Event History Analysis in Longitudinal Cohort Studies with Intermittent Inspection Times

by

Yayuan Zhu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics - Biostatistics

Waterloo, Ontario, Canada, 2016

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Event history studies based on disease clinic data often face several complications. Specifically, patients visit the clinic irregularly, and the intermittent inspection times depend on the history of disease-related variables; this can cause event or failure times to be dependently interval-censored. Furthermore, failure times could be truncated, treatment assignment is non-randomized and can be confounded, and there are competing risks of the failure time outcomes under study. I propose a class of inverse probability weights applied to estimating functions so that the informative inspection scheme and confounded treatment are appropriately dealt with. As a result, the distribution of failure time outcomes can be consistently estimated. I consider parametric, non- and semi-parametric estimation. Monotone smoothing techniques are employed in a two-stage estimation procedure for the non- or semi-parametric estimation. Simulations for a variety of failure time models are conducted for examining the finite sample performances of proposed estimators. This research is initially motivated by the Psoriatic Arthritis (PsA) Toronto Cohort Study at the Toronto Western Hospital and the proposed methodologies are applied to this cohort study as an illustration.

Acknowledgements

First and foremost I want to offer my sincerest gratitude to my supervisors, Dr. Jerry Lawless and Dr. Cecilia Cotton, who have supported me throughout this thesis with their patience, extensive knowledge and experience whilst encouraging and inspiring me to develop independent research capability. I attribute the completion of this thesis to their enormous and comprehensive support. They provided me with direction, theoretical and technical guidance, scientific application and writing suggestions, and even career advices. They became more of mentors and friends, than supervisors, which I truly appreciate. One simply could not wish for a better or friendlier supervisor, but I am so fortunate to have two.

I would also like to thank my thesis committee, Dr. Richard Cook, Dr. Leilei Zeng, Dr. Mark Oremus, and Dr. Eleanor Pullenayegum for their helpful suggestions, insightful comments and inspiring questions at the defence. Especially, Dr. Pullenayegum read the thesis thoroughly and gave valuable suggestions which are greatly helpful for thesis and papers revision.

I must also acknowledge Dr. Dafna Gladman, Dr. Vinod Chandrad and their team from the Psoriatic Arthritis (PsA) Clinic, Toronto Western Hospital, who kindly provided the PsA data and helped me with background learning and data cleaning. Additionally, a very special thanks goes out to Ker-Ai Lee who selflessly shared her experience on this data and statistical programming. Meanwhile, I want to thank Dr. Cook for his sponsorship of my account on Compute Canada (sharcnet) to use the computing facilities.

I would also like to thank other PhD students and professors at the University of Waterloo for our discussion on research, exchanges of knowledge, skills, which helped me broaden my horizons and enrich my experience. Furthermore, my friends at Waterloo are

the sources of enthusiasm and happiness. Particularly, my friend and roommate, Ying Wu, my officemates, Dr. Yujie Zhong and Kexin Ji, as well as Dr. Hua Shen, Dr. Liqun Diao and her Husband Dr. Chengguo Weng, Lu Cheng, Dr. Min Chen, Dr. Narges Nazeri Rad, Yingying Xu, among many more.

Last but not least, I would also like to thank my parents for their unconditional love and support and in particular, I warmly appreciate their understanding, emotional support and encouragement during the most engaged and stressful days. To them I dedicate this thesis.

Dedication

To my loving and caring parents Ruiling Wang and Zhenxiang Zhu.

Table of Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

Studies based on disease clinic data often face several complications. Patients may visit the clinic irregularly, and the intermittent inspection times may depend on disease-related variables. Intermittent observation can cause failure time outcomes to be dependently interval-censored, and failure times may also be left-truncated by clinic enrolment time. Additionally, treatment assignments are frequently not randomized and even can be affected by disease-related variables. In this thesis, a class of inverse probability weighted estimating function approaches will be proposed to consistently estimate failure time distributions by adjusting for the informative observation and measured confounders. Simulation studies are conducted to empirically examine the finite sample performances of proposed methods. Data from the Psoriatic Arthritis (PsA) Toronto Cohort Study is used for illustration.

In this chapter, first we provide some background of the PsA cohort study, which mainly motivates this thesis research. Secondly, some basic concepts of survival analysis and causal inference will be briefly introduced. Last but not least, problems and challenges

arising from intermittent outcome-dependent observation, model marginalization, and collapsibility of association measures will be addressed.

1.1 Motivating Example

In clinical, epidemiological and sociological research, longitudinal studies, which involve repeated observations on subjects over long periods of time, constitute a primary source of information on outcomes of interest. For example, researchers may be interested in quantifying the association between air pollution and lung function, where air pollution might be measured weekly and lung function of individuals is evaluated at periodic clinic visits. In clinical experiments, variables of interest are usually measured at regular and prespecified time points, e.g. in months, for lung function assessments. However, in practice, many longitudinal studies are observational studies in which subjects may miss scheduled visits, or may visit a clinic at arbitrary time points. In *"regular"* longitudinal studies, visit times are prespecified and often common to every subject, so they do not carry any information related to the outcome of interest. However, in *"irregular"* longitudinal studies, visit times are often associated with the outcome or outcome-related variables. If we do not take this into account in the analysis, estimates can be severely biased. We will discuss this in detail in a subsequent section. At present, irregular longitudinal data based on intermittent observation or dependent follow-up times are insufficiently studied, but they are very common in practice, especially in health-related research. In the following, we will see a real example in which participants are interviewed or evaluated in continuous time, but the frequency and timing of visits vary greatly and may be highly associated with the values of previous outcomes or outcome-related variables.

1.1.1 Data on Psoriatic Arthritis

Psoriasis is a chronic immune-mediated inflammatory skin disease affecting approximately 2% of the general population (Langley et al., 2005). Additionally, about $10-30\%$ of patients with psoriasis have psoriatic arthritis (PsA), which is defined as seronegative inflammatory musculoskeletal disease associated with psoriasis. Recent studies indicate that PsA is a progressive disease, leading to considerable joint pain, inflammation and destruction which can ultimately cause serious disability and poor quality of life (Chandran et al., 2010; O'Keeffe et al., 2011). The etiology of PsA is multifactorial, with genetic, environmental, and immunologic factors involved in its development (Gladman, 1998; Mease and Goffe, 2005).

The Toronto Psoriatic Arthritis Clinic was established by Professor Dafna Gladman at the University of Toronto in 1978. Since October 1995, it has been at The Centre for Prognosis Studies in the Rheumatic Diseases (CPSRD) at the Toronto Western Hospital. During the past 35 years, the clinic has collected comprehensive longitudinal information on the course and prognosis of PsA. So far, it has enrolled over 1000 patients with PsA who have been followed over many years. It constitutes one of the largest cohorts of PsA in the world. The study is approved by the Research Ethics Board of the University Health Network, Toronto, Ontario, Canada. Patients are assessed about every 6-12 months according to a defined protocol and data is collected on clinical history, pharmacotherapy, physician examination, laboratory evaluations such as routine blood and urine tests, and biennially performed X-ray tests. Physician examination includes the rheumatological assessment, which assesses the activity and clinical damage of peripheral joints and spine. Demographic information and family disease history are also registered at recruitment.

Clinical damage of a joint is defined by the presence of a limitation in the range of

movement of more than 20% of the range when there is no active inflammation, or if the joint is deformed, flail, ankylosed or has undergone surgery (Siannis et al., 2006). Clinical joint damage is determined on physical examination of the patient, which is done at each visit. In general, damage is an irreversible process, while disease activity, which is reflected by tenderness and/or effusion, is reversible. Therefore, most recent therapies aim at reducing signs and symptoms of active arthritis so as to inhibit the progression of structural damage. So far, a variety of therapies have been adopted to control the disease activity of PsA. There are three main types of treatment. Non-steroidal anti-inflammatory drugs (NSAIDs) and disease modifying anti-rheumatic drugs (DMARDs) are typically used as the first- and the second-line treatments. More recently, due to the immunologic basis of PsA, biologics have attracted increasing interest for treating disease activity. In addition, if these front-line therapies are not effective at reducing inflammation, other treatments such as intra-articular steroids injected directly into the specific active joint(s) may also be considered (O'Keeffe et al., 2011).

At present, a major research objective is to identify genetic and genomic variants associated with PsA disease progression. In addition, since joint damage mainly characterizes the disease severity of PsA, much attention has been paid to investigating the link between joint damage and the dynamic courses of pharmacotherapy and disease activity. Many interesting questions can be addressed from the PsA cohort, for example, evaluating the effects of recent therapies on certain joint damage events. Studying an event time outcome defined with respect to joint damage and biologics will be the focus of Chapter 5.

1.2 Introduction to Survival Analysis

This section aims to provide an overview of terminology, concepts and techniques for survival analysis. Survival time (also referred to as lifetime or failure time) is defined as a positive-valued random variable which typically represents the time to some specific event. This event can be death, the development of some disease, recurrence of a disease or the failure of a physical (or mechanical, electrical) component. It could also be a good event, such as disease remission, cessation of smoking, and so forth. In the PsA example, one event of interest is time to the appearance of the first joint damage since onset. Section 1.2.1 aims to introduce some basic measures commonly used in survival analysis. Section 1.2.2 focuses on various types of censoring which frequently occur in practice. Following that, Section 1.2.3 discusses the likelihood function construction based on failure time data. Finally, some widely used survival models are introduced in Section 1.2.4.

1.2.1 Basic Quantities

Let T_i be a nonnegative random variable that represents the failure time, i.e. time to the event under study, for subject i, where $i = 1, ..., n$. Characterization of the distribution of T_i and discussion of the association between failure time and potential risk factors are often of interest. In addition to the cumulative distribution function (CDF) and probability density (or probability mass) function, other functions including the survival function and hazard function can be used to characterize the distribution of a failure time random variable. It can be shown that if any one of these four quantities is known, then the others are uniquely determined. In the following, we will introduce these basic quantities by considering two cases: when T_i is a univariate continuous random variable and when T_i is a univariate discrete random variable, by referring to Lawless (2003). All functions in the

following, unless stated otherwise, are defined over $(0, \infty)$.

Continuous Quantities

Assume T_i's are i.i.d. nonnegative random variables from some continuous distribution. The CDF, denoted by $F(t)$, of T_i, a continuous survival time variable, is defined by

$$F(t) = Pr(T \leq t) = \int_0^t f(s)ds,$$

where $f(s) = dF(s)/ds$ is the probability density function of T_i at time s.

The survival function, denoted by $S(t)$, is the probability of an individual surviving beyond time t, i.e. experiencing the event after time t. It is defined as

$$S(t) = Pr(T > t) = 1 - F(t) = \int_t^\infty f(s)ds,$$

and hence,

$$f(t) = -\frac{dS(t)}{dt}.$$

Another basic quantity, the most commonly used in survival analysis, is the hazard function denoted by $\lambda(t)$, which is the probability an individual experiences the event in the next instant of time given the individual has not experienced an event by time t. The hazard function is defined by

$$\begin{aligned}\lambda(t) &= \lim_{\Delta \to 0^+} \frac{Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\ &= \frac{f(t)}{S(t)} \\ &= -\frac{d\ln[S(t)]}{dt}.\end{aligned}$$

Roughly speaking, $\lambda(t)\Delta t$ provides the approximate probability of failure during the time period $[t, t + \Delta t)$, given survival up to t.

The cumulative hazard function, $\Lambda(t)$, is defined by

$$\Lambda(t) = \int_0^t \lambda(s)ds,$$

which is related to the survival function by $S(t) = \exp\{-\Lambda(t)\}$. It is clear that any of $f(t)$, $F(t)$, $S(t)$, $\lambda(t)$ and $\Lambda(t)$ uniquely determines the distribution of T_i.

Discrete Quantities

Sometimes, discrete random variables arise due to rounding off measurements or when survival times refer to an integral number of units (Klein and Moeschberger, 2003). Suppose T (subscript i suppressed) can take on values t_1, t_2,..., where $0 = t_0 < t_1 < t_2 < ...$, with probability mass function $p(t_j) = Pr(T = t_j)$, $j = 1, 2, ...$ Then, the CDF of T is defined as $F(t) = Pr(T \leq t) = \sum_{t_j \leq t} p(t_j)$, and the corresponding survival function is given by

$$S(t) = Pr(T > t) = \sum_{t_j > t} p(t_j).$$

Note, when T is continuous, $S(t)$ is a monotone decreasing continuous function with $S(0) = 1$, while when T is discrete, under the above definition, $S(t)$ is a right-continuous, non-increasing step function, with $S(0) = 1$ and $S(\infty) = 0$.

The discrete time hazard function is given by

$$\begin{aligned}
\lambda(t_j) &= Pr(T = t_j | T \geq t_j) \\
&= \frac{p(t_j)}{S(t_{j-1})} \\
&= 1 - \frac{S(t_j)}{S(t_{j-1})}, \qquad j = 1, 2, ...
\end{aligned}$$

Since $S(t_j) = [1 - \lambda(t_j)]S(t_{j-1})$ and $S(t_0) = S(0) = 1$, we have

$$S(t) = \prod_{t_j \leq t} [1 - \lambda(t_j)].$$

Moreover, as an analog of the continuous case, a discrete cumulative hazard function $\Lambda(t)$ equals $\sum_{j:t_j \leq t} \lambda(t_j)$ and $-\ln S(t)$. A general formulation of $\Lambda(t)$ can be given by a Riemann-Stieltjes integral to unify continuous, discrete, and mixed survival time distribution in one framework of the form (Lawless, 2003):

$$\Lambda(t) = \int_0^t d\Lambda(s) = \int_0^t \lambda(s)ds + \sum_{j:t_j \leq t} \lambda_j,$$

where $\lambda(s) = f(s)/S(s)$ represents the hazard function for T at points where $F(s)$ (or $S(s)$) is continuous, and $\lambda_j = Pr(T = t_j | T \geq t_j)$ is the discrete hazard value at time t_j for which a jump in F occurs.

1.2.2 Varieties of Censoring and Truncation

In practice, survival data are often subject to censoring, which, broadly speaking, occurs when some event is only known to have occurred within a certain interval but the exact time is unknown. There are three primary types of censoring: right censoring, left censoring and interval censoring. Each type leads to a certain likelihood structure which forms the basis for likelihood-based inference.

Right Censoring

First, we introduce right censoring which occurs most often in practice, since the survival data are always under observation for a finite period of time. *Right censoring* happens when a failure has not been observed during follow-up and it may occur later. Let C_i be the right censoring time of subject i imposed by the follow-up period. Then, $(0, C_i]$ is the interval over which the failure time of subject i, i.e. T_i, can be observed. Obviously, only the minimum of failure time T_i and right censoring time C_i can be observed, so define

$X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$. Then, the observed data are pairs of the realizations of random variables (X_i, δ_i), $i = 1, 2, ..., n$, where X_i is referred to as an observed time and δ_i is referred to as an event indicator.

Left Censoring

If it is known that the event of interest has already occurred by some time L_i, but the exact failure time is unknown, this is called *left censoring*. The observations still can be characterized by (X_i, δ_i). But, in contrast to right censoring, here the observed time is defined as $X_i = \max(T_i, L_i)$ and the event indicator is defined as $\delta_i = I(T_i \geq L_i)$, where L_i denotes the left censoring time of subject i.

Interval Censoring

Interval censoring means that the failure time of interest is only known to lie within a finite interval instead of being observed exactly. Such censoring usually happens in clinical trials, industrial experiments or longitudinal studies where periodic follow-ups are assigned and a patient's failure time is only known to fall in a certain interval $(t_{il}, t_{ir}]$ between two visits, i.e. $t_{il} < T_i \leq t_{ir}$. Note that if the event of interest occurs exactly at the moment of one visit, then we have $t_{il} = T_i = t_{ir}$, which rarely happens in practice.

Left Truncation

In contrast to censoring where at least partial information is known about failure time, another feature of failure time data, truncation, restricts the inference to conditional estimation. Truncation of failure times occurs when only subjects whose failure times lie within certain observational window (W_{il}, W_{ir})'s are included in the analysis (Klein and

Moeschberger, 2003). When $W_{ir} = \infty$, it is called *left truncation* which often occurs when a subject's study entry time is later than the origin of failure time. A left truncation time W_{il} is also called a *delayed entry time*.

1.2.3 Likelihood Construction for Censored and Truncated Data

Although data may be subject to a variety of types of censoring, the methods for constructing likelihood functions are similar. Generally, suppose that data are subject to all kinds of censoring such as right censoring, left censoring and interval censoring. Then, under the assumption that censoring is independent and non-informative, the likelihood function can be constructed as

$$L \propto \prod_{i \in E} f(T_i) \prod_{i \in R} S(C_i) \prod_{i \in L} [1 - S(L_i)] \prod_{i \in I} [S(t_{il}) - S(t_{ir})], \tag{1.1}$$

where E is the set of exactly observed failure times, R is the set of right-censored observations, L is the set of left-censored observations, and I is the set of interval-censored observations. Here, $f(\cdot)$ and $S(\cdot)$ denote the density function and survival function of failure time T_i, respectively, and C_i, L_i, t_{il}, t_{ir} are defined as before.

Specifically, for right-censored data, the likelihood is of the form

$$L \propto \prod_{i=1}^{n} [f(T_i)]^{\delta_i} [S(C_i)]^{1-\delta_i}, \tag{1.2}$$

where $\delta_i = I(T_i \leq C_i)$ is the event indicator. For interval-censored data, the likelihood is given by

$$L \propto \prod_{i=1}^{n} [S(t_{il}) - S(t_{ir})]^{\delta_i^I} [S(C_i)]^{\delta_i^R}, \tag{1.3}$$

where δ_i^R is the indicator for right censoring and δ_i^I is the indicator for interval censoring.

When data is left truncated, all the quantities included in likelihood function (1.1) would be conditional on T_i being greater than the left truncation time, say $t_{i0} > 0$. Then, the general likelihood function of left-truncated data can be constructed as

$$L \propto \prod_{i \in E} \frac{f(T_i)}{S(t_{i0})} \prod_{i \in R} \frac{S(C_i)}{S(t_{i0})} \prod_{i \in L} \left[1 - \frac{S(L_i)}{S(t_{i0})}\right] \prod_{i \in I} \frac{[S(t_{il}) - S(t_{ir})]}{S(t_{i0})}, \tag{1.4}$$

where T_i, C_i, L_i, t_{il} and t_{ir} are all greater than t_{i0}. Specifically, if data is subject to interval-censoring as well as left-truncation, the corresponding likelihood function can be given by

$$L \propto \prod_{i=1}^{n} \left[\frac{S(t_{il})}{S(t_{i0})} - \frac{S(t_{ir})}{S(t_{i0})}\right]^{\delta_i^I} \left[\frac{S(C_i)}{S(t_{i0})}\right]^{\delta_i^R}. \tag{1.5}$$

1.2.4 Models in Survival Analysis

In this section, some foundational and widely used models in survival analysis will be briefly introduced. First, accelerated failure time (AFT) model is usually applied parametrically, while the other three, proportional hazards (PH) model, additive hazards (AH) model and proportional odds (PO) model, are often known as being semiparametric. Semiparametric model assumptions are usually more flexible and more robust than fully parametric models but bring difficulties in inference due to an unspecified component in the model.

Accelerated Failure Time Regression Model

Suppose the logarithm of T_i follows a location-scale distribution with mean $\beta_0 + \boldsymbol{\beta}'\boldsymbol{Z}_i$ and standard deviation σ, where $\boldsymbol{Z}_i = (Z_{i1}, ..., Z_{ip})'$ is a p-dimensional vector of covariates for subject i, β_0 is the intercept and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ is the p-dimensional vector of coefficients of covariates $\boldsymbol{Z}_i$. That is,

$$ln(T_i) = \beta_0 + \boldsymbol{\beta}'\boldsymbol{Z}_i + \sigma W_i,$$

where W_i is assumed to follow a standard location-scale distribution. If, for example, W_i has a standard extreme value distribution, then the corresponding T_i has a Weibull distribution with shape parameter $1/\sigma$.

Cox Proportional Hazards Regression Model

Suppose that the hazard function is given by

$$\lambda(t; \boldsymbol{Z}_i) = \lambda_0(t) \exp(\boldsymbol{\beta}' \boldsymbol{Z}_i),$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ and $\boldsymbol{Z}_i = (Z_{i1}, ..., Z_{ip})'$. Here, $\lambda_0(t)$, known as the baseline hazard function, is unspecified. The regression parameter, β_j, can be interpreted as the log hazard ratio when Z_{ij} is increased by one unit and other variables are kept unchanged. If the baseline hazard is specified, a parametric PH model can be obtained. For example, when $\lambda_0(t) = \kappa\rho(\rho t)^{\kappa-1}$, we have a Weibull proportional hazards model, where κ is the shape parameter and $1/\rho$ is the scale parameter of a Weibull distribution.

Additive Hazards Regression Model

The hazard is given by

$$\lambda(t; \boldsymbol{Z}_i) = \lambda_0(t) + \boldsymbol{\beta}' \boldsymbol{Z}_i,$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$, $\boldsymbol{Z}_i = (Z_{i1}, ..., Z_{ip})'$, and $\lambda_0(t)$ also denotes the baseline hazard and can be unspecified. Here, β_j represents the hazard difference when Z_{ij} is increased by one unit and other variables remain unchanged. In addition, similar to the PH model, the baseline hazard function can be specified to obtain a parametric AH model.

Proportional Odds Regression Model

This model takes the form

$$O(t) = O_0(t)\ \exp(\boldsymbol{\beta}'\boldsymbol{Z}_i),$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ and $\boldsymbol{Z}_i = (Z_{i1}, ..., Z_{ip})'$. Here, $O(t) = F(t)/\{1 - F(t)\}$ is the odds for distribution function $F(t)$, and $O_0(t)$ is the odds for the baseline distribution $F_0(t)$. Similar to the PH model, β_j is interpreted as the log odds ratio when Z_{ij} is increased by one unit, with other variables fixed. Also, $O_0(t)$ can be unspecified. If $F_0(t)$ is specified by a parametric distribution, e.g. log-logistic, a parametric PO model can be obtained.

1.3 Introduction to Observational Studies and Causal Inference

In practice, observational studies are often used to study human health, especially in epidemiological research. The PsA cohort we introduced in Section 1.1 is an observational study. Patients may be prescribed treatments with NSAIDs, DMARDs and/or biologics based on their clinical assessments. It is of interest to estimate the effect of specific treatment such as biologics, but in an observational study, this is challenging because treatment is prescribed according to a person's condition. This is the focus of Chapter 3, but here we review some causality concepts.

1.3.1 Observational Studies and Causality

The key feature of *randomized experiments* is that treatments or interventions are randomly allocated across individuals or experimental units. The simplest situation is that subjects are assigned to be treated or untreated by the flip of a fair coin, i.e. patients are allocated to be in the treatment arm or control arm with the same probability of 1/2. On the contrary, non-randomized treatments are commonly seen in *observational studies*. A random assignment of treatment ensures balance across study groups in terms of measured and unmeasured risk factors and allows the greatest reliability and validity of statistical estimates of *causal effects*. In fact, in a randomized experiment, *association* between treatment and outcome implies a causal effect of the treatment on outcome.

