.F., R., Brizz, B., Lalloo, U. G., Murphy, R. L., Swindells, S., Havlir, D., and Mellors, J. (2008). Class-sparing regimens for initial treatment of HIV-1 infection. *New England Journal of Medicine*, 358:2095–2106.

Robbins, G., De Gruttola, V., Shafer, R., Smeaton, L., Snyder, S., Pettinelli, C., Dub, M., Fischl, M., Pollard, R., Delapenha, R., Gedeon, L., van der Horst, C., Murphy, R., Becker, M., D'Aquila, R., Vella, S., Merigan, T., and Hirsch, M. (2003). Comparison of sequential three-drug regimens as initial therapy for HIV-1 infection. *New England Journal of Medicine*, 349:2293–2303.

Rosenbaum, P. (1984). The consquences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society. Series A (General)*, 147:656–666.

101

Sax, P., Tierney, C., Collier, A., Fischl, M., Mollan, K., and Peeples, L. e. a. (2009). Abacavir-lamivudine versus tenofovir-emtricitabine for initial HIV-1 therapy. *New England Journal of Medicine*, 361:2230–2240.

Scheike, T. H. and Zhang, M. J. (2003). Extensions and applications of the Cox-Aalen survival model. *Biometrics*, 59:1036–1045.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241.

Shen, Y. and Cheng, S. C. (1999). Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics*, 55:1093–1100.

Subramanian, S. (2006). Survival analysis for the missing censoring indicator model using kernel density estimation techniques. *Statistical Methodology*, 3(2):125 – 136.

Subramanian, S. and Bean, D. (2008). Hazard function estimation from homogeneous right censored data with missing censoring indicators. *Statistical Methodology*, 5(6):515 – 527. DOI:10.1016/j.stamet.2008.01.003.

Thiebaut, A. C. M. and Benichou, J. (2004). Choice of time-scale in Coxs model analysis of epidemiologic cohort data: a simulation study. *Statistics in Medicine*, 23:3803–3820.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Nat. Acad. Sci. U.S.A.*, 72:20–22.

Wang, Q., Dinse, G., and Liu, C. (2012). Hazard function estimation with cause-of-death data missing at random. *Annals of the Institute of Statistical Mathematics*, 64(2):415 – 438.