In an observational study, we can attempt to estimate the effect of a treatment or an exposure by comparing outcomes when "it is not feasible to use controlled experimentation, in the sense of being able to impose the procedures or treatments whose effects are desired to be discovered, or to assign subjects at random to different procedures" (Cochran, 1965). In an observational study, there is no control over the treatment assignment, so treated and untreated subjects may be quite different with respect to disease-related or outcome-dependent characteristics: some subjects could be more likely than others to receive the treatment due to these characteristics. These characteristics which determine if an individual will receive the treatment are referred to as *confounders (or confounding variables)* if they are also risk factors of outcomes. For instance, if doctors are more likely to assign a surgical treatment to sicker patients, while relatively healthier patients are more likely to be assigned standard care. Then, while studying the effect of surgery on survival, health status before treatment is a confounder. In an observational study, associations cannot be generally interpreted as causal effects. Removing the selection biases caused

by confounders is a central objective in the analyses of treatment effects in observational studies.

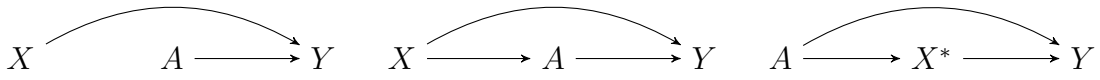
1.3.2 Causal Diagrams

This section aims to introduce some graphical devices, which are often referred to as causal diagrams. The graphical approach is helpful to summarize what we know about the study and what we assume about the relationships between variables relevant to our particular causal inference problem of interest. In practice, it is common to combine two approaches: using causal diagrams to conceptualize problems and using the counterfactual approach, e.g. the marginal structural model (MSM) and inverse probability weighting (IPW) that we will introduce later, to analyze data and do inference.

A diagram like the ones in Figure 1.1 is known as a directed acyclic graph (DAG)(Pearl, 1995), which is a visual summary of the likely (known, suspected or hypothesized) causal links between variables. They are called "directed" because one edge implies a direction, i.e. X may cause A, but not the other way around. The term "acyclic" implies that there are no cycles, i.e. a variable cannot cause itself, either directly or through another variable (Hernán and Robins, 2016).

In a DAG, each variable is represented by a node (vertex), e.g. X, X^*, A and Y in Figure 1.1. Relationships between variables are represented through edges (the arrows). Directed edges represent causal associations. We adopt the convention that time flows from left to right. Thus, Figure 1.1(a) represents a randomized experiment where there is not an arrow from X to A. Figure 1.1(b) represents an observational study where X is a common cause of outcome Y and treatment A. In this case, we say X is a known confounder of the effect of A on Y. Figure 1.1(c) displays that there is an observed variable

denoted by X^* that is affected by treatment A and also predicts the outcome. Here, X^* is called an intermediate variable which is on the causal pathway between A and Y. The path $A \to Y$ represents the direct effect of treatment on outcome, while $A \to X^* \to Y$ represents an indirect effect of treatment on outcome. To unbiasedly estimate the overall effect of A on Y, one should consider both direct and indirect effects. More details about DAGs can be found in a comprehensive book on this subject written by Pearl (2003).



(a) (b) (c)

Figure 1.1: Examples of randomized treatment (a), confounded treatment (b) and intermediate variable (c) by DAGs, where Y is the outcome variable, A is the treatment whose effect on Y is of interest, X is a confounder, and X^* is an intermediate variable.

1.3.3 Causal Inference and Some Important Assumptions

Nowadays, two main competing perspectives on causal inference have risen: the counterfactual perspective and the non-counterfactual perspective. Rubin (1974) proposed a counterfactual approach for causal inference which is the focus of most recent statistical research. For simplicity, assume only two interventions are compared: treated and untreated. Define Y_i^0 as the outcome that would have resulted if subject i was untreated and Y_i^1 as the outcome that would have resulted had subject i been treated. The causal effect of this treatment (denoted by $A = 0$ or 1) on outcome is based on the comparison of these two counterfactual or potential outcomes, Y_i^0 and Y_i^1. However, a subject usually can only

receive one treatment status, i.e. either be treated or be untreated, so only one of these two potential outcomes is observable. Therefore, one of the potential outcomes is counter-to-the-fact, i.e. counterfactual. The causal effect, $Y_i^1 - Y_i^0$, on one single subject can not be observed, but the *average causal effect (ACE)*, $E[Y_i^1] - E[Y_i^0]$, can be estimated and is often evaluated, under certain assumptions in causal inference. If outcome Y_i is a binary variable, the effect defined by $E[Y_i^1] - E[Y_i^0]$ is called the causal risk difference. In addition, a causal risk ratio and causal odds ratio are defined by $\frac{Pr(Y_i^1=1)}{Pr(Y_i^0=1)}$ and $\frac{Pr(Y_i^1=1)/Pr(Y_i^1=0)}{Pr(Y_i^0=1)/Pr(Y_i^0=0)}$, respectively. Furthermore, there can be other effect measures which depend on the context of a study.

The key difference between randomized experiments and observational studies is that randomized experiments can balance observed risk factors as well as unobserved factors. However, in observational studies, even applying appropriate analytical adjustments, we can only balance the known risk factors and have to rely on some assumptions about the unknown risk factors. Let $\boldsymbol{L}$ be a vector of measured covariates which describe the characteristics of a subject prior to treatment assignment. We adopt the convention that upper cases represent random variables, lower cases represent the realized values of random variables, and upper cases in bold fonts denote a vector of random variables. Let Y^a be the counterfactual outcome of treatment status a and let A be a random variable which denotes the actual treatment the subject received. As presented in Robins and Hernán (2009), the important assumptions needed in causal inference are listed below.

(A0) Consistency: If $A = a$ for a given subject, then $Y^a = Y$ for that subject.

(A1) Strongly Ignorable Treatment Assignment/No Unmeasured Confounders: The exposure or treatment assignment must be independent of the counterfactual

outcomes given the observed risk factors, i.e. for all a, we have

$$Y^a \amalg A|\boldsymbol{L}, \qquad (\textit{conditional exchangeability}),$$

and

$$0 < Pr(A = a|\boldsymbol{L}) < 1, \quad \forall\ \boldsymbol{L} \qquad (\textit{positivity}).$$

(A2) Stable Unit-Treatment Value Assumption (SUTVA): Each subject's potential outcomes are not influenced by the actual exposure of another subject.

Consistency, conditional exchangeability and positivity, described in Rosenbaum and Rubin (1983), were referred to as three *identifiability conditions* by Robins and Hernán (2009). The stable unit-treatment value assumption was labeled as "no interaction between units" by Cox in 1958 and was referred to as "no interference between subjects" by Hernán and Robins (2016). Assumptions (A0)-(A2) are essential in casual inference, under which consistently estimating causal effects from observational data is possible. Additionally, correct model specification, accurate data measurement and data missing at random (MAR) are generally required in statistical analyses. Note that all assumptions above are generally untestable. However, investigators' expert knowledge is helpful to enhance the plausibility of these assumptions. In addition, *sensitivity analysis* could be a useful tool to study the magnitude of hidden bias, if the proposed assumptions were violated.

1.3.4 Structural Models

Models for counterfactual outcomes are referred to as structural models (Hernán and Robins, 2016). For example, Robins et al. (2000) proposed a class of marginal structural generalized linear models, i.e.

$$E[Y_a] = g^{-1}(a; \boldsymbol{\beta}), \tag{1.6}$$

where Y_a denotes the counterfactual outcome under the treatment a and g is the link function. Hernán et al. (2000) developed a class of marginal structural Cox proportional hazards models for failure time outcomes, e.g.

$$\lambda_{T^a}(t|V) = \lambda_0(t) \exp(\beta_1 a + \beta_2 V), \tag{1.7}$$

where $\lambda_{T^a}(t|V)$ is the hazard of failure at t among subjects with baseline covariate V in the population had, contrary to fact, all subjects received treatment a at $t = 0$; $\lambda_0(t)$ is an unspecified baseline hazard. Model (1.6) and model (1.7) are called marginal structural models (MSMs) (Robins, 1999; Robins et al., 2000; Hernán et al., 2000; Hernán and Robins, 2016). Model (1.6) and model (1.7) with a time-varying treatment variable can be found in Robins et al. (2000) and Hernán et al. (2000), respectively.

In addition, Hernán et al. (2005) introduced a class of structural accelerated failure time models, e.g. for a time-fixed treatment a, which have the form

$$ln(T_a) = \beta_0 + \beta_1 a + \sigma W, \tag{1.8}$$

or

$$T_a = T_0 \exp(\beta_1 a), \tag{1.9}$$

where T_a is the counterfactual outcome under treatment a and W is a random variable that follows a standard location-scale distribution. Model (1.9) can be developed for a time-varying treatment and then the model is referred to as a structural nested AFT model (SNAFTM) (Hernán et al., 2005; Young et al., 2008, 2010).

1.3.5 Inverse-Probability-of-Treatment (IPT) Weighting

Inverse probability weighting (IPW) was first proposed by Horvitz and Thompson (1952) for surveys in which subjects are sampled with unequal probabilities; Zhao and Lipsitz

(1992) applied that to designs and analysis of two-stage studies; later, Xie and Liu (2005) applied the IPT weighting method to the Kaplan-Meier estimator (Kaplan and Meier, 1958) and log-rank test for survival data; Robins et al. (2000) and Hernán et al. (2000) further applied the IPT weighting to marginal structural models with time-varying treatment. In sampling theory, a hypothetical population (often referred to as a pseudo-population) in which characteristics are balanced across groups can created by weighting. For example, in an observational study where a treatment effect is of interest, a randomized experiment is imitated in the pseudo-population, and therefore associations can be used to estimate causal effects. In practice, the pseudo-population is created by weighting each subject in the original population by the inverse probability of the treatment this subject actually received conditional on measured confounders denoted by $\boldsymbol{X_i}$, i.e. $w_i^* = \frac{1}{Pr(A_i=a|\boldsymbol{X_i}=\boldsymbol{x})}$. The denominator is referred to as a propensity score and also known as a balancing score (Rosenbaum and Rubin, 1983). That is, the pseudo-population consists of w_i^* copies of subject i from the original population. In this sense, estimators constructed by the inverse probability weighting method are called inverse-probability-of-treatment weighted (IPTW) estimators.

1.4 Introduction to Intermittent Observation Schemes and Outcome-Dependent Follow-up

In this section, we discuss problems and challenges in longitudinal cohort studies with intermittent observation schemes and introduce the situation of outcome-dependent follow-up. A preliminary analysis of gap times between consecutive clinic visits in the PsA case will be provided as an illustration. In later chapters, methodology will be proposed to deal

with the outcome-dependent follow-up problem in survival analysis.

1.4.1 Intermittent Observation

In clinical experiments or planned longitudinal cohort studies, individuals are usually scheduled to be evaluated at regular and pre-specified time points during their follow-up. However, in practice, it is frequently found that individuals may miss some scheduled visits. They could return later at a scheduled or a non-scheduled time point, or they could even come to visit at arbitrary time points. That is, observation is intermittent, and the frequency and timing of visits may vary greatly across individuals. This could happen when a planned visit schedule is not adhered to by everyone, or when additional information is available from unplanned observation visits, or when studies are designed with no regular observation schedule (Bůžková and Lumley, 2007). Therefore, broadly speaking, irregular longitudinal studies could comprise discrete or continuous visit times or even a mixture of them. For example, in the PsA example, X-ray tests are scheduled every two years for assessing radiographical joint damage, but patients miss scheduled tests for various reasons. This is an example of discrete time observation scheme with missingness. In addition, patients come to visit the clinic for lab tests and clinical assessments at non-homogeneous times, though visits are planned every 6–12 months by the protocol. This is an example of intermittent visits in continuous time. In addition, Lin et al. (2004) studied a randomized trial comparing several housing interventions for homeless people with mental illness. Although investigators attempted to conduct follow-up interviews every 3 months, participants often missed and showed up between scheduled interviews. In their case, the actual visit times are a mixture of continuous random times and discrete prespecified times.

1.4.2 Outcome-Dependent Follow-up

When the observation times are uniformly prespecified, e.g. participants were planned to be assessed every month and actually adhered to the schedule, observation times would be marginally independent of outcomes and other variables. Then, observation times are automatically balanced among subjects, so they do not need to be adjusted for. On the other hand, in longitudinal studies with intermittent inspection times, the frequency and timing of visits are subject-specific. They could be highly associated with outcomes or outcome-related variables including the past outcome history and past observation history. As a result, follow-up times are unbalanced and could be dependent on the outcome process. Terms used for this problem in literature are informative follow-up, biased follow-up, personalized follow-up or observation, and outcome-dependent follow-up or observation. Here, we adopt the term "outcome-dependent" follow-up. Pullenayegum and Lim (2014) provided a detailed review of methods in longitudinal studies with irregular observation times with a focus on visit processes, assumptions, and study design.

One illustration of outcome-dependent follow-up is the hypothetical example described in Bůžková and Lumley (2007), where interest lies in quantifying the effect of air pollution $A(t)$ on lung function $P(t)$, e.g. to estimate β in the (marginal) outcome model, $E[P(t)|A(t);\beta]$, where β is the regression coefficient of $A(t)$. Define $N(t)$ as the counting process of the cumulative number of observations or visits by time t. Then, $dN(t) = 1$ means there is an observation at time t, $dN(t) = 0$ otherwise. We know air pollution can trigger an asthma attack, and someone with asthma attacks usually has lower lung function. Let $L(t)$ indicate an asthma attack at time t, where $L(t) = 0$ or 1. It is shown in Figure 1.2 that asthma attack behaves as an *intermediate variable* between air pollution $A(t)$ and lung function $P(t)$. Therefore, if we want to estimate the overall effect of air

pollution on lung function, $L(t)$ should not be directly controlled in the outcome model. On the other hand, a patient with an asthma attack is more likely to visit the doctor so that her/his lung function can be measured, which means that asthma attack $L(t)$ is a common risk factor between observation $dN(t)$ and outcome $P(t)$. The DAG for this hypothetical example is exhibited in Figure 1.2. Investigators study the distribution of lung function $P(t)$ based on the observed value $P^{obs}(t)$. In this example, if they ignored the informative observation scheme, i.e. analyze the observed data only, they would very likely to overestimate the influence of air pollution on lung function, since a high proportion of observable data is contributed by persons who had asthma attacks. In other words, the dependent observation scheme acts as a biased selection of the outcomes to be observed and the resulting bias is similar to the bias induced by informative missing data. In this case, an inverse probability weighting method can assist to eliminate the selection bias via appropriately adjusting for the common risk factors between the outcome process and the observation process.

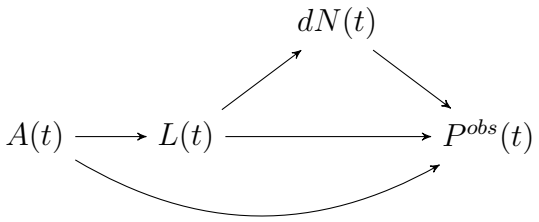


Figure 1.2: DAG for the air pollution and lung function example. $A(t)$: air pollution measured at t; $L(t)$: asthma attack indicator at t; $P^{obs}(t)$: lung function measured at t; $dN(t)$: indicator of a clinic visit at t.

1.4.3 Preliminary Analysis of Visit Times in the PsA Example

To date, over one thousand patients have been followed up over years in the PsA Toronto Cohort Study. Of the 1020 subjects who have at least two recorded clinic visits, we first consider a subcohort of 880 patients with complete information on key disease and treatment variables for a preliminary study of intermittent clinic visits. Among the 880 subjects, calendar dates of visits range from 1973-12-12 to 2013-03-25, because the administrative end of follow-up is Nov. 2013. Demographic information and disease onset times are collected at enrolment. Time-varying variables such as joint activity, joint damage and biomarkers are measured only at clinic visits, except treatments (i.e. NSAIDs, DMARDs and biologics). Therapy history is recalled retrospectively at visits, so the full history of taking a drug is ascertained. Figure 1.3 describes the visit process in this example. People who have PsA are recruited in this cohort study, so clinic enrolment time t_{i0}, which is the first visit, is some time point past the PsA onset time. Clinical evaluation and lab tests are conducted at visits, i.e. $t_{i0}, ..., t_{i,m_i}$ for subject i. Meanwhile, therapy history, e.g. names of specific drugs, the start date and stop date of usage, and the reasons for termination or switch, is recalled at visits.

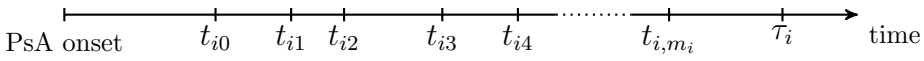


Figure 1.3: Scenario of the visit process in the PsA example, where t_{i0} is the clinic enrolment time and $t_{i1}, ..., t_{i,m_i}$ are the m_i intermittent clinic visits, and τ_i is the administrative end of follow-up for subject i.

To analyse the visit process, we consider the following stratified semi-Markov propor-

tional hazards model for visit gap times:

$$\lambda_N(t|\boldsymbol{Z_i}(t^-);\boldsymbol{\alpha}) = \lambda^s_{N0}(B(t))\,\exp(\boldsymbol{\alpha}'\boldsymbol{Z_i}(t^-)), \qquad s = 1, ..., S, \tag{1.10}$$

where t is the chronological time since the time origin, i.e. t_{i0}, of the visit process; $B(t)$ is the gap time between the most recent past visit and time t; λ^s_{N0} is the stratified baseline hazard function which is unspecified, where s denotes the strata defined by the decades of the most recent visit prior to t, i.e. 1970, 1980,..., 2010; $\boldsymbol{Z_i}(t^-)$ represents some features of the observed history of risk factors prior to t where t^- denotes the instant prior to t. Later we will see from the analysis results that $\boldsymbol{Z_i}(t^-)$ could include the history of outcome or outcome-related variables. On average, patients have about 11 visits from enrolment to τ_i. The length of follow-up ranges from 35 days to about 36 years, with a median of approximately 7 years, and a standard deviation (SD) of about 8 years. The number of visits ranges from 1 to 56, with a median of 7 and a SD of approximately 12. A "visit gap" is defined by the time gap between two successive visits and has a median of 196 days ($SD = 439$ days), with a range from 5 days to about 25 years. Although 53% of visit gaps are between 6 months and 12 months, as expected, there are some extreme cases such as visit gaps longer than 20 years. In total, 34% of visit gaps are shorter than 6 months, 13% are longer than 12 months, and about 1.6% are longer than 5 years. Table 1.1 shows the summary of fitting model (1.10) to the data. The attributes of variables that are considered in the analysis are given in Table A.1 in Appendix A.1 at the end of the thesis. From Table 1.1, we see that visit intensities are strongly associated with age, erythrocyte sedimentation rate (ESR), treatment status of NSAIDs, DMARDs and biologics and the median length of past visit gaps. Also, there is some evidence that family history of PsA is also related to visit times. This preliminary analysis indicates that in the PsA example, visit times are strongly dependent on disease-related variables, especially disease status represented by biomarkers (e.g. ESR) and treatments, and the history of past visits. That

is, the visit process will be informative to studies of disease progression.

1.5 Marginal or Partially Conditional Regression Models

Although multivariate regression models are widely used in observational data analyses, scientific interests may also include the association between a particular risk factor and outcome. As for the example described in Figure 1.2, controlling for an asthma attack by including it as a regressor in the outcome model for lung function could avoid biased analysis results due to a dependent observation scheme, but researchers' interest might lie in studying the marginal effect of air pollution on lung function. Then, the pathway $A(t) \longrightarrow L(t) \longrightarrow P(t)$ should not be blocked. Otherwise, not a marginal effect but a direct effect of $A(t)$ on $P(t)$ will be given, since an asthma attack, $L(t)$, plays the role of an intermediate variable between air pollution and lung function. In this section, we discuss marginal (or partially conditional, i.e. only conditional on a primary covariate, e.g. treatment, of interest) models and the collapsibility of association measures in regression models.

Suppose $A(t)$ (time-fixed or time-varying) is a covariate of prime research interest. The marginal effect of $A(t)$ on outcome is a population-averaged association measure, but when another risk factor, say $L(t)$, is controlled, the effect of $A(t)$ on outcome is interpreted as the conditional effect of $A(t)$ for a particular subset of individuals given $L(t)$. The adjusted exposure effect conditional on $L(t)$ and the unadjusted effect can differ, when the expected value of outcome is modeled as a nonlinear function of the exposure (Greenland et al., 1999). The difference between the adjusted and unadjusted association measures is

Table 1.1: Summary of fitting the stratified semi-Markov PH model (1.10) for visit gap times in the analysis set composed of 880 PsA patients. Variable med.gap denotes the median length of past visit gap times; coef denotes the coefficient estimate of a regressor, and exp(coef) is interpreted as a relative risk or hazard ratio of one unit change of the regressor, and se(coef) is the standard error of the coefficient estimate.

	coef	exp(coef)	se(coef)	z	$Pr(>\|z\|)$	
ESR	-2.05E-03	9.98E-01	6.30E-04	-3.251	0.00115	**
sex	1.98E-02	1.02E+00	2.15E-02	0.924	0.35572	
age	4.20E-03	1.00E+00	9.42E-04	4.461	8.17E-06	***
PS duration	1.01E-03	1.00E+00	1.02E-03	0.997	0.31869	
PsA duration	-1.25E-03	9.99E-01	1.42E-03	-0.876	0.38086	
family history of PS	-1.26E-02	9.88E-01	2.21E-02	-0.567	0.57053	
family history of PsA	-6.28E-02	9.39E-01	3.34E-02	-1.879	0.06025	.
number of active joints	-1.94E-03	9.98E-01	1.28E-03	-1.516	0.12951	
number of damaged joints	2.88E-04	1.00E+00	8.76E-04	0.328	0.74267	
NSAIDs	1.21E-01	1.13E+00	2.16E-02	5.581	2.40E-08	***
DMARDs	2.16E-01	1.24E+00	2.14E-02	10.105	$<$ 2e-16	***
biologics:$I(B(t) \leq 180)$	-3.13E-02	9.69E-01	6.39E-02	-0.49	0.62418	
biologics:$I(B(t) > 180)$	1.74E-01	1.19E+00	3.32E-02	5.244	1.57E-07	***
biologics:$I(B(t) > 365)$	2.34E-01	1.26E+00	1.02E-01	2.304	0.02121	*
med.gap:$I(B(t) \leq 180)$	-8.07E-04	9.99E-01	1.53E-04	-5.281	1.28E-07	***
med.gap:$I(B(t) > 180)$	-8.64E-04	9.99E-01	8.17E-05	-10.573	$<$ 2e-16	***
med.gap:$I(B(t) > 365)$	-8.66E-05	1.00E+00	5.50E-05	-1.574	0.11542	
—						
Signif. codes:	*** 0.001	** 0.01	* 0.05	. 0.1		

referred to as non-linearity by Janes et al. (2010) and as non-collapsibility by Greenland et al. (1999). Many widely used regression models are not collapsible due to non-linearity of association measures. Specifically, Janes et al. (2010) discussed non-collapsibility for logistic regression models, and Martinussen and Vansteelandt (2013) focused on the Cox models and additive hazards models. In addition, Aalen et al. (2015) indicated that even in a randomized survival study, a hazard model, $\lambda(t)$, is not generally collapsible.

For example, suppose that a conditional Aalen's additive model (Aalen, 1980, 1989) is defined by

$$\lambda(t|A, L) = \beta_0(t) + \beta_A(t)A + \beta_L(t)L, \tag{1.11}$$

then it was shown in Martinussen and Vansteelandt (2013) that the hazard model given A alone is given by

$$\lambda(t|A) = \beta_0(t) + \beta_A(t)A + \beta_L(t)\frac{E(e^{-B_L(t)L}L|A)}{E(e^{-B_L(t)L}|A)}, \tag{1.12}$$

where $B_L(t) = \int_0^t \beta_L(s)ds$. If L and A are independent and we define a new intercept as

$$\tilde{\beta}_0(t) = \beta_0(t) + \beta_L(t)\frac{E(e^{-B_L(t)L}L)}{E(e^{-B_L(t)L})},$$

the collapsibility of $\beta_A(t)$ in model (1.11) is shown by

$$\lambda(t|A) = \tilde{\beta}_0(t) + \beta_A(t)A.$$

Martinussen and Vansteelandt (2013) also showed that for a Cox conditional effect model, in general the marginal effect of A and the conditional effect of A with L controlled are not equal, and the proportional hazards assumption does not hold for marginal hazards $\lambda(t|A)$, even if L and A are independent. Therefore, another issue is that some model assumptions can be violated in the marginal effect models even though they hold in the conditional effect models.

The goal of this thesis is to study a marginal treatment effect when the observation scheme is intermittent and informative and the treatment or exposure in an observational study is likely confounded. A class of inverse probability weighting methods will be proposed to eliminate the selection bias which arises from irregular inspection times and confounded treatment for estimation of marginal effect models. Simulation studies will be employed to investigate the performances of the resulting weighted estimates. However, there is very limited literature on data generation mechanisms for inverse probability weighted estimation because of the difficulty of model marginalization and non-collapsibility. For example, suppose that the objective is to estimate the marginal distribution of T given $A(t)$, which is defined by a hazard model $\lambda(t|A(t))$. In addition, there are some other known risk factor(s) of T, denoted by $\boldsymbol{L}(t)$. A conditional model of T given both $A(t)$ and $\boldsymbol{L}(t)$ is defined by $\lambda(t|A(t), \boldsymbol{L}(t))$. The marginal model given $A(t)$ alone can be obtained theoretically by marginalizing the conditional model $\lambda(t|A(t), \boldsymbol{L}(t))$ over $\boldsymbol{L}(t)$. If $\boldsymbol{L}(t)$ is a time-fixed discrete variable, the marginalization is relatively feasible, but for a time-varying $\boldsymbol{L}(t)$, the integration over $\boldsymbol{L}(t)$ is not easy and rarely results in a neat form of model. Most recent papers on this problem either make certain assumptions about T and the covariate process $\boldsymbol{L}(t)$ which might not be plausible in practice or give approximate relationships of simple forms, e.g. Young and Tchetgen Tchetgen (2014), Havercroft and Didelez (2012), or do not result in a simple form of marginal model. Our simulations in Chapter 2 are based on a mechanism suggested by Young and Tchetgen Tchetgen (2014), which allows a time-varying ancillary variable $L(t)$. Another simulation design of log-normal failure time distribution and time-fixed covariates will be introduced in Chapter 3. Additionally, Aalen's additive hazards models will be a focus of Chapter 4, on semiparametric estimation with intermittent observation.

1.6 Outline of the Following Chapters

In Chapter 2 and Chapter 4, we will propose an inverse-intensity-of-visit (IIV) weighted estimating function approach to adjust for intermittent and outcome-dependent inspection times so that a marginal outcome model for failure time data can be consistently estimated. In Chapter 3, the estimation of causal effects of exposures or treatments on failure time outcomes will be considered. In Chapter 5, the association between treatment with biologics and a joint damage event in the PsA Toronto Cohort Study will be analysed as an illustration of the methodologies proposed in the preceding chapters. Finally, concluding remarks and future work will be discussed in Chapter 6.

Chapter 2

PARAMETRIC ANALYSIS OF INTERVAL-CENSORED FAILURE TIME DATA WITH DEPENDENT INSPECTION TIMES

In Section 1.5, we discussed the marginalization and collapsibility of regression models. Marginal failure time distributions or the overall associations with some particular factors are often of substantive interest, and then other covariates that are not the targets of inference should not be conditioned on, especially when they act as intermediate variables on the pathway between the primary factor and outcome. For example, in the Toronto PsA Cohort Study introduced in Section 1.1, one interesting question is how treatment with biologics is associated with disease-related outcomes such as joint activity or joint damage. Thus, usage of biologics is the primary factor for scientific interest, and other risk factors

like gender, age, health status, disease duration, and family history should not be included in the key outcome model. However, when other factors affect the timing of clinic visits, marginal analysis of observed data may lead to biased results, which we have discussed in Section 1.4.2 and will show by simulations in Section 2.3.1. Although some information obtained at or prior to treatment initiation may affect disease progression as well as the treatment assignment, we will discuss this later in Chapter 3 which focuses on causal inference. This chapter aims at the parametric estimation of the marginal distribution of a failure time outcome variable and its marginal association with a time-fixed exposure or treatment, like biologics, in the presence of intermittent and outcome-dependent inspection times.

In the PsA study, patients are planned to be inspected every 6-12 months according to a protocol. Although the median length of gap times between consecutive clinic visits is about 6 months, the visit gaps are highly variable and range from 5 days to 25 years. In Section 1.4.3, we have seen that how often patients come to visit the clinic depends on demographic information, biomarkers, treatments, family disease history, and the history of past visits. Therefore, the subject-specific visit times are informative or disease-dependent. At each visit, disease status such as joint activity or damage is assessed, but the exact onset time of a joint condition is not observable. That is, a joint event is subject to dependent interval-censoring due to the irregular clinic visits, so standard estimation methods such as maximum likelihood estimation (MLE) or generalized estimating equations (GEEs) could lead to biased estimates. In this chapter, we propose an inverse-intensity-of-visit (IIV) weighting method applied to estimating equations which can appropriately adjust for outcome-dependent follow-up times and provide consistent estimation in parametric survival models. Before Section 2.2.4, we assume that the visit process is not discontinued by the occurrence of failure or event and that visit times are continuous. For the case

where failure terminates visits, we convert to a discrete time visit process and pretend to observe the responses after failure is known to have occurred. This will be discussed in detail in Section 2.2.4.

2.1 Estimating Equations for Interval-Censored Failure Time Data

In survival analysis, if a continuous failure time is monitored at periodic visits, it is often interval-censored. For example, consider a study of the time to onset of bladder cancer, where participants are scheduled to visit clinic annually. Investigators know that a patient was first diagnosed with bladder cancer at the jth visit and that bladder cancer was still absent at the $(j-1)$th visit, but the exact onset time is unobservable. In the PsA example, because joint damage is evaluated only at clinic visits, the exact time of appearance of a damaged joint is not observable. If the inspection times of subject i, denoted by t_{ij} where $j = 1, ..., m_i$, are completely independent of outcomes or are conditionally independent of outcomes given the covariates which have been included in the outcome model, the likelihood for interval-censored data given below can produce consistent estimators:

$$L = \prod_{i=1}^{n} \prod_{j=0}^{m_i} \left[S_T(t_{ij}) - S_T(t_{i,j+1}) \right]^{\delta_{ij}} , \tag{2.1}$$

where $t_{i0} = 0$ and $t_{i,m_i+1} = +\infty$, $\delta_{ij} = I\{t_{ij} < T_i \leq t_{i,j+1}\}$, and $S_T(t) = Pr(T_i > t)$ is the survival function of T_i, $i = 1, ..., n$. For convenience, we suppress the dependency of T_i on covariates in the notation. However, when inspection times are informative or outcome-dependent, the likelihood in (2.1) can lead to biased estimates. One way to control for the intermittent inspection times is to regress on all the risk factors that are

related to the outcome process as well as the visit process, but the resulting regression coefficient indexing the variable of interest will be interpreted as an association with failure time conditional on the values of all other risk factors. Section 2.2 will introduce an inverse probability weighting approach which can produce consistent estimates of marginal regression parameters. This approach is based on estimating functions (White, 1982), so we will first introduce some estimating functions for failure time data before considering the adjustment for dependent inspection times.

As an aid to discussion and interpretation, notation will be defined in the context of the PsA example. Let $\boldsymbol{A_i} = (A_{i1}, ..., A_{iq})'$ be a q-dimensional vector of time-fixed exposure or treatment variables measured at $t = 0$ for subject i, where $i = 1, ..., n$. For simplicity, we assume that the time origin corresponds to a subject's clinic entry time, unless stated otherwise. In the PsA example, these variables of particular interest could be gender, age at PsA's onset time, family history of disease or treatments received at clinic entry. Exposure can be considered fixed in three settings: first, if every subject's exposure occurs only once at the start of follow-up (e.g. vaccination, radiation from a bomb explosion, a surgical intervention); second, if the exposure remains constant over time (e.g. genotype); or third, if the exposure evolves over time in a deterministic way (e.g. age) (Robins and Hernán, 2009). Then, define a time-to-event variable denoted by T_i, which can be the time to the presence of some joint event from clinic entry for subject i, e.g. time to an increase in the number of damaged joints since enrolment.

Instead of assessing T_i directly, we define a longitudinal binary outcome $P_i(t) = I(T_i > t)$ whose mean given $\boldsymbol{A_i}$ would be the marginal (or partially conditional) survival function of T_i given $\boldsymbol{A_i}$, i.e. $E[P_i(t)|\boldsymbol{A_i}] = S_T(t|\boldsymbol{A_i})$. Then, $P_i(t)$ is a monotone function from 1 to 0. In addition, let C_i be a random drop-out time and τ_i be the administrative end of follow-up for subject i, and we define $C_i \leq \tau_i$. Additionally, let $0 < t_{i1} < t_{i2} < ... < t_{im_i} \leq C_i$

be the m_i intermittent inspection times of subject i. Moreover, let $\boldsymbol{L_i}(t)$ be a vector of time-varying auxiliary variables which affect the risk of failure as well as the timing of visits. Some of $\boldsymbol{L_i}(t)$ may be defined for all t but measured only at clinic visits, e.g. inflammation evaluated by lab tests. On the other hand, some factors change at certain known time points, e.g. the exact start dates and stop dates of treatments such as NSAIDs, DMARDs and biologics are reported retrospectively at visits, so their whole history is known. Later, we will introduce an important assumption that the dependent inspection times are based only on the "observed history" of relevant variables. Finally, we need to introduce some notation for a counting process: let $\{N_i(t), t > 0\}$ be the cumulative number of visits for the ith individual through time t, and write $N_i(t) = \int_0^t dN_i(s)$, for $t > 0$. Let $C_i(t) = I(C_i > t)$ be the at-risk process and define $dN_i^*(t) = dN_i(t)C_i(t)$, so we have $\{t_{i1}, ..., t_{im_i}\} = \{t > 0 : dN_i^*(t) = 1\}$.

To introduce the estimating function method for failure time data, firstly, we assume that a subject can be followed up to the last visit, t_{im_i}, before loss to follow-up, even if failure occurred before t_{im_i}. This is realistic in some cases, e.g. in the PsA example where patients who have already been diagnosed with joint damage can still visit the clinic before the administrative end of follow-up.

2.1.1 Estimating Function Methods for Interval-Censored Failure Times

We assume in this section that unless stated otherwise, the visit times, t_{ij} where $j = 1, ..., m_i$, are conditionally independent of T_i given $\boldsymbol{A_i}$. The dependent visit times case will be considered in Section 2.2, but it is convenient to introduce the type of estimating functions for failure time outcomes under independent visit scheme first. Suppose that T_i

follows a parametric model whose survival function given $\boldsymbol{A_i}$ is denoted by $S_T(t|\boldsymbol{A_i};\boldsymbol{\theta})$, and our objective is to estimate the parameter $\boldsymbol{\theta}$. For example, we could assume that T_i has a parametric proportional hazards model, i.e. its hazard function is given by

$$\lambda_T(t|\boldsymbol{A_i};\boldsymbol{\theta}) = \lambda_{T,0}(t;\boldsymbol{\gamma})\,\exp\left[\boldsymbol{\beta}'\boldsymbol{A_i}\right], \tag{2.2}$$

where $\boldsymbol{\theta} = (\boldsymbol{\gamma}', \boldsymbol{\beta}')'$, and $\lambda_{T,0}(t;\boldsymbol{\gamma})$ is a parametric baseline hazard function, which could be the hazard of a Weibull, log-normal, log-logistic, Gompertz, etc, and e^{β_l} is interpreted as the marginal hazard ratio contributed by one unit change of A_{il}, $l = 1, ..., q$, with other variables unchanged. The corresponding marginal survival function is $S_T(t|\boldsymbol{A_i};\boldsymbol{\theta}) = \exp\left[-\int_0^t \lambda_T(s|\boldsymbol{A_i};\boldsymbol{\theta})ds\right]$, provided that T_i is a continuous variable. Consider $P_i(t)$, $t = t_{i,1}, ..., t_{im_i}$, as the repeated measures of survival status; an estimating function for the parameter $\boldsymbol{\theta}$ can then be defined by

$$\begin{aligned}\boldsymbol{U}(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\int_0^{\tau_i} \boldsymbol{c}(t|\boldsymbol{A_i};\boldsymbol{\theta})[P_i(t) - S_T(t|\boldsymbol{A_i};\boldsymbol{\theta})]dN_i^*(t) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} \boldsymbol{c}(t_{ij}|\boldsymbol{A_i};\boldsymbol{\theta})[P_i(t_{ij}) - S_T(t_{ij}|\boldsymbol{A_i};\boldsymbol{\theta})],\end{aligned} \tag{2.3}$$

where $dN_i^*(t) = dN_i(t)C_i(t)$ and $C_i(t) = I(C_i > t)$ as defined earlier, and $\boldsymbol{c}(t|\boldsymbol{A_i};\boldsymbol{\theta})$ is a vector of known functions of t conditional on $\boldsymbol{A_i}$, with the same dimension as $\boldsymbol{\theta}$; components of $\boldsymbol{c}(t|\boldsymbol{A_i};\boldsymbol{\theta})$ are linearly independent functions of t for all $\boldsymbol{\theta}$. Note that the unbiasedness of estimating function (2.3) holds regardless of the specification of $\boldsymbol{c}(\cdot)$. Each $P_i(t)$ has a Bernoulli (Binomial) distribution with mean $S_T(t_{i,j}|\boldsymbol{A_i};\boldsymbol{\theta})$, so the estimating function (2.3) is equivalent to the score function of $n \times m_i$ independent binary outcome $P_i(t)$'s, if the function $\boldsymbol{c}(t|\boldsymbol{A_i};\boldsymbol{\theta})$ is defined by

$$\boldsymbol{c}(t|\boldsymbol{A_i};\boldsymbol{\theta}) = \frac{\partial S_T(t|\boldsymbol{A_i},\boldsymbol{\theta})/\partial\boldsymbol{\theta}}{S_T(t|\boldsymbol{A_i},\boldsymbol{\theta})[1 - S_T(t|\boldsymbol{A_i},\boldsymbol{\theta})]}. \tag{2.4}$$

Generalized linear models or GEEs (Liang and Zeger, 1986) are widely used for survival analysis, e.g. for current status data by Jewell and Shiboski (1990); Shiboski and Jewell (1992), and for general interval-censored data by Sun (1997); Huang and Rossini (1997); Zhang et al. (2005), and for competing risk models or multi-state models by Andersen et al. (2003); Klein and Andersen (2005). If $dN_i^*(t)$ is independent of the value of outcome $P_i(t)$ given $\boldsymbol{A_i}$, i.e. the condition (B0) to be introduced in Section 2.2.1 and independent random drop-out are satisfied, it is obvious that (2.3) is an unbiased estimation function, i.e. $E\{\boldsymbol{U}_i(\boldsymbol{\theta})|\boldsymbol{A_i}\} = 0$. As a result, by White (1982), we know that the solution to $\boldsymbol{U}(\boldsymbol{\theta}) = \boldsymbol{0}$ is a consistent estimator of $\boldsymbol{\theta}$ under mild regularity conditions.

In estimating function (2.3), we have assumed that $S_T(t|\boldsymbol{A_i};\boldsymbol{\theta})$ is a parametric model, e.g. (2.2). Although we consider semi-parametric models or non-parametric models in Chapter 4, one simple way to gain flexibility for model (2.2) is to apply a piecewise constant baseline proportional hazards model. That is, for a prespecified set of cut-points $0 = a_0 < a_1 < ... < a_K = \infty$, we assume that

$$\lambda_T(t|\boldsymbol{A_i};\boldsymbol{\theta}) = \sum_{k=1}^{K} \rho_k I_k(t) \exp\left(\boldsymbol{\beta}'\boldsymbol{A_i}\right), \tag{2.5}$$

where $I_k(t) = I\{t \in (a_{k-1}, a_k]\}$ and ρ_k's are unknown positive constants; $\boldsymbol{\theta} = (\boldsymbol{\rho}', \boldsymbol{\beta}')'$ are the parameters we want to estimate. It has been discussed in Lawless (2003) that when $K \to \infty$ and $a_k - a_{k-1} \downarrow 0$, the profile likelihood function for $\boldsymbol{\beta}$ based on model (2.5) approaches the partial likelihood of Cox (semi-parametric) proportional hazards model and the parametric MLE of the baseline hazard estimate of model (2.5) approaches the Breslow or generalized Nelson-Aalen estimate for the Cox model. This good approximation to the Cox model makes model (2.5) more flexible than many ordinary parametric models. In practice, a moderate value of K is often chosen, because experience indicates that reduction of the grid fineness beyond a certain point in model (2.5) yields little change in

inferences. Actually, for many practical problems, choosing K to be 4 to 6 would be sufficient (Lawless and Zhan, 1998; He and Lawless, 2003). Other choices of flexible parametric proportional hazards models could be spline specifications (He and Lawless, 2003) or kernel specifications for baseline hazard functions, especially when smooth estimates of hazard functions are preferred. Likewise, an approximate non-parametric estimate of the survival function can by obtained by fitting a piecewise constant baseline hazards models like (2.5) without covariates. We note that for suitably defined parametric models, $S_T(t|A_i;\boldsymbol{\theta})$ is non-increasing in t. However, for non- or semi-parametric estimation, this is not implicit, and then constrained estimation may be necessary, which will be discussed in Chapter 4.

Another interesting question in the PsA example is how to assess the time to the appearance of the first clinical damaged joint from the onset of PsA with respect to some baseline risk factors such as gender fixed at the onset time. In this case, the time origin corresponds to the onset time of PsA rather than the clinic entry time, but the visit process starts after the onset, because only patients who have PsA are enrolled. Then, we can observe the time to the first damaged joint only if it occurs after clinic entry, so failure time might be left-truncated at the clinic entry time which is denoted by t_{i0} for subject i. An estimating function can still be developed to consistently estimate the parameters in the outcome model $S_T(t|\boldsymbol{A_i};\boldsymbol{\theta})$, if visit times, drop-out, and the delayed entry time (i.e. t_{i0}) are all independent of outcomes given $\boldsymbol{A_i}$. Note that $\boldsymbol{A_i}$ and T_i are defined at $t = 0$ ($\leq t_{i0}$), but the visit process $dN_i(t)$ is defined for $t > t_{i0}$. Here, we define a binary longitudinal outcome as $P_i(t) = I(T_i > t)$, for all $t > t_{i0}$, which indicates survival past t. Its mean conditional on $\boldsymbol{A_i}$ and $T_i > t_{i0}$ is $E[P_i(t)|T_i > t_{i0}, \boldsymbol{A_i};\boldsymbol{\theta}] = S_T[t|\boldsymbol{A_i};\boldsymbol{\theta}]/S_T[t_{i0}|\boldsymbol{A_i};\boldsymbol{\theta}]$, for all $t > t_{i0}$, if a parametric model for failure time T_i is assumed. Then, the estimating function

for $\boldsymbol{\theta}$ can be defined by

$$\boldsymbol{U}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_{t_{i0}}^{\tau_i} \boldsymbol{c}(t|\boldsymbol{A_i}; \boldsymbol{\theta}, t_{i0})[P_i(t) - S_T(t|\boldsymbol{A_i}; \boldsymbol{\theta})/S_T(t_{i0}|\boldsymbol{A_i}; \boldsymbol{\theta})]dN_i^*(t), \tag{2.6}$$

where $\boldsymbol{c}(\cdot)$ is a vector of known functions with the same dimension as $\boldsymbol{\theta}$. Again, $S_T(\cdot|\boldsymbol{\theta})$ could be any (flexible) parametric survival function for a failure time outcome.

2.1.2 Relation between Interval-Censored Maximum Likelihood and Estimating Function Methods

When failure time is subject to interval censoring, one standard estimation method for parametric models is to maximize likelihood (2.1), assuming that the inspection times, t_{ij} and $t_{i,j+1}$, which capture the occurrence of failure, are independent of outcome T_i. Alternatively, the estimating function given in (2.3) can be used. Given the m_i inspection times and the true covariance matrix of $P_i(t)$, at $t_{i1}, ..., t_{im_i}$, correctly specified for subject i, it can be shown that an estimating function in the form of (2.3) and the score function based on likelihood (2.1) are identical, as we discuss below.

Recall that $0 < t_{i1} < t_{i2} < ... < t_{im_i} \leq C_i$ are the inspection times of subject i. Now, we define variables $\delta_{ij} = I\{T_i \in (t_{ij}, t_{i,j+1}]\}$, where $j = 0, ..., m_i$ with $t_{i0} = 0$ and $t_{i,m_i+1} = +\infty$, to indicate interval censoring or right censoring at the last visit and let $\pi_{ij} = E(\delta_{ij}|\boldsymbol{A_i}; \boldsymbol{\theta})$, where $0 < \pi_{ij} < 1$. Then, we know that $\sum_{j=0}^{m_i} \pi_{ij} = 1$ and $\sum_{j=0}^{m_i} \delta_{ij} = 1$. The likelihood for independently interval-censored T_i given in (2.1) can be rewritten as (2.7) below:

$$\begin{aligned} L(\boldsymbol{\theta}) &\propto \prod_{i=1}^{n} \prod_{j=0}^{m_i} [S_T(t_{ij}) - S_T(t_{i,j+1})]^{\delta_{ij}} \\ &\propto \prod_{i=1}^{n} \prod_{j=0}^{m_i} \pi_{ij}^{\delta_{ij}}. \end{aligned} \tag{2.7}$$

Note that all the expectations or variances shown above are based on a parametric model given $\boldsymbol{A_i}$, but $\boldsymbol{A_i}$ and parameter $\boldsymbol{\theta}$ are suppressed in expressions for simplicity. We notice that equation (2.7) is the likelihood of n independent multinomial random variables $\boldsymbol{\delta_i} = (\delta_{i0}, ..., \delta_{im_i})'$ with the number of trials fixed as 1. It is known that the multinomial distribution with a fixed number of trials is a member of the exponential family whose probability density or mass function is of the form

$$f(\boldsymbol{\delta_i}; \boldsymbol{p_i}, \phi) = \exp\{[\boldsymbol{\delta_i'} \boldsymbol{p_i} - b(\boldsymbol{p_i}) - d(\boldsymbol{\delta_i}, \phi)]/a(\phi)\}. \tag{2.8}$$

Here, $p_{ij} = ln(\pi_{ij}/\pi_{im_i})$, $j = 0, 1, ..., m_i$, $\boldsymbol{p_i} = (p_{i0}, ..., p_{im_i})'$, and $b(\boldsymbol{p_i}) = -ln(\pi_{im_i})$, $a(\phi) = 1$ and $d(\boldsymbol{\delta_i}, \phi) = 0$. Therefore, by Wedderburn (1974) and McCullagh (1983), it is known that (2.7) is identical to the quasi-likelihood for $\boldsymbol{\delta_i}$, $i = 1, ..., n$, whose score function has the form:

$$\boldsymbol{U}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{D_{i(\delta)}'} \boldsymbol{V_{i(\delta)}^{-1}} [\boldsymbol{\delta_i} - \boldsymbol{\mu_{i(\delta)}}], \tag{2.9}$$

where $\boldsymbol{\mu_{i(\delta)}} = E(\boldsymbol{\delta_i} | \boldsymbol{A_i}, \boldsymbol{\theta})$, $\boldsymbol{D_{i(\delta)}} = \partial \boldsymbol{\mu_{i(\delta)}} / \partial \boldsymbol{\theta'}$, and $\boldsymbol{V_{i(\delta)}^{-1}}$ is a generalized inverse of $\boldsymbol{V_{i(\delta)}} = Var(\boldsymbol{\delta_i} | \boldsymbol{A_i}, \boldsymbol{\theta})$. Explicitly, for δ_{ij}, $\mu_{ij} = \pi_{ij}$, $Var(V_{ij}) = \pi_{ij}(1 - \pi_{ij})$ and $Cov(V_{ij}, V_{il}) = -\pi_{ij}\pi_{il}$, when $j \neq l$, $j, l = 0, 1, ..., m_i$.

Now, we will show the the quasi-likelihood score function (2.9) is identical to the generalization of an estimating function given in (2.3) via a variable transformation from $\boldsymbol{\delta_i}$

to $\boldsymbol{P_i}$. By their definitions, we know

$$\begin{aligned}
\begin{pmatrix} \delta_{i0} \\ \delta_{i1} \\ \vdots \\ \delta_{i,m_i-1} \\ \delta_{im_i} \end{pmatrix} &= \begin{pmatrix} 1-P_{i1} \\ P_{i1}-P_{i2} \\ \vdots \\ P_{i,m_i-1}-P_{im_i} \\ P_{im_i} \end{pmatrix} \\
&= \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 & -1 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} P_{i1} \\ P_{i2} \\ \vdots \\ P_{i,m_i-1} \\ P_{im_i} \end{pmatrix} \\
&\triangleq \boldsymbol{B_0} + \boldsymbol{B}\boldsymbol{P_i}, \qquad (2.10)
\end{aligned}$$

where $\boldsymbol{P_i} = (P_{i1}, ..., P_{im_i})'$, $P_{ij} = I(T_i > t_{ij})$, $j = 1, ..., m_i$, and the notation $\triangleq$ denotes equal to by definition. By the linear relation (2.10), we have $E(\boldsymbol{\delta_i}) = \boldsymbol{B_0} + \boldsymbol{B}E(\boldsymbol{P_i})$, i.e. $\boldsymbol{\mu_{i(\delta)}} = \boldsymbol{B_0} + \boldsymbol{B}\boldsymbol{\mu_i}$, where $\boldsymbol{\mu_i}$ denotes $E(\boldsymbol{P_i})$, and then $\boldsymbol{\delta_i} - \boldsymbol{\mu_{i(\delta)}} = \boldsymbol{B}(\boldsymbol{P_i} - \boldsymbol{\mu_i})$. Also, $\boldsymbol{D_{i(\delta)}} = \partial\boldsymbol{\mu_{i(\delta)}}/\partial\boldsymbol{\theta}' = \boldsymbol{B}\partial\boldsymbol{\mu_i}/\partial\boldsymbol{\theta}' \triangleq \boldsymbol{B}\boldsymbol{D_i}$, where $\boldsymbol{D_i} = \partial\boldsymbol{\mu_i}/\partial\boldsymbol{\theta}'$, and $\boldsymbol{V_{i(\delta)}} = Var(\boldsymbol{\delta_i}) = \boldsymbol{B}Var(\boldsymbol{P_i})\boldsymbol{B}'$, and then we have $\boldsymbol{V_{i(\delta)}^{-1}} = (\boldsymbol{B}')^{-1}\boldsymbol{V_i^{-1}}\boldsymbol{B}^{-1}$, where $\boldsymbol{V_i} = Var(\boldsymbol{P_i})$. Thus, the quasi-likelihood score function (2.9) can be written alternatively as

$$\begin{aligned}
U(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \boldsymbol{D'_{i(\delta)}}\boldsymbol{V_{i(\delta)}^{-1}}[\boldsymbol{\delta_i} - \boldsymbol{\mu_{i(\delta)}}] \\
&= \sum_{i=1}^{n} \boldsymbol{D'_i}\boldsymbol{B}'(\boldsymbol{B}')^{-1}\boldsymbol{V_i^{-1}}\boldsymbol{B}^{-1}\boldsymbol{B}(\boldsymbol{P_i} - \boldsymbol{\mu_i}) \\
&= \sum_{i=1}^{n} \boldsymbol{D'_i}\boldsymbol{V_i^{-1}}(\boldsymbol{P_i} - \boldsymbol{\mu_i}), \qquad (2.11)
\end{aligned}$$

where $\boldsymbol{\mu_i} = E(\boldsymbol{P_i})$, $\boldsymbol{D_i} = \partial\boldsymbol{\mu_i}/\partial\boldsymbol{\theta}'$, and $\boldsymbol{V_i} = Var(\boldsymbol{P_i})$, $i = 1, ..., n$. Note that it is an extension of (2.3) with $\boldsymbol{c}(t_{ij}|\boldsymbol{A_i};\boldsymbol{\theta})$ replaced by the jth column of $\boldsymbol{D_i'V_i^{-1}}$ and $S_T(t_{ij}|\boldsymbol{A_i};\boldsymbol{\theta})$ replaced by the jth element of $\boldsymbol{\mu_i}$. This is also equivalent to a GEE for $\boldsymbol{P_i}$ with the true covariance matrix of $\boldsymbol{P_i}$ specified for the working covariance matrix.

So far, we have shown that given the visit times, $t_{i1}, ..., t_{im_i}$, likelihood (2.1) for interval-censored T_i and a GEE with the true covariance matrix for $\boldsymbol{P_i}$ correctly specified are identical and lead to the same estimator for $\boldsymbol{\theta}$. Furthermore, when the random inspection times t_{ij}, $j = 1, ..., m_i$, are independent of T_i given $\boldsymbol{A_i}$, maximizing the likelihood (2.1) or solving an estimating equation given in (2.3) gives a consistent estimator for parameter $\boldsymbol{\theta}$. However, if the random inspection times are outcome-dependent, T_i could be dependently interval-censored. Then, maximizing the likelihood (2.1) or solving (2.3) may lead to biased estimates. In the next section, we will introduce an inverse probability weighting method that can be applied to (2.3) or (2.6) so that the resulting estimates obtained by solving the weighted estimating functions are consistent, provided that an important assumption which will be introduced in Section 2.2.1 is satisfied and that both the outcome model and the weight model are correctly specified.

2.2 Methodology of Inverse-Intensity-of-Visit Weighted Estimation

We have introduced intermittent and outcome-dependent observation in Section 1.4.1 and Section 1.4.2. In principle, if observation times are discrete and finite and individuals have a common set of prespecified potential visit times, the outcome-dependent follow-up problem can be dealt with as a longitudinal missing data problem. Robins et al. (1995) presented the

assumption of sequentially ignorable nonresponse and proposed using weighted estimating equations for the adjustment of monotone missing responses, i.e. censoring, with the weight at time t defined by the inverse probability of the outcome being observed at t. Imputation and expectation-maximization (EM) algorithms based on a joint model for outcomes and visits are also commonly employed for longitudinal studies in the presence of missing data. In particular, Chen et al. (2010) studied the PsA data via a likelihood-based approach based on multi-state models. They assumed that subjects are scheduled to be examined at a common set of times, and then applied the EM algorithm to deal with the missing data at unattended visits.

Most literature dealing with missingness under continuous observation schemes in longitudinal studies, especially for failure time outcome, considers monotone missingness, i.e. random drop-out, first presented by Wu and Carroll (1988). A classification of drop-out processes was defined by Diggle and Kenward (1994): completely random drop-out (CRD), random drop-out (RD) and informative drop-out (ID), following the terminology in Rubin (1976), and a class of inverse-probability-of-censoring (IPC) weighted estimating functions were proposed by Robins (1993), Robins et al. (1995), Scharfstein et al. (1999), Robins and Finkelstein (2000), and Satten et al. (2001) for various outcome models or censoring time models. However, there are few papers about the intermittent observation scheme in continuous time. When observation occurs in continuous time, missing data techniques do not provide a useful method for the dependent observation problem, unless continuous observation times are discretized by grouping.

Outcome-dependent observation in continuous time was first addressed by Lipsitz et al. (2002) where they focused on the repeated measures following a multivariate Gaussian distribution. They separated the likelihood into two parts: one for the outcome process and one for the observation process and proposed that the latter can be ignored if it is

likelihood-based and all the common risk factors between the outcome and observation processes are conditioned on. Therefore, as we mentioned before, one way to adjust for dependent intermittent observation is to introduce all common risk factors between the outcome process and the visit process as covariates or stratifying variables in the analysis of outcomes. Other methods could be like Sun et al. (2005) where they marginally modelled the visit process and then modelled the outcome process conditional on visit history, but this gives conditional regression parameters given the visit history. Alternatively, most recent literature links the outcome process and the visit process by introducing common (shared) latent variable(s), e.g. Sun et al. (2007) and Cai et al. (2012), or correlated latent variable(s), e.g. Liang et al. (2009) and Sun et al. (2012), between these the two processes. Their methods produce estimates of the regression parameters in the outcome model that are conditional on unobservable latent variable(s), i.e. random effect(s). Moreover, most joint modelling approaches for dependent visits via random effect(s) assume time-invariant random effect(s), which are hardly plausible in many situations. The advantage of such joint modelling methods is that they can handle the cases where the outcome process and the visit process are correlated via unknown factors, as long as these effects are of the assumed form.

However, in many applications, the regression parameters of a model for outcomes conditional on ancillary variables are not the target of inference. Instead, a model for the outcome given a smaller set of "primary" covariates is of interest (e.g. Bůžková and Lumley (2007)). To estimate the marginal effect of a set of primary factors on outcomes in the presence of time-varying ancillary variables, a more appropriate and convenient way to adjust for outcome-dependent inspection times is weighting an estimating equation for outcomes which is conditional on the primary factors only by an inverse intensity of visit. Explicitly, Lin et al. (2004) considered the intermittent inspection times as a recurrent event

process in continuous time; the visit intensity at t may depend on the history of previous outcomes, previous visits and external covariates prior to t, in addition to the primary factors included in the marginal outcome model. Bůžková and Lumley (2007) extended their method to involve time-varying covariates and discontinuous visit intensities so that visit schemes with a mixture of scheduled discrete time visits and unplanned continuous time visits can be dealt with.

Inverse probability weighting (IPW) is a general estimating function methodology for informative selection, e.g. missingness, censoring, sampling in surveys, treatment assignment in causal inference, etc, when the selection mechanism is ignorable, i.e. at random, following the notions and terminologies of missing at random (MAR) by Rubin (1976) and coarsened at random (CAR) by Heitjan and Rubin (1991) and Jacobsen and Keiding (1995). In longitudinal studies, there is a sequential ignorability assumption (Robins and Rotnitzky, 1992; Hogan et al., 2004; Cook and Lawless, 2014), which states that the missingness of outcome at time t is independent of the current outcome value given the past history. This is similar to the important assumption which the inverse-intensity-of-visit (IIV) weighting relies on. We will discuss this in the next section. Inverse probability weighting is a very useful approach for outcome-dependent selection problems. It standardizes the selected data to the whole underlying population by weighting each observation with the inverse of the probability that this subject is selected from the population. Horvitz and Thompson (1952) applied the IPW idea in sampling contexts; later, it was applied in a variety of studies by Manski and Lerman (1977), Kalbfleisch and Lawless (1988), and Zhao and Lipsitz (1992); Robins (1993) showed that IPW can also handle dependent censoring; Robins et al. (1994, 1995) developed the IPW estimation for missing data; Robins et al. (2000) and Hernán et al. (2000) applied the IPW to adjust for confounders in observational studies with time-varying treatment. In the studies of dependent follow-up

times, Lin et al. (2004) and Bůžková and Lumley (2007, 2009) incorporated the intensity of observation or visit as a weight into the estimation of marginal association measures for irregularly observed longitudinal or repeated measures data, to adjust for ancillary variables associated with the outcome process as well as the observation process in randomized experiments. Pullenayegum and Feldman (2013) further introduced the IIV weighting to increment-based methods for irregularly observed longitudinal data and discussed the optimal truncation of IIV weights and a doubly robust IIV weighted estimator. In addition, Pullenayegum and Lim (2014) gave a comprehensive literature review on longitudinal data analysis with irregular observation.

Lin et al. (2004) and Bůžková and Lumley (2007)'s IIV weighting method focuses on repeated responses over time based on parametric or semi-parametric linear or generalized linear models. We aim to extend this approach to failure time data analysis, where failure time status is periodically monitored until a known occurrence of failure or loss to follow-up. In this sense, a known failure discontinues the visit process. For example, this can happen when an individual who has experienced the failure event is withdrawn from the cohort or switched to another cohort. Furthermore, if commonly used monotone measures of a failure time such as the CDF, $F_T(t) = E\{I(T \leq t)\}$, and survival function, $S_T(t) = E\{I(T > t)\}$ are targeted, monotonicity is a challenge for non-/semi-parametric estimation by using estimating equations.

Outcome-dependent visit times cause failure time to be "dependently interval-censored", which makes standard analysis methods for interval-censored data inappropriate. In addition to likelihood-based methods, other approaches for interval-censored data include multiple imputation, e.g. Pan (2000); Hsu et al. (2007); Chen and Sun (2010), but irregular visits which cause large variation of visit times make imputation difficult. Earlier, van der Laan and Hubbard (1997) and van der Laan and Robins (1998) proposed locally

efficient estimation for interval-censored data or current status data. If interval censoring is independent and survival models are parametric, likelihoods involving time-varying covariates can be constructed as in Sparling et al. (2006). Additionally, Finkelstein et al. (2002) and Zhang et al. (2007) developed an EM algorithm for dependently interval-censored data. Finally, intermittent visits could also cause failure times to be left-truncated if the time origin is set before the first visit, and then failures prior to the first visit are not included in the analysis.

To sum up, the prime advantages of the IIV weighting method is that regression parameters indexing the marginal associations between the factors of primary interest and outcomes can be consistently estimated when time-varying ancillary variables are adjusted for but not directly regressed in the outcome model. Additionally, weighting methods can be conveniently implemented by existing software such as R functions *lm*, *glm*, and *geeglm*. The main constraint is that this weighting method relies on an important assumption of conditional independence between the outcome process and the visit process given the observed history of known variables which we will discuss in the subsequent section. This condition is one that cannot be avoided without making assumptions that are uncheckable given the type of data we consider, and other approaches such as imputation and EM algorithms have the same or equivalent constraints.

2.2.1 Required Assumptions

Dependent observation arises when in addition to the covariates in the regression model for outcomes, there are still some factors related to the observation process as well as the outcome process. The values of these variables affect how often and when an individual comes for a visit. For example, in the PsA example, patients who have more severe joint

pain could be more likely to visit the clinic. Meanwhile, those patients who suffer more from joint pain may have a higher risk of joint damage. The common factors between the outcome process and the visit process could induce a selection bias, if we fail to properly adjust for them. These common factors may include baseline or time-varying treatments for PsA, baseline or time-varying biomarkers, and the history of previous outcomes and previous visits, etc.

Let an overbar denote the history of a variable, i.e. $\bar{Z}(t) = \{Z(s) : 0 \leq s \leq t\}$ is the full history of a time-varying variable $Z(s)$ up to and including time t, and let $\bar{Z}^{obs}(t)$ be the corresponding observed history. Then, define $\boldsymbol{\mathcal{H}_i^{obs}}(t^-) = \{\bar{P}_i^{obs}(t^-), \bar{N}_i(t^-), \boldsymbol{A_i}, \bar{\boldsymbol{L}}_{\boldsymbol{i}}^{\boldsymbol{obs}}(t^-)\}$ be the observed history, which includes not only the outcome model covariates, $\boldsymbol{A_i}$, but also the observed history of auxiliary external (time-varying) variables, $\bar{\boldsymbol{L}}_{\boldsymbol{i}}^{\boldsymbol{obs}}(t^-)$, and, importantly, the observed history of the outcome process, $\bar{P}_i^{obs}(t^-)$ where $P_i(t) = I(T_i > t)$, and history of the visit process, $\bar{N}_i(t^-)$. We can let $\bar{\boldsymbol{L}}_{\boldsymbol{i}}^{\boldsymbol{obs}}(t^-)$ be left-continuous, i.e. $\bar{\boldsymbol{L}}_{\boldsymbol{i}}^{\boldsymbol{obs}}(t^-) = \bar{\boldsymbol{L}}_{\boldsymbol{i}}^{\boldsymbol{obs}}(t)$ for all t. In general, $\boldsymbol{\mathcal{H}^{obs}}(t^-)$ can include the observed history of everything except the current outcome value and current visit status. Then, let $\boldsymbol{Z_i}(t^-) = \boldsymbol{h}\{\boldsymbol{\mathcal{H}_i^{obs}}(t^-)\}$ represent some features of the observed history $\boldsymbol{\mathcal{H}_i^{obs}}(t^-)$, where $\boldsymbol{h}(\cdot)$ is a vector of certain known functions. The target of inference is to estimate parameter $\boldsymbol{\theta}$ in a parametric model for T_i:

$$S_T(t|\boldsymbol{A_i}; \boldsymbol{\theta}) = Pr(T_i > t|A_i; \boldsymbol{\theta}) = E\{P_i(t)|\boldsymbol{A_i}\}. \tag{2.12}$$

For the adjustment of intermittent visits, we consider the following two conditions for the visit process:

(B0) Independent Observation Scheme

$$E\{dN_i(t)|\boldsymbol{A_i}, \boldsymbol{P_i}(t), C_i(t)\} = C_i(t)E\{dN_i(t)|\boldsymbol{A_i}\}, \quad \forall t > t_{i0}, \qquad \text{(a)}$$

$$E\{dN_i(t)|\boldsymbol{A_i}\} > 0, \quad \text{for all } \boldsymbol{A_i} \text{ and } t > t_{i0}. \qquad \text{(b)}$$

(B1) Conditionally Independent Observation Scheme

$$\begin{aligned} E\{dN_i(t)|\boldsymbol{\mathcal{H}_i^{obs}}(t^-), \boldsymbol{P_i}(t), C_i(t)\} &= C_i(t)E\{dN_i(t)|\boldsymbol{\mathcal{H}_i^{obs}}(t^-)\}, \quad \forall t > t_{i0}, \quad \text{(a)} \\ &= C_i(t)E\{dN_i(t)|\boldsymbol{Z_i}(t^-)\}, \end{aligned}$$

$$E\{dN_i(t)|\boldsymbol{Z_i}(t^-)\} > 0, \quad \text{for all } \boldsymbol{Z_i}(t^-) \text{ and } t > t_{i0}. \qquad \text{(b)}$$

Note $\boldsymbol{P_i}(t) = \{P_i(s) : s \geq t\}$ denotes the current and future outcomes, $C_i(t) = I(C_i > t)$ is the at-risk indicator at t where C_i is a random drop-out time, and $t_{i0} = 0$ if not stated otherwise, for $i = 1, ..., n$. When $t_{i0} > 0$, it indicates that T_i is left-truncated and t_{i0} is the delayed entry time. Condition B0 (b) and B1 (b) are referred to as positivity conditions which are needed to guarantee the existence of $n^{1/2}$-consistent estimators of $\boldsymbol{\theta}$ (Robins et al., 1995). In principle, we can weaken condition (b) in (B1) to allow $E\{dN_i(t)|\boldsymbol{Z_i}(t^-)\} = 0$ at certain t^- values, as long as this holds for all possible $\boldsymbol{Z_i}(t^-)$, but we will ignore this in our development. In addition, we need to assume that the visit process distribution and the outcome process distribution have distinct parameters, as for the ignorability of coarsening mechanism discussed by Heitjan and Rubin (1991). Condition (B1) was referred to as "sequential ignorability" by Robins and Rotnitzky (1992) and Robins et al. (1995) for assuming that nonresponse at time t is independent of current and future outcomes given the history through t^- for the case of monotone missing responses in discrete time. This is satisfied when data is missing at random in the sense of Rubin (1976) for the longitudinal setting. Pullenayegum and Lim (2014) referred to the condition (B1) as visiting at random (VAR), and a similar assumption was made for random drop-out by Diggle and Kenward (1994) which is essential for IPC weighting methods.

Condition (B0) indicates that given the covariates $\boldsymbol{A_i}$ controlled for in the targeted outcome model (2.12), visits are outcome-independent, which is an analog to the condition of "missing completely at random" (MCAR) by Rubin (1976). This was assumed for the methods discussed in Sections 2.1.1 and 2.1.2. Condition (B1) means that given some features of the observed past history, intermittent visits are ignorable. This is an analog to the "missing at random" (MAR) by Rubin (1976). If condition (B0) is not satisfied but (B1) holds, then consistent estimation of $\boldsymbol{\theta}$ still can be achieved via appropriate adjustment for $\boldsymbol{Z_i}(t^-)$.

Condition (B1) requires that visits depend only on the observed history and known factors. As Pullenayegum and Lim (2014) indicated, history-dependent protocol visits and physician-driven visits usually satisfy (B1), but patient-driven visits are likely to be not at random. In that case, (B1) may not hold, because a patient's decision about visit attendance may depend on some information which is not provided at past visits, e.g. the true history $\boldsymbol{\mathcal{H}}(t^-)$. Condition (B1) is essential for the adjustment of irregular visits by weighting, though it is usually untestable, like many other MAR conditions. If (B1) is violated, it means that irregular visits are non-ignorable, and it is similar to the case of "missing not at random", which means visit times depend on some unknown outcome-related factors. Missing not at random may result in weighting methods not being applicable.

When condition (B1) is satisfied, weighting with an inverse-intensity-of-visit is useful to adjust for informative past history so that characteristics are balanced between the observed data and the unobserved data, and the marginal regression parameter $\boldsymbol{\theta}$ in the targeted outcome model (2.12) can be estimated consistently. As Bůžková and Lumley (2009) emphasized, the outcome model covariates should be picked on scientific grounds to answer a question of particular interest, while the weight model must be determined by

the nature of the observation process. In this section, we will extend a class of inverse-intensity-of-visit (IIV) weighted estimators proposed by Lin et al. (2004) and Bůžková and Lumley (2007, 2009) for irregularly observed longitudinal data to failure time data based on parametric models.

Lin et al. (2004) considered the visit or observation process $\{N_i(t), t > 0\}$ as a continuous time recurrent event process. As in Lin et al. (2004) and Bůžková and Lumley (2007, 2009), we can consider a continuous-time parametric or semi-parametric intensity model for the visit process conditional on $\boldsymbol{Z_i}(t^-)$, with intensity denoted by $d\Lambda_N(t|\boldsymbol{Z_i}(t^-); \boldsymbol{\alpha}) = E\{dN_i(t)|\boldsymbol{Z_i}(t^-)\}$, provided that condition (B1) holds. We will further discuss the modelling of visit times and the estimation of IIV weights in Section 2.2.3.

2.2.2 Weighted Estimating Functions

As mentioned before, if (B0) holds, no weighting is required to adjust for visit times, given the targeted outcome model $S_T(t|\boldsymbol{A_i}; \boldsymbol{\theta}) = \exp\left[-\int_0^t \lambda_T(s|\boldsymbol{A_i}; \boldsymbol{\theta})ds\right]$ is correctly specified. However, if (B0) is not satisfied, (2.3) or (2.6) takes only the observed outcomes, i.e. when $dN_i^*(t) = 1$, into account and the intermittent inspection times predict which values of outcomes would be observed. Thus, the resulting estimators obtained by solving (2.3) or (2.6) could be inconsistent. However, if (B0) is not satisfied, but (B1) holds, then weighting the observed outcomes by the inverse of the probability or intensity of being observed at that time can balance the characteristics between observed and unobserved outcomes, and as a result, selection bias caused by dependent observation is eliminated. First, consider the case where there is no left truncation of T_i. By solving the following unbiased estimating function weighted by the inverse-intensity-of-visit, we can obtain a consistent estimator of $\boldsymbol{\theta}$ under certain regularity conditions, provided that the outcome model and weight model

are both correctly specified. We consider

$$\boldsymbol{U}^w(\boldsymbol{\theta},\boldsymbol{\alpha}) = \sum_{i=1}^{n} \boldsymbol{U}_i^w(\boldsymbol{\theta},\boldsymbol{\alpha}) = \sum_{i=1}^{n} \int_0^{\tau_i} w_i(t;\boldsymbol{\alpha}) \boldsymbol{c}(t|\boldsymbol{A_i};\boldsymbol{\theta})[P_i(t) - S_T(t|\boldsymbol{A_i};\boldsymbol{\theta})] dN_i^*(t), \quad (2.13)$$

where the weight $w_i(t;\boldsymbol{\alpha})$ is defined by

$$\begin{aligned} w_i(t;\boldsymbol{\alpha}) &= a(t|\boldsymbol{A_i})dt / E[dN_i(t)|\boldsymbol{Z_i}(t^-)], \qquad \forall t > 0 \\ &= a(t|\boldsymbol{A_i}) / \lambda_N(t|\boldsymbol{Z_i}(t^-);\boldsymbol{\alpha}), \end{aligned} \quad (2.14)$$

where $a(t|\boldsymbol{A_i})$ is a stabilizing weight; we note that if $\boldsymbol{c}_i(t)$ in the proof given below is multiplied by an additional time-varying function $a(t|\boldsymbol{A_i})$, (2.13) is still an unbiased estimating function. We let $a(t|\boldsymbol{A_i}) = 1$ in the following discussion of the thesis, unless stated otherwise. Notation, e.g. $P_i(t)$, $dN_i^*(t)$, and $\boldsymbol{c}(t|\boldsymbol{A_i};\boldsymbol{\theta})$, is defined as in (2.3), and $\lambda_N(t)$ is the intensity of the visit process at time t. To show that (2.13) is an unbiased estimating function, one needs to show $E[\boldsymbol{U}_i^w(\boldsymbol{\theta},\boldsymbol{\alpha})|\boldsymbol{A_i}] = 0$. For simplicity, we consider $\boldsymbol{A_i}$ as known constants and suppress $\boldsymbol{A_i}$ and the parameter notation, $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$, in the following proof,

e.g. writing $\boldsymbol{c_i}(t)$ for $\boldsymbol{c}(t|A_i;\boldsymbol{\theta})$, $S_{Ti}(t) = S_T(t|A_i;\boldsymbol{\theta})$, etc. The required expectation is then:

$$
\begin{aligned}
& E\left\{\int_0^{\tau_i} \frac{1}{\lambda_N(t|\boldsymbol{Z_i}(t^-))}\boldsymbol{c}_i(t)[P_i(t) - S_{Ti}(t)]dN_i^*(t)\right\} \\
&= E\left\{\int_0^{\tau_i} \frac{1}{\lambda_N(t|\boldsymbol{Z_i}(t^-))}\boldsymbol{c}_i(t)[P_i(t) - S_{Ti}(t)]C_i(t)dN_i(t)\right\} \\
&= \int_0^{\tau_i} \boldsymbol{c}_i(t)E_{Z,P,C}\left\{1/\lambda_N(t|\boldsymbol{Z_i}(t^-))[P_i(t) - S_{Ti}(t)]C_i(t)E\left[dN_i(t)|\boldsymbol{\mathcal{H}_i^{obs}}(t^-), P_i(t), C_i(t)\right]\right\} \\
&= \int_0^{\tau_i} \boldsymbol{c}_i(t)E_{Z,P,C}\left\{1/\lambda_N(t|\boldsymbol{Z_i}(t^-))[P_i(t) - S_{Ti}(t)]C_i(t)E\left[dN_i(t)|\boldsymbol{\mathcal{H}_i^{obs}}(t^-)\right]\right\} \qquad \text{by (B1)} \\
&= \int_0^{\tau_i} \boldsymbol{c}_i(t)E_{Z,P,C}\left\{1/\lambda_N(t|\boldsymbol{Z_i}(t^-))[P_i(t) - S_{Ti}(t)]C_i(t)\lambda_N(t|\boldsymbol{Z_i}(t^-))dt\right\} \\
&= \int_0^{\tau_i} \boldsymbol{c}_i(t)E_{P,C}\left\{[P_i(t) - S_{Ti}(t)]C_i(t)\right\}dt \\
&= \int_0^{\tau_i} \boldsymbol{c}_i(t)E[P_i(t) - S_{Ti}(t)]E[C_i(t)]dt \qquad \text{by independent drop-out} \\
&= \int_0^{\tau_i} \boldsymbol{c}_i(t)[S_{Ti}(t) - S_{Ti}(t)]E[C_i(t)]dt \\
&= \mathbf{0}
\end{aligned}
$$

as desired. Note $C_i(t) = I(C_i > t)$ is the at risk indicator as defined earlier, and E with subscripts denotes the expectation with respect to relevant variables. The last third line depends on the assumption that random drop-out is independent of outcome given $\boldsymbol{A_i}$, i.e. $P_i(t) \amalg C_i(t)|\boldsymbol{A_i}$, $\forall t > 0$. Otherwise, dependent drop-out should be adjusted for as well. A so-called inverse-probability-of-censoring (IPC) weighted estimator can be applied in that case.

Secondly, when event time T_i is left-truncated, the estimating function (2.6) can be weighted by the inverse-intensity-of-visit as well, i.e.

$$\boldsymbol{U}^w(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \int_{t_{i0}}^{\tau_i} w_i(t;\boldsymbol{\alpha})\boldsymbol{c}(t|\boldsymbol{A_i};\boldsymbol{\theta}, t_{i0})[P_i(t) - S_T(t|\boldsymbol{A_i};\boldsymbol{\theta})/S_T(t_{i0}|\boldsymbol{A_i};\boldsymbol{\theta})]dN_i^*(t), \quad (2.15)$$

where $P_i(t)$, $dN_i^*(t)$ and $\boldsymbol{c}(t|\boldsymbol{A_i};\boldsymbol{\theta}, t_{i0})$ are defined as in (2.6) and $w_i(t;\boldsymbol{\alpha}) = 1/\lambda_N(t|\boldsymbol{Z_i}(t^-);\boldsymbol{\alpha})$,

for all $t > t_{i0}$. Note that, here $\boldsymbol{A_i}$ is a vector of time-fixed variables defined at $t = 0$. Assuming that condition (B1) is true, and random drop-out (i.e. $P_i(t) \amalg C_i(t)|\boldsymbol{A_i}$, for $t > t_{i0}$) and left truncation (i.e. $t_{i0} \amalg T_i|\boldsymbol{A_i}$) are independent, the unbiasedness of estimating function (2.15) can be proven in a similar manner as for (2.13).

2.2.3 Models for the Visit Process and Estimation of IIV Weights

The IIV weighted estimation is a two-step procedure. First, we fit a model for the visit process to estimate weights. Then, by solving the weighted estimating equation (2.13) or (2.15) with the estimated IIV weights, a consistent estimator of the parameter $\boldsymbol{\theta}$ from the outcome model $S_T(t|\boldsymbol{A_i}, \boldsymbol{\theta})$ can be obtained under mild regularity conditions, provided that the outcome process model and the visit process model are correct.

In the first step, to estimate weights, we need to assume a model for the visit process. First, we introduce a semi-parametric proportional intensities model employed by Lin et al. (2004) and Bůžková and Lumley (2007, 2009). The visit intensity is assumed to be of the form:

$$\begin{aligned} d\Lambda_N(t|\boldsymbol{Z_i}(t^-); \boldsymbol{\alpha}) &= E\{dN_i(t)|\boldsymbol{Z_i}(t^-)\} \\ &= d\Lambda_{N0}(t)\ \exp\{\boldsymbol{Z_i}(t^-)'\boldsymbol{\alpha}\} \\ &= \lambda_{N0}(t)\ \exp\{\boldsymbol{Z_i}(t^-)'\boldsymbol{\alpha}\}dt \end{aligned} \tag{2.16}$$

where $\lambda_{N0}(t)$ is an unspecified non-negative function of t, which is known as the baseline visit intensity function. More intensity-based models and theories for recurrent events can be found in Cook and Lawless (2007).

Then, the weight for subject i at time t can be estimated by

$$w_i(t; \widehat{\boldsymbol{\alpha}}, \widehat{\lambda}_{N0}) = \frac{1}{\widehat{\lambda}_{N0}(t)\ \exp(\boldsymbol{Z_i}(t^-)'\widehat{\boldsymbol{\alpha}})}, \tag{2.17}$$

or

$$w_i(t;\widehat{\boldsymbol{\alpha}}) = \frac{1}{\exp(\boldsymbol{Z_i}(t^-)'\widehat{\boldsymbol{\alpha}})}, \tag{2.18}$$

where $\widehat{\boldsymbol{\alpha}}$ can be obtained by maximizing the partial likelihood for the Cox model or by using existing software for the Cox model such as R function *coxph* or SAS procedure *PHREG*. Lin et al. (2004) proposed a kernel-smoothed Breslow estimator to estimate the baseline intensity $\lambda_{N0}(t)$ in formula (2.17). However, Bůžková and Lumley (2007) suggested that baseline intensity $\lambda_{N0}(t)$ can be exempted from the weight formula, i.e. they proposed to use (2.18). This is analogous to introducing a stabilizing weight $a(t|\boldsymbol{A_i}) = \lambda_0(t)$. Furthermore, omitting $\widehat{\lambda}_{N0}(t)$ in weight estimation can avoid the smoothing techniques applied to $\widehat{\lambda}_{N0}(t)$ and achieve $\sqrt{n}$ consistency of the final estimator of $\boldsymbol{\theta}$.

So far, we have been assuming that the visit process (2.16) satisfies a modulated Markov proportional intensities assumption given covariates $\boldsymbol{Z_i}(t^-)$. However, in the PsA example, the clinic visits were nominally scheduled to be a certain length of time apart (e.g. 6 months), but actual visit gaps often deviated substantially from this. In this case, modeling the gap times or the inter-arrival times between consecutive clinic visits may be more plausible. That is, the visit intensity at t could be related to the elapsed time since the most recent visit prior to t. An alternative assumption would be that, given the most recent past visit time $T_{N(t^-)}$, visit intensity has the form of a modulated renewal process (or semi-Markov process) (Cook and Lawless, 2007), i.e.

$$\lambda_N(t|\boldsymbol{Z_i}(t^-);\boldsymbol{\alpha}) = \lambda^{\dagger}_{N0}(B(t))\,\exp(\boldsymbol{\alpha}'\boldsymbol{Z_i}(t^-)), \tag{2.19}$$

where t is the chronological time, e.g. time from the clinic entry in PsA data (no left truncation case) or the onset of PsA (left truncation case), and $B(t)$ is the gap time or elapsed time from the most recent past visit, i.e. $B(t) = t - T_{N(t^-)}$. Once again $\boldsymbol{Z_i}(t^-)$ may include the features of prior visit history and outcome history. When we estimate

the IIV weights based on (2.19), one difference from the former visit model (2.16) is that the renewal baseline intensities $\lambda^{\dagger}_{N0}(B(t))$ can not be exempted from the weight formula, because $\lambda^{\dagger}_{N0}(B(t))$ depends on the individual's previous visit time $T_{N(t^-)}$ which is random and also informative for outcomes due to variables other than $\boldsymbol{A_i}$. Thus, $\lambda^{\dagger}_{N0}(B(t))$ also needs to be estimated while estimating the weights $w_i(t)$, i.e.

$$w_i(t; \widehat{\boldsymbol{\alpha}}, \widehat{\lambda}^{\dagger}_{N0}) = \frac{1}{\widehat{\lambda}^{\dagger}_{N0}[B(t)] \exp(\boldsymbol{Z_i}(t^-)'\widehat{\boldsymbol{\alpha}})}. \tag{2.20}$$

One way is to adopt the Breslow nonparametric estimator, which can be implemented by most software, and then smooth the resulting baseline estimate to satisfy positivity condition and achieve a certain convergence rate, like Lin et al. (2004) did. Also, parametric visit time models can be considered to avoid smoothing of the baseline estimate. For example, in the simulation studies we will show in Section 2.3.1, a semi-Markov gap times model (2.19) with a piecewise-constant baseline is considered. Furthermore, in Section 2.2.5, large sample theory of the proposed estimator of $\boldsymbol{\theta}$ from a parametric outcome model with the IIV weights estimated by fitting a parametric semi-Markov visit gap time model (2.19) will be discussed.

2.2.4 Discrete Observation Process Models

So far, we assume that the occurrence of failure does not terminate the visit process, e.g. events of interest such as time to joint damage, time to relapse after surgery, or time to the appearance of a tumor. Individuals are followed up to the end of follow-up, even if it is known that she/he has already experienced a failure. In this section, we will consider the case where a known failure occurrence terminates the visit process.

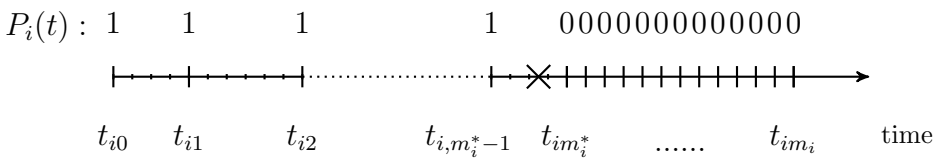


Figure 2.1: Graphical demonstration of a continuous failure time under a discrete visit process, with a potential visit time increment of 0.25 units, where $\{t_{i1}, ..., t_{im_i}\} = \{t > t_{i0} : dN_i^*(t) = 1\}$; $t_{i1}, ..., t_{im_i^*}$ are the actual visits and $t_{i,m_i^*+1}, ..., t_{im_i}(= C_i)$ are the pseudo visits after failure occurrence; t_{i0} denotes the start of follow-up for subject i; responses $P_i(t)$ at $\{t_{i0}, t_{i1}, ..., t_{im_i}\}$ are given above the time axis; and the cross denotes a failure occurrence.

We assume that after a known failure, individuals discontinue being inspected. That is, the "probability" of attending a visit at any t past the known failure occurrence is zero. In this sense, the positivity assumptions in conditions (B0) and (B1) discussed in Section 2.2.1 are violated. On the other hand, we know a person's survival status: $P_i(t) = 0$ for all $t > T_i$. In this case, it is necessary to artificially continue "observation" of the individual so as to satisfy the positivity condition. One approach would be to randomly generate pseudo visit times from the visit process model, and then apply the IIV weighted estimating function approach as introduced in Section 2.2.2. A simpler alternative, which we adopt, is to discretize the visit process and assume an individual "visits" at each possible time following their observed failure, i.e. let $dN_i(t) = 1$ when $\bar{P}_i^{obs}(t^-) = 0$, with probability 1 up to the end of follow-up to create pseudo visits. Suppose there are M_i prespecified potential visit times, $t_{i0} = a_{i0} < a_{i1} < ... < a_{iM_i} = C_i$, for subject i, where $t_{i0} \geq 0$ is the start of follow-up and $t_{i0} > 0$ indicates that T_i is left-truncated. In Figure 2.1, we set the time increment $\Delta a = 0.25$ for illustration, but M_i should be a fairly large integer to make the grid fine enough to approximate a continuous visit process in cases where ties of visit times rarely exist or we want to use existing software for continuous failure time

or recurrent event time data to fit a model to estimate the weights. An actual continuous visit time t_{ij} which falls in $(a_{ik}, a_{i,k+1}]$ will be carried forward to the upper bound $a_{i,k+1}$, $k = 0, ..., M_i - 1$. Since $P(t) = I(T > t)$ is a monotone response from 1 to 0, it is obvious that after the visit subsequent to the occurrence of failure, e.g. the visit t_{i,m_i^*} in Figure 2.1, $P(t) = 0$ is known. Therefore, $dN_i(t)$ can be considered equal to 1 with probability 1 between the first visit after failure occurrence (i.e. t_{i,m_i^*}) and the drop-out time (i.e. C_i). That is, we consider the visits $t_{i,1}, ..., t_{i,m_i^*}$ as actual visits and $t_{i,m_i^*+1}, ..., t_{im_i}$ as pseudo visits. Because the distribution of the visit process changes after failure occurrence, i.e. visits are associated with the observed outcome history, condition (B0) which states that visits and outcomes are correlated only through $\boldsymbol{A_i}$ does not hold in this case. However, condition (B1) can be modified as follows:

(B1*) Let $\bar{P}_i^{obs}(a_{i,k-1})$ be the most recently observed status of response P_i through $a_{i,k-1}$, and then when $\bar{P}_i^{obs}(a_{i,k-1}) = 1$ for any $k = 1, ..., M_i$, we assume

$$\begin{aligned} E\{dN_i(a_{ik})|\boldsymbol{\mathcal{H}_i^{obs}}(a_{i,k-1}), \boldsymbol{P_i}(a_{ik}), C_i(a_{ik})\} &= C_i(a_{ik})E\{dN_i(a_{ik})|\boldsymbol{\mathcal{H}_i^{obs}}(a_{i,k-1})\} \qquad \text{(a1)} \\ &= C_i(a_{ik})E\{dN_i(a_{ik})|\boldsymbol{Z_i}(a_{i,k-1})\}, \end{aligned}$$

$$E\{dN_i(a_{ik})|\boldsymbol{Z_i}(a_{i,k-1})\} > 0, \quad \text{for all } \boldsymbol{Z_i}(a_{i,k-1}), \qquad \text{(b)}$$

and when $\bar{P}_i^{obs}(a_{i,k-1}) = 0$, we have

$$E\{dN_i(a_{ik})|\boldsymbol{\mathcal{H}_i^{obs}}(a_{i,k-1}), \boldsymbol{P_i}(a_{ik}), C_i(a_{ik})\} = C_i(a_{ik}), \qquad \text{(a2)}$$

to create pseudo visits, where $\boldsymbol{P_i}(a_{ik})$, $C_i(a_{ik})$, $dN_i(a_{ik})$, and $\boldsymbol{Z_i}(a_{i,k-1}) = \boldsymbol{h}\{\boldsymbol{\mathcal{H}_i^{obs}}(a_{i,k-1})\}$ are defined similarly as in (B1), and we note that $\boldsymbol{\mathcal{H}_i^{obs}}(a_{ik}^-) = \boldsymbol{\mathcal{H}_i^{obs}}(a_{i,k-1})$ in the discrete time setting. If (B1*) is satisfied, we still say the observation scheme is conditionally independent, and we see the independence is conditional on the observed history of outcome, i.e. $\bar{P}_i^{obs}(a_{i,k-1})$.

We can still consider the following two visit process models, a modulated Markov proportional intensities model (2.21) or a modulated renewal (semi-Markov) process model (2.22) given below, to estimate the IIV weights, when $\bar{P}_i^{obs}(a_{i,k-1}) = 1$, for $k = 1, ..., M_i$:

$$\lambda_N(a_{ik}; \boldsymbol{\alpha}) = \lambda_{N0}(a_{ik}) \exp\{\boldsymbol{Z_i}(a_{i,k-1})'\boldsymbol{\alpha}\}, \tag{2.21}$$

or

$$\lambda_N(a_{ik}; \boldsymbol{\alpha}) = \lambda_{N0}^{\dagger}(B(a_{ik})) \exp(\boldsymbol{Z_i}(a_{i,k-1})'\boldsymbol{\alpha}), \tag{2.22}$$

where $\lambda_N(a_{ik}; \boldsymbol{\alpha})$ denotes a discrete visit intensity at a_{ik}. The IIV weights are defined by $\widehat{w}_i(a_{ik}) = 1/\lambda_N(a_{ik}; \widehat{\boldsymbol{\alpha}})$, when $\bar{P}_i^{obs}(a_{i,k-1}) = 1$, and by 1, when $\bar{P}_i^{obs}(a_{i,k-1}) = 0$, if the weight estimate is given by (2.17) based on model (2.21) or by (2.20) based on model (2.22). Note that if model (2.21) is assumed for the visit process and $\widehat{w}_i(a_{ik})$ is formulated as (2.18) when $\bar{P}_i^{obs}(a_{i,k-1}) = 1$, then $\widehat{w}_i(a_{ik}) = \widehat{\lambda}_{N0}(a_{ik})$ when $\bar{P}_i^{obs}(a_{i,k-1}) = 0$.

The IIV weighted estimating functions (2.13) and (2.15) can be defined for a discrete visit process by

$$\begin{aligned} \boldsymbol{U}(\boldsymbol{\theta}; \boldsymbol{\alpha}) &= \sum_{i=1}^{n} \sum_{k=1}^{M_i} w_i(a_{ik}; \boldsymbol{\alpha}) \boldsymbol{c}(a_{ik}|\boldsymbol{A_i}; \boldsymbol{\theta}, t_{i0})[P_i(a_{ik}) - \mu_T(a_{ik}|\boldsymbol{A_i}; \boldsymbol{\theta}, t_{i0})] dN_i^*(a_{ik}), \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_i(t_{ij}; \boldsymbol{\alpha}) \boldsymbol{c}(t_{ij}|\boldsymbol{A_i}; \boldsymbol{\theta}, t_{i0})[P_i(t_{ij}) - \mu_T(t_{ij}|\boldsymbol{A_i}; \boldsymbol{\theta}, t_{i0})], \end{aligned} \tag{2.23}$$

where $dN_i^*(a_{ik}) = dN_i(a_{ik}) I(C_i \geq a_{ik})$ and $t_{i1} < t_{i2} < ... < t_{im_i}$ are the m_i visits, including the m_i^* actual visit times before failure occurrence and the $m_i - m_i^*$ pseudo visit times after failure occurrence, among the M_i potential visit times, $a_{i1}, ..., a_{iM_i}$, for subject i. The mean function $\mu_T(a_{ik}|\boldsymbol{A_i}; \boldsymbol{\theta}, t_{i0})$ is defined by $S_T(a_{ik}|\boldsymbol{A_i}; \boldsymbol{\theta})/S_T(t_{i0}|\boldsymbol{A_i}; \boldsymbol{\theta})$, where $t_{i0} \geq 0$ is the start of follow-up of subject i, and when $t_{i0} > 0$ indicates that T_i is left-truncated and $t_{i0} = 0$ otherwise.

2.2.5 Large Sample Theory

In this section, we will discuss the asymptotic distribution of the proposed IIV weighted estimator $\widehat{\boldsymbol{\theta}}$ for a parametric failure time model where T_i may be subject to dependent interval censoring. In the following, the case where failure occurrence does not stop visits and the visit process is continuous will be considered. The case where failure terminates visits and the visit process is discretized will be discussed at the end.

Let $\boldsymbol{U}_{i1}(\boldsymbol{\theta}, \boldsymbol{\alpha})$ and $\boldsymbol{U}_{i2}(\boldsymbol{\alpha})$ be the estimating functions contributed by the outcome process and the visit process of subject i, respectively, and let $\boldsymbol{U}_i(\boldsymbol{\theta}, \boldsymbol{\alpha}) = (\boldsymbol{U}'_{i1}(\boldsymbol{\theta}, \boldsymbol{\alpha}), \boldsymbol{U}'_{i2}(\boldsymbol{\alpha}))'$ be the combined estimating functions for all the parameters, $(\boldsymbol{\alpha}', \boldsymbol{\theta}')'$.

First, $\boldsymbol{U}_{i1}(\boldsymbol{\theta}, \boldsymbol{\alpha})$ for the outcome process can be written as

$$\boldsymbol{U}_{i1}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{j=1}^{m_i} w_i(t_{ij}; \boldsymbol{\alpha}) \boldsymbol{c}(t_{ij} | \boldsymbol{A_i}; \boldsymbol{\theta}, t_{i0}) [P_i(t_{ij}) - \mu_T(t_{ij} | \boldsymbol{A_i}; \boldsymbol{\theta}, t_{i0})], \tag{2.24}$$

where $t_{i0} \geq 0$ is the start of follow-up for subject i, and the IIV weight $w_i(t_{ij}; \boldsymbol{\alpha}) = [\lambda_N(t_{ij} | \boldsymbol{Z_i}(t_{ij}^-); \boldsymbol{\alpha})]^{-1}$ is defined in (2.14). The above estimating function can be applied for the left truncation case where some t_{i0}'s are greater than zero, and the mean function $\mu_T(t | \boldsymbol{A_i}; \boldsymbol{\theta}, t_{i0}) = E[P_i(t) | T_i > t_{i0}, \boldsymbol{A_i}; \boldsymbol{\theta}] = S_T(t | \boldsymbol{A_i}; \boldsymbol{\theta}) / S_T(t_{i0} | \boldsymbol{A_i}; \boldsymbol{\theta})$ reduces to $S_T(t | \boldsymbol{A_i}; \boldsymbol{\theta})$ when $t_{i0} = 0$ for the no left truncation case.

Secondly, for the visit process, assume that $t_{i1} < ... < t_{im_i}$ are the m_i intermittent visit times for subject i. Here, we consider a parametric semi-Markov model (2.19) for illustration, but any parametric recurrent event or gap time model can be employed, e.g. the proportional intensities model (2.16) if $\lambda_{N0}(t)$ is parametric. Let $\boldsymbol{U}_{i2}(\boldsymbol{\alpha})$ be the log likelihood contribution of the m_i+1 visit gap times based on model (2.19) with a parametric baseline $\lambda^{\dagger}_{N0}(B(t); \boldsymbol{\alpha_0})$, i.e.

$$\lambda_N(t | \boldsymbol{Z_i}(t^-); \boldsymbol{\alpha}) = \lambda^{\dagger}_{N0}[B(t); \boldsymbol{\alpha_0}] \exp(\boldsymbol{\alpha}'_{\boldsymbol{1}} \boldsymbol{Z_i}(t^-)), \tag{2.25}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_0', \boldsymbol{\alpha}_1')'$, and $B(t)$ denotes the gap time at t from the previous visit, e.g. $B(t_{i,j+1}) = t_{i,j+1} - t_{ij}$. We notice that visit gap times are exactly observed except for the last one, which is right-censored at the end of follow-up, i.e. C_i. Therefore, the score function based on the above semi-Markov model (2.25) is given by the log likelihood (4.47) in Cook and Lawless (2007) for general parametric multiplicative intensity models for gap times:

$$\boldsymbol{U}_{i2}(\boldsymbol{\alpha}) = \sum_{j=1}^{m_i} \frac{\partial ln\{\lambda[B(t_{ij})]\}}{\partial \boldsymbol{\alpha}} - \sum_{j=1}^{m_i+1} \int_0^{B(t_{ij})} \frac{\partial \lambda(s)}{\partial \boldsymbol{\alpha}} ds, \tag{2.26}$$

where $\lambda(B(t)) = \lambda(B(t); \boldsymbol{\alpha}) = \lambda_{N0}^{\dagger}[B(t); \boldsymbol{\alpha_0}] \exp(\boldsymbol{\alpha_1'} \boldsymbol{Z_i}(t^-))$ and $t_{i,m_i+1} = C_i$.

Since $\boldsymbol{U}_{i1}(\boldsymbol{\theta}, \boldsymbol{\alpha})$ and $\boldsymbol{U}_{i2}(\boldsymbol{\alpha})$ are both unbiased estimating functions, given the condition (B1) is true and both the outcome process and the visit process are correctly modelled, by Theorem 2.2 and Theorem 3.2 in White (1982), we have the following theorem:

Theorem 1. *Given the regularity conditions in White (1982),* $(\widehat{\boldsymbol{\theta}}', \widehat{\boldsymbol{\alpha}}')'$ *by solving the IIV weighted estimating equations* $\sum_{i=1}^n \boldsymbol{U}_i(\boldsymbol{\theta}, \boldsymbol{\alpha}) = (\sum_{i=1}^n \boldsymbol{U}_{i1}'(\boldsymbol{\theta}, \boldsymbol{\alpha}), \sum_{i=1}^n \boldsymbol{U}_{i2}'(\boldsymbol{\alpha}))' = \boldsymbol{0}$*, where* $\boldsymbol{U}_{i1}(\boldsymbol{\theta}, \boldsymbol{\alpha})$ *and* $\boldsymbol{U}_{i2}(\boldsymbol{\alpha})$ *are defined by (2.24) and (2.26) respectively, is a consistent estimator of* $(\boldsymbol{\theta}', \boldsymbol{\alpha}')'$ *and is asymptotically normal distributed with covariance matrix* $\Gamma^{-1}\Sigma\ (\Gamma^{-1})'$*, where* Γ *and* Σ *are defined by*

$$\Gamma = \begin{pmatrix} -E(\partial \boldsymbol{U}_{i1}/\partial \boldsymbol{\theta}') & -E(\partial \boldsymbol{U}_{i1}/\partial \boldsymbol{\alpha}') \\ \boldsymbol{0} & -E(\partial \boldsymbol{U}_{i2}/\partial \boldsymbol{\alpha}') \end{pmatrix} \triangleq \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \boldsymbol{0} & \Gamma_{22} \end{pmatrix},$$

and

$$\Sigma = \begin{pmatrix} E(\boldsymbol{U}_{i1}\ \boldsymbol{U}_{i1}') & E(\boldsymbol{U}_{i1}\ \boldsymbol{U}_{i2}') \\ E(\boldsymbol{U}_{i2}\ \boldsymbol{U}_{i1}') & E(\boldsymbol{U}_{i2}\ \boldsymbol{U}_{i2}') \end{pmatrix} \triangleq \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}.$$

Then, the IIV estimator $\widehat{\boldsymbol{\theta}}$ *has the following asymptotic distribution*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} N(0, V_\theta),$$

where

$$V_\theta = \Gamma_{11}^{-1}(\Sigma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Sigma_{12}' - \Sigma_{12}\Gamma_{22}^{-1}\Gamma_{12}' + \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{12}')\ \Gamma_{11}^{-1}. \tag{2.27}$$

If estimating function (2.24) is a GEE with an independent working covariance matrix and is written in a matrix form by

$$\boldsymbol{U}_{i1} = \boldsymbol{D}_{\boldsymbol{i}}'\boldsymbol{V}_{\boldsymbol{i}}^{-1}\boldsymbol{W}_{\boldsymbol{i}}[\boldsymbol{P}_{\boldsymbol{i}} - \boldsymbol{\mu}_{\boldsymbol{i}}],$$

where $\boldsymbol{P}_{\boldsymbol{i}} = (P_i(t_{i1}), ..., P_i(t_{im_i}))'$, $\boldsymbol{\mu}_{\boldsymbol{i}} = (\mu_{i1}, ..., \mu_{im_i})'$, $\mu_{ij} = \mu_T(t_{ij}|\boldsymbol{A}_{\boldsymbol{i}}; \boldsymbol{\theta}, t_{i0})$, $\boldsymbol{D}_{\boldsymbol{i}} = \partial\boldsymbol{\mu}_{\boldsymbol{i}}/\partial\boldsymbol{\theta}'$, $\boldsymbol{V}_{\boldsymbol{i}} = \text{diag}\{\mu_{i1}(1-\mu_{i1}), ..., \mu_{im_i}(1-\mu_{im_i})\}$, and $\boldsymbol{W}_{\boldsymbol{i}} = \text{diag}\{w_i(t_{i1}; \boldsymbol{\alpha}), ..., w_i(t_{im_i}; \boldsymbol{\alpha})\}$, then we have

$$\widehat{\Gamma}_{11} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{D}_{\boldsymbol{i}}'\boldsymbol{V}_{\boldsymbol{i}}^{-1}\boldsymbol{W}_{\boldsymbol{i}}\ \boldsymbol{D}_{\boldsymbol{i}}\big|_{\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\theta}}}, \tag{2.28}$$

and

$$\widehat{\Gamma}_{12} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \lambda^{-2}(B(t_{ij}); \widehat{\boldsymbol{\alpha}})\ \frac{\partial\lambda(B(t_{ij}); \boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}'}\bigg|_{\widehat{\boldsymbol{\alpha}}} \boldsymbol{c}(t_{ij}|\boldsymbol{A}_{\boldsymbol{i}}; \widehat{\boldsymbol{\theta}}, t_{i0})[P_i(t_{ij}) - \mu_T(t_{ij}|\boldsymbol{A}_{\boldsymbol{i}}; \widehat{\boldsymbol{\theta}}, t_{i0})]. \tag{2.29}$$

Additionally, since $\boldsymbol{U}_{i2}$ is likelihood-based, $\Sigma_{22} = \Gamma_{22}$ is the Fisher information matrix of $\boldsymbol{U}_{i2}$. If we employ likelihood-based software to estimate $\boldsymbol{\alpha}$ in $\boldsymbol{U}_{i2}$, e.g. R function *phreg* with a piecewise constant baseline intensity, then Σ_{22} and Γ_{22} can be estimated by the observed Fisher information matrix reported by the software divided by sample size n. Finally, we can use $\widehat{\Sigma}_{11} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{U}_{i1}\ \boldsymbol{U}_{i1}'$ and $\widehat{\Sigma}_{12} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{U}_{i1}\ \boldsymbol{U}_{i2}'$.

Now, consider the case where failure occurrence terminates the visit process as we have discussed in the preceding section. The IIV weighting approach is still applicable if we discretize the visit process and set the weights equal to 1 when $\bar{P}_i^{obs}(t^-) = 0$. Then the asymptotic theory discussed above still can be applied. For the outcome process, the IIV

weighted estimating function is (2.23) and the required assumption is (B1*). For the visit process, only the visits that occurred when $P_i^{obs}(t^-) = 1$ contribute to the estimation of $\boldsymbol{\alpha}$ using model (2.25), i.e. t_{i1},...,t_{i,m_i^*} in Figure 2.1, and the score function for visit gap times becomes

$$\boldsymbol{U}_{i2}^*(\boldsymbol{\alpha}) = \sum_{j=1}^{m_i^*} \left\{ \frac{\partial ln\{\lambda[B(t_{ij})]\}}{\partial \boldsymbol{\alpha}} - \int_0^{B(t_{ij})} \frac{\partial \lambda(s)}{\partial \boldsymbol{\alpha}} ds \right\} - I(\bar{P}_i^{obs}(t_{im_i^*}) = 1) \int_0^{C_i - t_{im_i^*}} \frac{\partial \lambda(s)}{\partial \boldsymbol{\alpha}} ds. \tag{2.30}$$

This follows because if T_i is interval-censored, all visit gap times when $P_i^{obs}(t^-) = 1$, which contribute to the estimation of $\boldsymbol{\alpha}$, are exactly observed, while if T_i is right-censored at the last actual visit $t_{im_i^*}$, the last visit gap time is right-censored at C_i. Theorem 1 still holds with $\boldsymbol{U}_{i2}$ replaced by $\boldsymbol{U}_{i2}^*$. Note that since IIV weight $w_i(t)$ equals 1 when $P_i^{obs}(t^-) = 0$, (2.29) is modified to

$$\widehat{\Gamma}_{12} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i^*} \lambda^{-2}(B(t_{ij}); \widehat{\boldsymbol{\alpha}}) \left. \frac{\partial \lambda(B(t_{ij}); \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha'}} \right|_{\widehat{\boldsymbol{\alpha}}} \boldsymbol{c}(t_{ij} | \boldsymbol{A_i}; \widehat{\boldsymbol{\theta}}, t_{i0}) [P_i(t_{ij}) - \mu_T(t_{ij} | \boldsymbol{A_i}; \widehat{\boldsymbol{\theta}}), t_{i0}]. \tag{2.31}$$

2.3 Simulation Studies

This section aims to empirically examine the finite sample performances of the proposed IIV weighted estimators in the presence of dependent follow-up times and to compare with standard estimation approaches such as the unweighted GEE and MLE. To simplify, in this section, we assume that random drop-out is absent. We will discuss three cases: (i) subjects visit intermittently until the administrative end of follow-up; (ii) subjects visit intermittently but stop visiting after a known failure; and (iii) failure stops visits and the

failure time may be left-truncated at the first visit. We will also investigate how the IIV weighted estimators perform under misspecification of the visit process model.

2.3.1 Performances of the IIV Weighted Estimators and Standard Estimators

In this simulation study, we examine the IIV weighted estimation of a parametric proportional hazards model like (2.2) with a time-fixed treatment variable A_i of interest, and a time-varying variable $L_i(t)$ which affects the visit process as well as the failure time T_i as an intermediate variable. Another simulation study based on an AFT marginal outcome model with a time-fixed intermediate variable L_i will be introduced in Chapter 3. As we discussed in Section 1.5, a specific model form for outcomes will not necessarily keep the same structure or even have a neat closed form after some regressors have been marginalized over, e.g. proportional hazards models like (2.2). However, for some particular distributions of the outcomes and under certain assumptions about the ancillary variable process, proportional hazards can still hold, though the marginal regression parameters and the conditional regression parameters differ. The following empirical studies will be based on a model developed in Young and Tchetgen Tchetgen (2014).

Here, we consider a binary A_i, i.e. $A_i = 0$ or 1. In Section 2.2.4, we have introduced the theory and methods under a discrete visit process. To make the simulated data applicable for all the three cases, (i)-(iii), throughout this section and simulation studies in Chapter 3 and Chapter 4, the discretization of continuous time is based on a grid of 100 per time unit. We set the administrative censoring time for each subject as $\tau = 5$. Let $a_0 = 0$, $a_1 = 0.01$, ..., $a_M = \tau$ be the universal prespecified potential visit times for all individuals. For example, if we aim to study the effect of biologics on time to a joint dam-

age event during the first episode of biologics, since the median length of the treatment durations of the first episode of biologics in the PsA cohort study is about $3 \sim 3.5$ years, then Δa approximates 2 days. Additionally, since the median length of visit gap times is approximately 6 months, we assume there is never more than one visit within 2 days. Discretization in this way makes using the software for continuous survival data, e.g. R functions *survreg*, *phreg*, *coxph* and *coxreg* a reasonable approach for estimating the parameter $\boldsymbol{\alpha}$ in the visit process model, though data are simulated and "observed" in discrete time. In the following, a variable with subscript m denotes the value or level of it at time a_m, $m = 0, 1, ..., M$, where $M = 500$. We assume the whole history of $L_i(t)$ is known, i.e. $\bar{L}_i^{obs}(t) = \bar{L}_i(t)$, for simplicity, e.g. physical temperature or blood pressure that can be measured by patients, though it is more plausible in practice that the visit process is associated with $\bar{L}_i^{obs}(t)$. Subscript i may be dropped for simplicity as well. In each of the following simulation studies in this section, sample size is set as $n = 1000$ and the number of simulation replicates is $N = 500$.

CASE I: Event occurrence does not terminate visits

Firstly, we consider a case where event occurrence does not prohibit a patient from visiting the clinic. In the present simulation, we imitate a randomized trial with $Pr[A = 1] = 0.5$ and $Pr[A = 0] = 0.5$, and the ancillary variable L_m at a_m is assumed to be independent of its previous history and be normally distributed given A, i.e.

$$L_m | A, \bar{L}_{m-1} \sim N(\beta_1 A, \sigma_l), \tag{2.32}$$

where β_1 is the effect of treatment $A = 1$ versus $A = 0$ on the mean of L_m for any m and we let $\sigma_l = 1$. Here, we assume that the process L_m retains the same distribution before and after the occurrence of the outcome event. Now, to generate the outcome process, it

is assumed that the discrete time hazard of T at time a_{m+1} has the following form:

$$Pr[P_{m+1} = 0|\bar{L}_m, A, P_m = 1] = \Phi(\theta_0 + \theta_1 L_m + \theta_2 A), \tag{2.33}$$

where Φ denotes the CDF of a standard normal distribution $N(0,1)$ and $P_m = I(T > a_m)$. Once a zero has been generated for some P_m, we have $P_s = 0$ for all $s > m$, because survival status $P(t)$ is a decreasing function. Note that $P_0 = 1$ for all individuals. We set $\theta_0 = -2$, $\theta_1 = 1$, $\theta_2 = -0.1$ and $\beta_1 = -1$ so that the treatment A has a negative effect on the risk of failure and L_m behaves as an intermediate variable which is inhibited by treatment and is positively related to the risk of failure. Model (2.33) can be rewritten proportionally as

$$Pr[P_{m+1} = 0|\bar{L}_m, A, P_m = 1] = e^{\eta_0} \exp(\eta_1 L_m + \eta_2 A + \eta_3 A L_m),$$

where $e^{\eta_0} = \Phi(\theta_0)$ is the baseline hazard, and

$$\begin{aligned}
\eta_1 &= ln\left[\frac{\Phi(\theta_0 + \theta_1)}{\Phi(\theta_0)}\right], \\
\eta_2 &= ln\left[\frac{\Phi(\theta_0 + \theta_2)}{\Phi(\theta_0)}\right], \\
\eta_3 &= ln\left[\frac{\Phi(\theta_0)\Phi(\theta_0 + \theta_1 + \theta_2)}{\Phi(\theta_0 + \theta_1)\Phi(\theta_0 + \theta_2)}\right].
\end{aligned}$$

Then, the marginal hazard function of T given A alone can be obtained by marginalizing over L_m as

$$Pr[P_{m+1} = 0|A, P_m = 1] = \Phi\left\{c \cdot [\theta_0 + (\theta_2 + \theta_1\beta_1)A]\right\}, \tag{2.34}$$

where $c = 1/\sqrt{1 + \theta_1^2\sigma_l^2}$ and equation (2.34) can be rewritten in proportional hazards form as well, i.e.

$$Pr[P_{m+1} = 0|A, P_m = 1] = e^{\psi_0} \exp(\psi_1 A), \tag{2.35}$$

where $\psi_0 = ln\left\{\Phi(c \cdot \theta_0)\right\}$, $\psi_1 = ln\left\{\frac{\Phi[c\cdot(\theta_0+\theta_2+\theta_1\beta_1)]}{\Phi(c\cdot\theta_0)}\right\}$. Here ψ_0 and ψ_1 are the parameters we wish to estimate. Based on the values of θ_0, θ_1, θ_2, β_1 and σ_l given above, the baseline

hazard rate $e^{\psi_0} \doteq 0.08$ and the marginal hazard ratio $e^{\psi_1} \doteq 0.18$ for $A = 1$ versus $A = 0$, i.e. treatment results in about an 82% risk reduction.

The discrete visit process, $\{dN_m : m = 1, ..., M\}$, is generated by a Markov proportional intensities model with a constant baseline intensity:

$$\begin{aligned} Pr[dN_{m+1} = 1|A, L_m] &= \exp(\gamma_0 + \gamma_1 L_m + \gamma_2 A), \\ &= e^{\gamma_0} \exp(\gamma_1 L_m + \gamma_2 A) \end{aligned} \tag{2.36}$$

where $\gamma_0 = -4$, $\gamma_1 = 0$ or 1, and $\gamma_2 = -0.2$, so that the average visit gap time is approximately 0.59 in the untreated group and 0.63 in the treated group, when $\gamma_1 = 0$, and is about 0.33 in the untreated group and 1 in the treated group, when $\gamma_1 = 1$. Since we assume that the outcome event does not stop visits, the visit process generated by model (2.36) is non-informative unless $\gamma_1 \neq 0$. Thus, the scenario where $\gamma_1 = 0$ satisfies the independent observation assumption (B0), while the scenario with $\gamma_1 = 1$ gives dependent visit times. In fact, when $\gamma_1 = 1$ and $\theta_1 = 1$, L_m is a strong outcome-dependent risk factor of the visit process which fails to be adjusted for when standard analysis methods such as MLE and unweighted GEE are applied, and as a result, estimates could be biased. On the other hand, model (2.36) ensures that the conditionally independent observation assumption (B1) holds when both of A and L_m have been adjusted for via weighting, so that resulting estimators are consistent. The scenario of this simulation study is illustrated by a DAG shown in Figure 2.2, assuming $M = 2$ for simplicity.

Three methods, MLE, unweighted GEE and IIV weighted GEE, will be compared for the estimation of parameters ψ_0 and ψ_1 from model (2.35). The unweighted GEE estimator

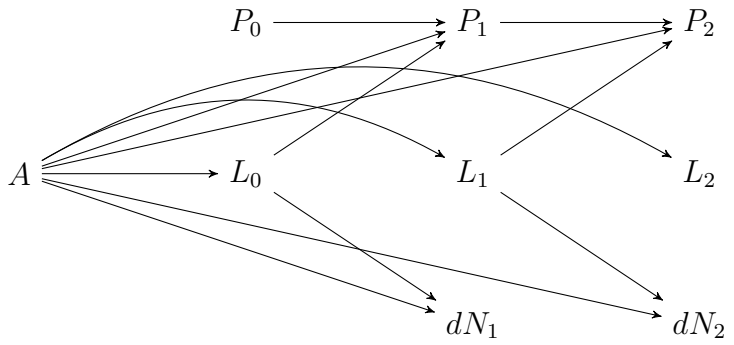


Figure 2.2: DAG for the simulation setting when individuals can still be followed up even after event occurrence to the administrative end τ, where $P_m = I(T > a_m)$, $P_0 = 1$ and dN_m indicates a clinic visit at a_m, where $m = 0, 1, 2$.

will be obtained by solving the following estimating function:

$$\boldsymbol{U}(\boldsymbol{\psi}) = \sum_{i=1}^{n} \sum_{m=1}^{M} \boldsymbol{c}(a_m|A_i; \boldsymbol{\psi}, t_{i0})[P_i(a_m) - \mu_T(a_m|A_i; \boldsymbol{\psi}, t_{i0})]dN_i^*(a_m), \tag{2.37}$$

where $\mu_T(a_m|A_i; \boldsymbol{\psi}, t_{i0}) = S_T(a_m|A_i; \boldsymbol{\psi})/S_T(t_{i0}|A_i; \boldsymbol{\psi})$ and t_{i0} is the start of follow-up and we assume $dN_i(a_m) = 0$ for all $a_m \leq t_{i0}$. Here, we have $t_{i0} = 0$, since we are not considering left-truncated failure times in the present case. In addition,

$$\boldsymbol{c}(a_m|A_i; \boldsymbol{\psi}, t_{i0}) = \frac{\partial \mu_T(a_m|A_i; \boldsymbol{\psi}, t_{i0})/\partial \boldsymbol{\psi}}{\mu_T(a_m|A_i; \boldsymbol{\psi}, t_{i0})[1 - \mu_T(a_m|A_i; \boldsymbol{\psi}, t_{i0})]},$$

i.e. the GEE given in (2.11) with an independent working covariance matrix specified. In addition, based on model (2.35), $S_T(a_m|A_i; \boldsymbol{\psi})$ is given by

$$S_T(a_m|A_i; \boldsymbol{\psi}) = [1 - e^{\psi_0} \exp(\psi_1 A_i)]^m.$$

The IIV weighted GEE estimator will be obtained by similarly solving the estimating

function (2.38) with the same $\boldsymbol{c}(\cdot)$ function defined above:

$$\boldsymbol{U}(\boldsymbol{\psi}, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \sum_{m=1}^{M} w_i(a_m; \boldsymbol{\alpha}) \boldsymbol{c}(a_m | A_i; \boldsymbol{\psi}, t_{i0}) [P_i(a_m) - \mu_T(a_m | A_i; \boldsymbol{\psi}, t_{i0})] dN_i^*(a_m). \quad (2.38)$$

The IIV weights $w_i(a_m; \boldsymbol{\alpha})$ can be estimated by fitting the following semi-Markov piecewise-constant baseline proportional hazards model for visit gap times:

$$\lambda_N(a_{m+1}; \boldsymbol{\rho}, \boldsymbol{\alpha}) = \sum_{k=1}^{K} \rho_k I_k[B(a_{m+1})] \exp(\alpha_1 L_m + \alpha_2 A), \quad m = 0, 1, ..., M-1, \quad (2.39)$$

where $I_k(t) = I\{t \in (d_{k-1}, d_k]\}$, $0 = d_0 < d_1 <, ..., < d_K = +\infty$ are the cut-points of visit gap times and $\boldsymbol{\rho} = (\rho_1, ..., \rho_K)'$ are positive unknown constants and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$. We do this in order to reflect the desirability of using flexible models for the visit time process. The actual visit time process model (2.36) in the simulation is of this form but with $K = 1$. Here, we chose $K = 4$ and $d_1 = 0.40$, $d_2 = 0.75$ and $d_3 = 1.00$, roughly based on the quartiles of visit gap times. In general, K and cut-points can be chosen by a graphical comparison of the resulting estimate of $\boldsymbol{\rho}$ based on model (2.39) with a non-parametric estimate of the baseline intensity function based on the Cox model. Here, $\boldsymbol{\rho} = (\rho_1, ..., \rho_4)'$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$ are estimated by the R function *phreg*. Because the common risk factors of failure risk and visit intensity at a_{m+1} have been adjusted for by model (2.39), given the condition (B1) with $\boldsymbol{Z_i}(a_m) = (A_i, L_m)'$ is true, the IIV weighted GEE estimators should be consistent.

The MLE for $\boldsymbol{\psi}$ can be obtained by solving the score function based on the likelihood given in (2.1). All the estimating equations for $\boldsymbol{\psi}$, GEEs or score functions, are numerically solved by using R function *nleqslv* with the Newton-Raphson method specified, setting $\boldsymbol{\psi} = (-1, -1)'$ as the initial value. The resulting estimates by IIV weighted GEE, unweighted GEE and MLE are summarized in Table 2.1 for the case of independent follow-up times ($\gamma_1 = 0$) and the case of dependent follow-up times ($\gamma_1 = 1$). Bias and mean squared error

(MSE) are respectively estimated by

$$\widehat{BIAS} = \frac{1}{N}\sum_{r=1}^{N}(\widehat{\psi}_{lr} - \psi_{lr}), \quad l = 0, 1,$$

and

$$\widehat{MSE} = \frac{1}{N}\sum_{r=1}^{N}(\widehat{\psi}_{lr} - \psi_{lr})^2, \quad l = 0, 1.$$

where $\widehat{\psi}_{lr}$ is the rth estimate of ψ_l and $N = 500$ is the number of simulations. The coverage probability (CP) for nominal 95% confidence intervals of an estimator of ψ_l, $l = 0, 1$, is estimated by:

$$\widehat{CP} = \frac{1}{N}\sum_{r=1}^{N} I\{\widehat{\psi}_{lr} - 1.96\ se(\widehat{\psi}_{lr}) < \psi_l < \widehat{\psi}_{lr} + 1.96\ se(\widehat{\psi}_{lr})\}, \quad l = 0, 1,$$

where *se* denotes the asymptotic standard error of an estimate, and ASE in Tables 2.1-2.4 denotes the average of asymptotic standard errors. The asymptotic distribution (2.27) and asymptotic standard error of the proposed IIV weighted GEE estimator of $\boldsymbol{\psi}$ is given in Theorem 1 in Section 2.2.5. The asymptotic standard error of an unweighted GEE estimator can be estimated by the ordinary sandwich form variance formula given in White (1982) and Liang and Zeger (1986), i.e. $V_\psi = \Gamma^{-1}\Sigma\ (\Gamma^{-1})'$, where $\Gamma = -E(\partial \boldsymbol{U}_i(\boldsymbol{\psi})/\partial \boldsymbol{\psi}')$ and $\Sigma = E\{\boldsymbol{U}_i(\boldsymbol{\psi})\boldsymbol{U}_i(\boldsymbol{\psi})'\}$, where $\boldsymbol{U}_i(\boldsymbol{\psi})$ is the ith subject's contribution for the estimating function given in (2.37). The MLE's asymptotic variance V_ψ is the inverse of its Fisher information matrix. It can be estimated by the negative Jacobian matrix of the score function divided by the sample size. When we use *nleqslv* to solve the score function, the Jacobian matrices at the resulting MLEs are outputted. In addition, the empirical standard error (ESE) is the sample standard deviation of estimates across the 500 simulated data sets.

Table 2.1: Bias, average of asymptotic standard errors (ASE), empirical standard error (ESE), mean squared error (MSE) and coverage probability (CP) for $(\psi_0, \psi_1)'$ in model (2.35) when individuals still can be followed up even after failure occurrence. In (2.36), $\gamma_1 = 0$, which means that visit times are independent, i.e. assumption (B0) is satisfied; $\gamma_1 = 1$, which means that visit times are dependent, but we assume that (B1) is satisfied. Sample size: $n = 1000$, and simulation replicates: $N = 500$.

			TRUE VALUE	BIAS	ASE	ESE	MSE	CP
$\gamma_1 = 0$	MLE	ψ_0	-2.543	0.006	0.069	0.071	0.005	0.94
		ψ_1	-1.713	-0.002	0.085	0.085	0.007	0.95
	Unweighted	ψ_0	-2.543	0.010	0.072	0.075	0.006	0.93
		ψ_1	-1.713	-0.004	0.090	0.089	0.008	0.95
	IIV Weighted	ψ_0	-2.543	0.010	0.072	0.076	0.006	0.93
		ψ_1	-1.713	-0.004	0.090	0.089	0.008	0.95
$\gamma_1 = 1$	MLE	ψ_0	-2.543	0.111	0.060	0.062	0.016	0.53
		ψ_1	-1.713	-0.072	0.081	0.081	0.012	0.87
	Unweighted	ψ_0	-2.543	0.119	0.068	0.066	0.019	0.59
		ψ_1	-1.713	-0.075	0.091	0.088	0.013	0.87
	IIV Weighted	ψ_0	-2.543	0.004	0.085	0.090	0.008	0.93
		ψ_1	-1.713	-0.002	0.113	0.115	0.013	0.93

From Table 2.1, when $\gamma_1 = 0$, we see that when visit times are outcome-independent, bias of all the three methods, MLE, unweighted GEE and IIV weighted GEE, is negligible relative to the sampling standard error. The ASEs agree well with the corresponding ESEs, which indicates that the asymptotic distribution and sandwich variance estimator given in Theorem 1 in Section 2.2.5 provides satisfactory estimation of the asymptotic variances of the proposed IIV weighted estimators for sample size $n = 1000$. Additionally, it is seen that weighted GEE and unweighted GEE estimates have approximately the same ASEs, and MLEs give slightly smaller ASEs, which suggests that the MLEs are a little more efficient. In addition, all the three methods give coverage probabilities around $93\% - 95\%$ which are close to the nominal level, 95%, so the overall performances of these three methods with an independent observation scheme are satisfactory. We found that when sample sizes are increased, the asymptotic variances can be more accurately estimated so that coverage probabilities can be improved. Some simulation results for a smaller sample size n, e.g. $n = 500$, can be found in Table 2.4 and Table 2.5, where we can see that coverage probabilities become relatively lower.

The lower part of Table 2.1 summarizes the simulation results for the settings with outcome-dependent visit times ($\gamma_1 = 1$). We see that the bias of the IIV weighted estimator is still negligible and the coverage is close to the nominal level. However, MLE and unweighted GEE estimator show large bias (over 80% of the ESE) and poor coverage probabilities. The estimation of the asymptotic variances of IIV weighted estimators by the sandwich form estimator given in Theorem 1 is satisfactory.

To summarize this simulation study for the case where failure does not prohibit visits, we see when the intermittent visit times are non-informative for the inference of outcomes, i.e. condition (B0) is true, standard estimation methods result in sound inference. However, when the visit times are not independent, but adjustable by known factors, i.e.

condition (B1) is true, standard methods fail. The proposed IIV weighting method results in consistent estimates and comparable efficiency to the MLE when visits are independent, and also leads to consistent estimates even when visits are dependent.

CASE II: Event occurrence terminates visits

Now, suppose subjects who have experienced the event of interest will not come to visit the clinic any longer, i.e. follow-up discontinues once a failure is known. In this case, we assume that the visit process follows a proportional intensity model conditional on $\boldsymbol{Z_m} = (A, L_m, \bar{P}_m^{obs})'$ under assumption (B1*), i.e.

$$\begin{aligned} & Pr[dN_{m+1} = 1 | A, L_m, \bar{P}_m^{obs}] \\ &= \begin{cases} \exp(\gamma_0 + \gamma_1 L_m + \gamma_2 A), & \text{if } \bar{P}_m^{obs} = 1 \\ 1, & \text{if } \bar{P}_m^{obs} = 0 \end{cases}, \quad m = 0, 1, ..., M-1, \end{aligned} \tag{2.40}$$

where subscripts for subjects are suppressed for simplicity, and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)'$ is specified the same as in CASE I. Data generation for A, L_m, P_m and relevant parameter values are also the same as in CASE I. The DAG of this scenario is displayed in Figure 2.3 where we see that current visit intensity is associated with the previous outcome and previous visit status. Therefore, the observed past outcome should be adjusted for to achieve consistent estimation by using estimating function estimation methods, though L_m might not be a common risk factor between outcomes and visit times. That is, even if γ_1 from the visit process model (2.40) equals zero the unweighted GEE without considering the difference of visit intensities before and after failure occurrence leads to biased estimates, which we can see from Table 2.2. On the other hand, comparing Table 2.1 with Table 2.2 and Table 2.3, the MLE based on likelihood (2.1) gives the same estimates, because only the two actual visits $t_{ij} < T_i \leq t_{i,j+1}$ are needed in the likelihood (2.1). To obtain the IIV weighted GEE

estimates, the IIV weights needed in (2.23) are estimated by $\widehat{w}_i(a_{m+1}) = 1/\lambda_N(a_{m+1}; \widehat{\boldsymbol{\rho}}, \widehat{\boldsymbol{\alpha}})$ by fitting the semi-Markov (discrete) visit gap times model given in (2.39) when $\bar{P}_m^{obs} = 1$, and $\widehat{w}_i(a_{m+1}) = 1$ when $\bar{P}_m^{obs} = 0$, and function $\boldsymbol{c}(\cdot)$ is the same as for CASE I. The unweighted GEE estimates are obtained similarly by letting $\widehat{w}_i(a_{m+1}) = 1$ for all $m = 0, ..., M-1$, and MLEs are computed by maximizing likelihood (2.1). The simulation results for $\gamma_1 = 0$ and $\gamma_1 = 1$ are summarized in Table 2.2 and Table 2.3, respectively.

From Table 2.2, when $\gamma_1 = 0$ in (2.40), we see that the bias of MLEs and IIV weighted GEE estimator is negligible and their coverage probabilities are overall satisfactory. However, unweighted GEE estimates are biased, as we expected. For the unweighted GEE method, we pretend to observe the outcome after failure occurrence, i.e. the visit process changes distribution after failure occurrence, but the past observed outcome history fails to be adjusted for, so the resulting estimates are biased. In addition, although both MLE and the IIV weighted GEE estimator are consistent, MLE shows less variability than the weighted GEE estimator, i.e. MLE is more efficient.

When $\gamma_1 = 1$ in (2.40), it is seen from Table 2.3 that the bias of IIV weighted GEE estimator is still negligible, the corresponding ASEs and ESEs are close to each other, and coverage probabilities are around the nominal 95% level. On the other hand, likelihood (2.1) leads to biased estimates and unsatisfactory coverage probabilities for MLE, because the information (i.e. L_m) carried by visit times fails to be taken into account. The unweighted GEE method results in heavily biased estimates and nearly zero coverage probabilities, because more information fails to be adjusted for.

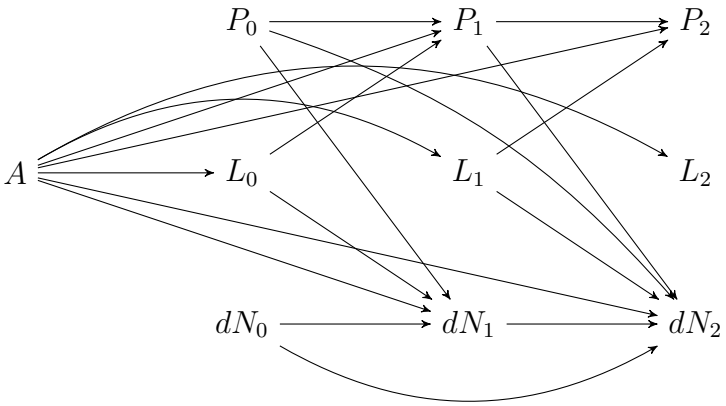


Figure 2.3: DAG for the simulation setting when individuals are assumed to stop visiting after a known failure, where $P_m = I(T > a_m)$ and let $P_0 = 1$ and dN_m indicates a clinic visit at a_m, where $m = 0, 1, 2$.

Table 2.2: Bias, average of asymptotic standard errors (ASE), empirical standard error (ESE), mean squared error (MSE) and coverage probability (CP) for $(\psi_0, \psi_1)'$ in model (2.35) when individuals stop visiting after a known failure, with $\gamma_1 = 0$ in (2.40). Sample size: $n = 1000$, and replicates: $N = 500$.

		TRUE VALUE	BIAS	ASE	ESE	MSE	CP
MLE	ψ_0	-2.543	0.006	0.069	0.071	0.005	0.94
	ψ_1	-1.713	-0.002	0.085	0.085	0.007	0.95
Unweighted GEE	ψ_0	-2.543	0.748	0.054	0.057	0.562	0.00
	ψ_1	-1.713	0.434	0.069	0.070	0.193	0.00
IIV Weighted	ψ_0	-2.543	0.006	0.072	0.076	0.006	0.92
	ψ_1	-1.713	-0.001	0.089	0.090	0.008	0.95

Table 2.3: Bias, average of asymptotic standard errors (ASE), empirical standard error (ESE), mean squared error (MSE) and coverage probability (CP) for $(\psi_0, \psi_1)'$ in model (2.35) when individuals stop visiting after a known failure, with $\gamma_1 = 1$ in (2.40). Sample size: $n = 1000$, and replicates: $N = 500$.

		TRUE VALUE	BIAS	ASE	ESE	MSE	CP
MLE	ψ_0	-2.543	0.111	0.060	0.062	0.016	0.53
	ψ_1	-1.713	-0.072	0.081	0.081	0.012	0.87
Unweighted GEE	ψ_0	-2.543	0.844	0.051	0.054	0.715	0.00
	ψ_1	-1.713	0.425	0.071	0.069	0.185	0.00
IIV Weighted	ψ_0	-2.543	0.009	0.081	0.086	0.008	0.95
	ψ_1	-1.713	0.007	0.109	0.114	0.013	0.95

CASE III: Event occurrence terminates visits and failure times could be left-truncated

In the PsA example, only patients who have PsA are enrolled in the cohort, so onset times of PsA occurred before clinic enrolment times. If we are interested in the time to the appearance of the first joint damage since onset ($t = 0$), it may be left-truncated at clinic entry time t_{i0} (> 0). To model left-truncation in the setting discussed above, we simulate a delayed entry time t_{i0} as the maximum of 0 and a random number generated from $N(0.15, 0.1)$. As a result, $90\% - 95\%$ of t_{i0}'s are greater than zero and the delayed entry times are completely random. About 40% of subjects whose failures occurred before t_{i0} are excluded from the analysis. The outcome process, treatment A and intermediate variable L_m are generated the same as for CASE II. Visits are generated based on model (2.40) onward from t_0.

The IIV weighted GEE and unweighted GEE estimates are obtained the same as in CASE II but note that $t_{i0} > 0$ for most individuals. MLEs are found by maximizing the likelihood (1.5). All these estimates are summarized in Table 2.4 for the scenario where $\gamma_1 = 1$ in (2.40). Table 2.4 shows that the IIV weighted GEE performs satisfactorily with negligible bias and good coverage probabilities. Comparing with Table 2.3, ASEs and ESEs are higher, because sample size is smaller and fewer visits contribute to the analysis when failure times could be left-truncated. MLEs are still biased, because L_m is not adjusted for if only A is conditioned on in the likelihood. Finally, the unweighted GEE method still performs very poorly.

Table 2.4: Bias, average of asymptotic standard errors (ASE), empirical standard error (ESE), mean squared error (MSE) and coverage probability (CP) for $(\psi_0, \psi_1)'$ in model (2.35) when individuals stop visiting after a known failure and failure time could be left-truncated at t_{i0}, with $\gamma_1 = 1$ in (2.40). Initial sample size: $n = 1000$, analysis sample size: $n^* \doteq 600$ (i.e. $n^* = \sum_{i=1}^{n} I\{T_i > t_{i0}\}$), and simulation replicates: $N = 500$.

		TRUE VALUE	BIAS	ASE	ESE	MSE	CP
MLE	ψ_0	-2.543	0.124	0.098	0.107	0.027	0.77
	ψ_1	-1.713	-0.079	0.116	0.124	0.021	0.89
Unweighted GEE	ψ_0	-2.543	0.855	0.083	0.085	0.738	0.00
	ψ_1	-1.713	0.414	0.099	0.101	0.182	0.02
IIV Weighted	ψ_0	-2.543	0.020	0.128	0.134	0.018	0.93
	ψ_1	-1.713	0.001	0.150	0.159	0.025	0.94

2.3.2 Investigation of the IIV Weighted Estimator under Model Misspecification for the Visit Process

In Section 2.2.1, we mentioned that in addition to the conditionally independent observation scheme condition (B1) or (B1*), the outcome model as well as the visit model need to be correctly specified when the IIV weighting approach is applied. When a marginal outcome model is the inference target, we should be careful with the model collapsibility as discussed in Section 1.5. Moreover, if the model for the visit process is misspecified or certain assumptions which the visit time model relies on are violated, the finally resulting IIV weighted estimates could be biased. Therefore, this section aims to study the impact on the IIV weighted estimator when the visit time model fails to be correctly specified. Except for the model for generating intermittent visits, data are simulated from the same scenario as CASE II in Section 2.3.1, where a known failure occurrence is assumed to terminate visits and failure times are not left-truncated.

In CASE II of Section 2.3.1, we generated the visits from a constant baseline proportional intensities model (2.40) when $\bar{P}_m^{obs} = 1$, so both the weight formulated as (2.17) which was considered by Lin et al. (2004) and our proposed weight (2.20) can be applied to estimate the IIV weights when $\bar{P}_m^{obs} = 1$. However, here we generate the visit process from a semi-Markov model for (discrete) visit gap times given by

$$\begin{aligned} &Pr[dN_{m+1} = 1 | A, L_m, \bar{P}_m^{obs}] \\ &= \begin{cases} \{[\lambda_0 B(a_{m+1})]^\kappa - [\lambda_0 B(a_m)]^\kappa\} \exp\{\gamma_1 L_m + \gamma_2 A\}, & \text{when } \bar{P}_m^{obs} = 1, \\ 1, & \text{when } \bar{P}_m^{obs} = 0, \end{cases} \end{aligned} \tag{2.41}$$

where $1/\lambda_0$ is the scale parameter and κ is the shape parameter of a Weibull distribution. We set $\lambda_0 = 1.8$ and consider two values of shape parameter κ in model (2.41): $\kappa = 1.0$ and $\kappa = 1.5$ and note that when $\kappa = 1.0$, model (2.41) reduces to model (2.40), since we know

$a_{m+1} - a_m = 0.01$. Let $\gamma_1 = 1$ in (2.41), so L_m should be adjusted for, and let $\gamma_2 = -0.2$ as in CASE II. When $\kappa = 1.0$, the overall average length of visit gap times is 0.75: 0.34 for the untreated ($A = 0$) group and 1.06 for the treated ($A = 1$) group, and when $\kappa = 1.5$, the overall average length of visit gaps is 0.62: 0.35 for the untreated group and 0.80 for the treated group.

Given the assumption (B1*) is true with $\boldsymbol{Z}(a_m) = (A, L_m, \bar{P}_m^{obs})'$, we consider the following two working visit models for estimating the IIV weights when $\bar{P}_m^{obs} = 1$:

$$\lambda_N(a_{m+1}) = \lambda_{N0}(a_{m+1}; \boldsymbol{\rho}) \exp(\alpha_1 L_m + \alpha_2 A), \tag{2.42}$$

and

$$\lambda_N(a_{m+1}) = \lambda_{N0}^{\dagger}(B(a_{m+1}); \boldsymbol{\rho}) \exp(\alpha_1 L_m + \alpha_2 A), \tag{2.43}$$

where baselines $\lambda_{N0}(a_{m+1}; \boldsymbol{\rho})$ and $\lambda_{N0}^{\dagger}(B(a_{m+1}); \boldsymbol{\rho})$ are both piecewise constant. Although Lin et al. (2004) proposed to adopt the Breslow estimator to estimate a non-parametric baseline $\lambda_{N0}(t)$ in (2.42) and then apply kernel smoothing to attain a certain convergence rate, we consider a parametric baseline so that it can be fairly compared with our weight formula (2.43) based on a piecewise constant baseline for visit gap times. Cut-points for the piecewise constant baseline hazards are chosen to be equally spaced, e.g. $(0.30, 0.60, 0.90, 1.20, 1.50)$ for model (2.42) and $(0.3, 0.6, 0.9)$ for model (2.43).

In Table 2.5, we see that when either $\kappa = 1.0$ or $\kappa = 1.5$, estimate $\tilde{\boldsymbol{\psi}}$ based on weight model (2.42) and $\widehat{\boldsymbol{\psi}}$ based on weight model (2.43) have comparable mean squared errors. We assume condition (B1*) is satisfied and both weight model (2.42) and (2.43) have taken $\boldsymbol{Z}(a_m)$ into account, so they are supposed to give asymptotically unbiased estimates of $\boldsymbol{\psi}$, provided that the weight model is correctly specified. When $\kappa = 1.0$ in (2.41), i.e. visit times are generated from a constant baseline intensity model, (2.42) and (2.43) are actually equivalent. As a result, from Table 2.5, we can see that these two models result

Table 2.5: Investigation of the IIV weighted estimator under different model specifications for the estimation of IIV weights, where $\tilde{\boldsymbol{\psi}}$ denotes the estimator based on weight model (2.42) and $\widehat{\boldsymbol{\psi}}$ denotes the estimator based on weight model (2.43). Line 5 and line 6 represent the case of weight model misspecification. BSE denotes the mean of bootstrap estimated standard errors, and ECP denotes 95% empirical coverage probability; ESE denotes empirical standard error; MSE denotes mean squared error. Sample size: n=500, the number of replicates: 500, and bootstrap sample size: 100.

		TRUE VALUE	BIAS	BSE	ESE	MSE	ECP
$\kappa = 1.0$	$\tilde{\psi}_0$	-2.543	0.015	0.112	0.130	0.017	0.90
	$\tilde{\psi}_1$	-1.713	0.017	0.150	0.163	0.027	0.92
	$\widehat{\psi}_0$	-2.543	0.014	0.112	0.129	0.017	0.91
	$\widehat{\psi}_1$	-1.713	0.010	0.151	0.163	0.027	0.92
$\kappa = 1.5$	$\tilde{\psi}_0$	-2.543	0.034	0.121	0.127	0.017	0.92
	$\tilde{\psi}_1$	-1.713	0.067	0.149	0.155	0.028	0.91
	$\widehat{\psi}_0$	-2.543	0.010	0.126	0.134	0.018	0.93
	$\widehat{\psi}_1$	-1.713	0.009	0.158	0.166	0.028	0.92

in estimates with similar bias and standard errors. However, when $\kappa = 1.5$, $\widehat{\boldsymbol{\psi}}$ obtained by using the correct visit process model leads to smaller bias than $\tilde{\boldsymbol{\psi}}$ obtained by a wrong visit process model. For example, the bias for $\tilde{\psi}_1$ is about 0.43 times of the ESE, while the bias for $\widehat{\psi}_1$ is about 0.05 times of the ESE. To conclude, given the essential condition (B1*) is true, from this study, we see that models should be assessed carefully based on the characteristics of the real problem and data. Otherwise, biased estimates could result. Later, in Chapter 6, we will introduce a doubly robust estimator for the IIV weighting approach, which was also considered by Pullenayegum and Feldman (2013). In addition, overall low coverage probabilities should be caused by the small sample size, but when the weight model is correctly specified, coverage is a bit closer to the nominal level.

Chapter 3

CAUSAL INFERENCE FOR TREATMENT EFFECTS IN OBSERVATIONAL STUDIES WITH DEPENDENT INSPECTION TIMES

In the previous chapter, our objective was to validly assess a marginal association when an observational longitudinal cohort had features such as interval censoring, left truncation and outcome-dependent intermittent inspection times. We did not discuss the estimation of a (marginal) causal effect when these problems are present. In this chapter, we propose a double weighting method which leads to consistent estimation of the causal effect of a treatment variable on an interval-censored event time outcome under a outcome-dependent

observation scheme. This method can be extended to the case with a time-varying treatment which is assigned at a set of regular discrete times, but we will focus on a time-fixed treatment in the context of the PsA example. The doubly weighted estimator can eliminate the selection bias due to intermittent outcome-dependent follow-up times as well as potential confounding of the treatment effect. Similar methodology can be applied to the non- and semi-parametric estimation which will be introduced in Chapter 4, but parametric estimation is the focus of this chapter. We continue the discussion based on a parametric proportional hazards model (2.2) as an example. Then empirical studies will be employed to examine the finite sample performances of proposed estimators, with simulations of both a randomized experiment and an observational study. The proposed method will be applied to the PsA cohort in Chapter 5 where treatments such as biologics are confounded with disease risk factors and the event time outcome is dependently interval-censored due to intermittent visits.

3.1 A Doubly Weighted Estimator for Causal Inference with Intermittent Outcome-Dependent Inspection Times

From the previous chapter, we know that IIV weights eliminate the selection bias caused by outcome-dependent intermittent observation. Now, we will discuss how to deal with the selection bias due to confounders in observational studies. See Section 1.3.1 for the definition of causal effect and confounding. An example we mentioned earlier is to study the effect of biologics on the time to some joint damage event, e.g. time to an increase in the number of damaged joints since treatment started. In the PsA cohort study, the median

durations of the first episodes of biologics is over 3 years, so treatment with biologics can be considered as a time-fixed treatment, when the treatment effect of biologics on joint damage event times is of interest. It is known that the prescription of biologics could be determined by disease status, biomarkers, demographic information, family history, or other concurrent treatments such as NSAIDs and DMARDs, which were assessed at or prior to biologics' assignment. On the other hand, those variables may be risk factors of joint damage as well. The common causes between treatment assignment and outcomes are confounders and result in selection bias in the analysis of outcomes. For example, patients who have more joint activity are more likely to receive biologics and are also at higher risk of joint damage. A crude estimate of the association between biologics and the joint damage event by directly comparing the two treatment groups might be misleading, which we will discuss in Section 5.1.2 and Section 5.2. In this chapter, we consider a longitudinal observational study with confounding variables as well as dependent inspection times which are adjusted for by double weighting so that causal marginal treatment effect(s) can be validly evaluated and plausibly interpreted.

3.1.1 Structural Models versus Associational Models

Recall that model (2.2) which was introduced in Section 2.1.1 is a parametric proportional hazards model which allows the estimation of the marginal association of a variable of primary interest, e.g. A, and a failure time outcome, T, i.e.

$$\lambda_T(t|A=a;\boldsymbol{\theta}) = \lambda_{T,0}(t;\boldsymbol{\gamma})\,\exp(\beta a), \tag{3.1}$$

where we let A be a dichotomous treatment random variable, e.g. 1 if treated, or 0 otherwise; a denotes the realization value of A; $\lambda_T(t|A=a;\boldsymbol{\theta})$ is the hazard function of T conditional on the actual treatment $A=a$; $\lambda_{T,0}(t;\boldsymbol{\gamma})$ is a parametric baseline hazard

function, and $\boldsymbol{\theta} = (\boldsymbol{\gamma}', \beta)'$. An estimate of β from the above model represents association between A and T, but can not be interpreted as a causal effect of A on T unless treatment A is randomized, so models like (3.1) are referred to as associational models. Alternatively, Robins et al. (2000) introduced a class of so-called marginal structural models (MSMs) relating the hypothetical exposure or treatment, a, to the corresponding counterfactual outcome, T^a and allowing the unconditional or marginal effect of a to be estimated. The MSM corresponding to model (3.1) is defined by

$$\lambda_{T^a}(t; \boldsymbol{\theta}^*) = \lambda_{T^0}(t; \boldsymbol{\gamma}^*) \exp(\beta^* a), \tag{3.2}$$

where λ_{T^a} is the hazard function of the counterfactual outcome T^a under hypothetical treatment a. Here, $\boldsymbol{\theta}^* = (\boldsymbol{\gamma}^{*\prime}, \beta^*)'$ and λ_{T^0}, the parameter and baseline hazard in the counterfactual outcome model (3.2), are the targets of inference. Specifically, T^0 and T^1 represent the counterfactual outcomes under being untreated and treated, respectively. As we introduced in Section 1.3.3, usually only one of the counterfactual outcomes can be observed for each subject, but the average causal effect (ACE), e.g. the causal hazard ratio in MSM (3.2), can be evaluated for the population. In this MSM, e^{β^*} is interpreted as the casual hazard ratio of being treated ($a = 1$) versus being not untreated ($a = 0$), which is usually different from the crude hazard ratio, e^{β}, of $A = 1$ versus $A = 0$ from model (3.1), when treatment A is non-randomized.

In a randomized experiment, the association parameter β and the causal effect β^* are equal, since randomization ensures the absence of measured or unmeasured confounders. But in an observational study, they are usually different, unless it is evident that the treatment is not confounded. Greenland et al. (1999) defined the difference between the population-averaged causal effect, β^*, and the raw marginal association between treatment and outcome, β, as the amount of confounding bias. Standard methods to estimate a causal

effect, e.g. β^*, and eliminate the confounding bias are standardization, inverse probability weighting, and propensity score adjustment approaches such as matching, stratification and covariate adjustment. Among them, inverse probability weighting is commonly used in practical applications. Explicitly, in an observational study, under the assumption of no unmeasured confounders, i.e. (A1) in Section 1.3.3, parameters in MSMs can be consistently estimated by modifying the crude estimates obtained from an associational model, e.g. (3.1), by weighting each subject with the inverse probability of receiving that treatment. In this sense, treatment can be considered as unconfounded in the pseudo-population. As a result, weighted estimates based on the associational model have causal interpretations. In the following, we will introduce an inverse-probability-of-treatment (IPT) weight first and then combine that with the IIV weight we proposed in Section 2.2.2 for parametric survival models.

3.1.2 A Doubly Weighted Estimator

Section 1.3.5 provides a preliminary introduction to the inverse-probability-of-treatment (IPT) weighting approach, which is also referred to as propensity score weighting method, because the IPT weights are formulated as the inverse of propensity scores. This section will further introduce the IPT weighting method and combine that with the IIV weighting to remove various sources of selection bias. As we discussed earlier in Section 1.3.3, in addition to the consistency assumption and stable unit-treatment value assumption (SUTVA), an important assumption we need to assume is the conditional exchangeability of treatment, i.e. no unmeasured confounders, which can be defined for our context as

(C1) Strongly Ignorable Treatment Assignment/No Unmeasured Confounders

$$P^a(t) \amalg A|\boldsymbol{V}, \quad \forall t > 0 \text{ and } \forall a,$$

where $P^a(t) = I(T^a > t)$, $a = 0$ or 1 if treatment is a dichotomous variable, and its mean is $E[P^a(t)] = Pr(T^a > t) \triangleq S_{T^a}(t)$. Also, we assume that for any a, the following positivity assumption is satisfied

$$0 < Pr(A = a|\boldsymbol{V}) < 1, \quad \text{for all } \boldsymbol{V}.$$

Assuming that the vector $\boldsymbol{V_i}$ includes all the measured confounders between A_i and T_i and there are no unmeasured confounders, we can construct the following weight:

$$\begin{aligned} w_i^* &= \frac{1}{Pr(A_i = a|\boldsymbol{V_i}; \boldsymbol{\zeta})} \\ &\triangleq \frac{1}{f(a|\boldsymbol{V_i}; \boldsymbol{\zeta})}, \end{aligned} \tag{3.3}$$

where $f(a|\boldsymbol{V_i}, \boldsymbol{\zeta})$ denotes a parametric mass function or a density function of random treatment variable A_i at value or level a, conditional on confounders $\boldsymbol{V_i}$. The measured confounders $\boldsymbol{V_i}$ can be anything which predicts the assignment of treatment, affects the outcome, and was measured before the assignment of treatment. Some good candidates for the confounders of treatment with biologics could be age, gender, joint activity, PsA duration, results of lab tests, family history of disease, and other treatments used concurrently. Further detailed theory about the propensity scores and IPT weighting methods can be found in Rosenbaum and Rubin (1983) and Hernán and Robins (2006, 2016).

If A_i is binary, the parameter $\boldsymbol{\zeta}$ in weight formula (3.3) can be estimated by fitting a logistic regression:

$$ln\left[\frac{Pr(A_i = 1|\boldsymbol{V_i}; \boldsymbol{\zeta})}{1 - Pr(A_i = 1|\boldsymbol{V_i}; \boldsymbol{\zeta})}\right] = \boldsymbol{\zeta}'\boldsymbol{V_i}, \tag{3.4}$$

where $\boldsymbol{V_i} = (1, V_{i1}, ..., V_{iq})'$. Then, $\boldsymbol{\zeta}$ can be estimated by solving the quasi-likelihood score equation $\sum_{i=1}^{n} \boldsymbol{U_{i3}}(\boldsymbol{\zeta}) = \boldsymbol{0}$ based on the logistic regression model (3.4):

$$\boldsymbol{U_{i3}}(\boldsymbol{\zeta}) = \frac{\partial p(\boldsymbol{V_i}; \boldsymbol{\zeta})/\partial \boldsymbol{\zeta}}{p(\boldsymbol{V_i}; \boldsymbol{\zeta})[1 - p(\boldsymbol{V_i}; \boldsymbol{\zeta})]}[A_i - p(\boldsymbol{V_i}; \boldsymbol{\zeta})], \tag{3.5}$$

where $p(\boldsymbol{V_i};\boldsymbol{\zeta}) = Pr[A_i = 1|\boldsymbol{V_i};\boldsymbol{\zeta}]$. This can also be done by using software for generalized linear models such as R function *glm*. Next, the IPT weight w_i^* can be estimated by

$$w_i^*(\widehat{\boldsymbol{\zeta}}) = Pr[A_i|\boldsymbol{V_i};\widehat{\boldsymbol{\zeta}}]^{-1} = \{\exp(A_i\widehat{\boldsymbol{\zeta}}'\boldsymbol{V_i})/[1+\exp(\widehat{\boldsymbol{\zeta}}'\boldsymbol{V_i})]\}^{-1}. \tag{3.6}$$

To consistently estimate the causal effect, e.g. β^* in model (3.2), in an observational study where treatment is confounded and failure times may be dependently interval-censored, a double weight given below can be applied:

$$w_i^\dagger(t) = w_i^*\ w_i(t), \tag{3.7}$$

where $w_i(t)$ is the IIV weight defined by (2.14) for adjusting for the outcome-dependent follow-up times, and w_i^* is the IPT weight defined by (3.3) for adjusting for confounding. For example, we can solve the estimating function (2.3) incorporated with the double weight $w_i^\dagger(t)$ for the case where failure occurrence does not terminate visits (CASE I) and visit times are continuous, i.e.

$$\boldsymbol{U}^{ww}(\boldsymbol{\theta^*},\boldsymbol{\alpha},\boldsymbol{\zeta}) = \sum_{i=1}^{n}\sum_{all\ a}\int_0^{\tau_i} w_i^*(\boldsymbol{\zeta})w_i(t;\boldsymbol{\alpha})I(A_i=a)\boldsymbol{c}(t|a,\boldsymbol{\theta^*})[P_i(t)-S_{T^a}(t;\boldsymbol{\theta^*})]dN_i^*(t), \tag{3.8}$$

where $S_{T^a}(t;\boldsymbol{\theta^*})$ is a parametric survival function of the counterfactual outcome T^a, with assuming that a is a discrete treatment; $I(\cdot)$ is an indicator function; $dN_i^*(t) = dN_i(t)C_i(t)$ where $C_i(t)$ is the at risk indicator, i.e. $C_i(t) = I(C_i > t)$, and C_i is the random drop-out time of subject i; $\boldsymbol{c}(t|a,\boldsymbol{\theta^*})$ is defined similarly as in (2.3).

In the following, we will show the doubly weighted estimating function (3.8) is an unbiased estimating function to estimate the parameter $\boldsymbol{\theta^*}$, provided that assumption (B1) and assumption (C1) are satisfied and all the involved models are correctly specified. To show (3.8) is an unbiased estimating function, we need to prove that $E\{\boldsymbol{U}_i^{ww}(\boldsymbol{\theta^*},\boldsymbol{\alpha},\boldsymbol{\zeta})\} = 0$.

We assume that there is no random drop-out for simplicity, i.e. $C_i(t) = 1$ and $dN_i^*(t) = dN_i(t)$ for all $t > 0$. Variable $\boldsymbol{Z_i}(t^-)$ denotes some features of the observed history so that the conditionally independent observation scheme assumption (B1) is satisfied. We assume that $\{A_i, \boldsymbol{V_i}\} \subset \boldsymbol{\mathcal{H}_i^{obs}}(t^-)$ for all t and suppress the parameters $\boldsymbol{\theta^*}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}$ in the notation for convenience. Then,

$$
\begin{aligned}
& E\left\{\sum_{all\ a}\int_0^{\tau_i} w_i^* w_i(t) I(A_i = a)\boldsymbol{c}(t|a)[P_i(t) - S_{T^a}(t)]dN_i(t)\right\} \\
=& \sum_{all\ a}\int_0^{\tau_i} \boldsymbol{c}(t|a) E\left\{\frac{I(A_i = a)}{f(a|\boldsymbol{V_i})}\frac{dt}{E[dN_i(t)|\boldsymbol{Z_i}(t^-)]}[P_i(t) - S_{T^a}(t)]dN_i(t)\right\} \\
=& \sum_{all\ a}\int_0^{\tau_i} \boldsymbol{c}(t|a) E_{A,V,Z,P}\left\{\frac{I(A_i = a)}{f(a|\boldsymbol{V_i})}\frac{dt}{E[dN_i(t)|\boldsymbol{Z_i}(t^-)]}[P_i(t) - S_{T^a}(t)]E_{dN|A,V,P,\mathcal{H}^{obs}}[dN_i(t)]\right\} \\
=& \sum_{all\ a}\int_0^{\tau_i} \boldsymbol{c}(t|a) E_{A,V,Z,P}\left\{\frac{I(A_i = a)}{f(a|\boldsymbol{V_i})}\frac{dt}{E[dN_i(t)|\boldsymbol{Z_i}(t^-)]}\right. \\
& \left.\times [P_i(t) - S_{T^a}(t)]E\left[dN_i(t)|\boldsymbol{Z_i}(t^-)\right]\right\}, \qquad \text{by (B1)} \\
=& \sum_{all\ a}\int_0^{\tau_i} \boldsymbol{c}(t|a) E_{A,V,P}\left\{\frac{I(A_i = a)}{f(a|\boldsymbol{V_i})}[P_i(t) - S_{T^a}(t)]\right\}dt \\
=& \sum_{all\ a}\int_0^{\tau_i} \boldsymbol{c}(t|a) E_{A,V}\left\{\frac{I(A_i = a)}{f(a|\boldsymbol{V_i})}[E(P_i(t)|A_i = a, \boldsymbol{V_i}) - S_{T^a}(t)]\right\}dt \\
=& \sum_{all\ a}\int_0^{\tau_i} \boldsymbol{c}(t|a) E_{A,V}\left\{\frac{I(A_i = a)}{f(a|\boldsymbol{V_i})}[Pr(T_i > t|A_i = a, \boldsymbol{V_i}) - S_{T^a}(t)]\right\}dt \\
=& \sum_{all\ a}\int_0^{\tau_i} \boldsymbol{c}(t|a) E_{A,V}\left\{\frac{I(A_i = a)}{f(a|\boldsymbol{V_i})}[Pr(T_i^a > t|\boldsymbol{V_i}) - S_{T^a}(t)]\right\}dt, \qquad \text{by (C1)} \\
=& \sum_{all\ a}\int_0^{\tau_i} \boldsymbol{c}(t|a) E_V\left\{[Pr(T_i^a > t|\boldsymbol{V_i}) - S_{T^a}(t)]\frac{1}{f(a|\boldsymbol{V_i})}E[I(A_i = a)|\boldsymbol{V_i}]\right\}dt \\
=& \sum_{all\ a}\int_0^{\tau_i} \boldsymbol{c}(t|a) E_V\left\{[Pr(T_i^a > t|\boldsymbol{V_i}) - S_{T^a}(t)]\,\frac{1}{f(a|\boldsymbol{V_i})}f(a|\boldsymbol{V_i})\right\}dt
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{all\ a} \int_0^{\tau_i} \boldsymbol{c}(t|a) \left\{ E_V \left[Pr(T_i^a > t | \boldsymbol{V_i}) \right] - S_{T^a}(t) \right\} dt \\
&= \sum_{all\ a} \int_0^{\tau_i} \boldsymbol{c}(t|a) [Pr(T_i^a > t) - S_{T^a}(t)] dt \\
&= \sum_{all\ a} \int_0^{\tau_i} \boldsymbol{c}(t|a) [S_{T^a}(t) - S_{T^a}(t)] dt \\
&= \mathbf{0},
\end{aligned}
$$

where $P_i(t) = I(T_i > t)$ is the observed response, and $P_i^a(t) = I(T_i^a > t)$ denotes the counterfactual response under hypothetical treatment value or level a.

Let $\boldsymbol{U_{1i}}(\boldsymbol{\theta^*}, \boldsymbol{\alpha}, \boldsymbol{\zeta})$ be $\boldsymbol{U}_i^{ww}$ which is given in (3.8), the doubly weighted estimating function for the outcome process, $\boldsymbol{U_{i2}}(\boldsymbol{\alpha})$ be an unbiased estimating function for the IIV weight $w_i(t)$, and $\boldsymbol{U_{i3}}(\boldsymbol{\zeta})$ be an unbiased estimating function for the IPT weight w_i^*. Then, define $\boldsymbol{U_i} = (\boldsymbol{U}'_{\boldsymbol{1i}}, \boldsymbol{U}'_{\boldsymbol{i2}}, \boldsymbol{U}'_{\boldsymbol{i3}})'$. Since $\boldsymbol{U_i}$ is a vector of unbiased estimating functions, by White (1982), under mild regularity conditions, solving $\sum_{i=1}^n \boldsymbol{U_i}(\boldsymbol{\theta^*}, \boldsymbol{\alpha}, \boldsymbol{\zeta}) = \mathbf{0}$ leads to consistent estimators of parameters $\boldsymbol{\theta^*}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}$, given the outcome model $S_{T^a}(t; \boldsymbol{\theta^*})$ and the models for weight $w_i^*(\boldsymbol{\zeta})$ and weight $w_i(t; \boldsymbol{\alpha})$ are all correctly specified.

Similar to the Theorem 1 given in Section 2.2.5, the proposed doubly weighted estimator of $\boldsymbol{\theta^*}$ has an asymptotically Normal distribution with a sandwich form variance, i.e.

$$
\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta^*}) \xrightarrow{D} N(0, V_{\theta^*}), \tag{3.9}
$$

where V_{θ^*} is the $r \times r$ left upper block of $A^{-1}B\ (A^{-1})'$, r is the dimension of $\boldsymbol{\theta^*}$, and

$$
\begin{aligned}
A &= \begin{pmatrix} -E(\partial \boldsymbol{U}_{i1}/\partial \boldsymbol{\theta}^{*\prime}) & -E(\partial \boldsymbol{U}_{i1}/\partial \boldsymbol{\alpha}') & -E(\partial \boldsymbol{U}_{i1}/\partial \boldsymbol{\zeta}') \\ \mathbf{0} & -E(\partial \boldsymbol{U}_{i2}/\partial \boldsymbol{\alpha}') & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -E(\partial \boldsymbol{U}_{i3}/\partial \boldsymbol{\zeta}') \end{pmatrix} \\
&\triangleq \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ \mathbf{0} & A_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & A_{33} \end{pmatrix}
\end{aligned} \tag{3.10}
$$

and

$$
\begin{aligned}
B &= E\{\boldsymbol{U_i}\ \boldsymbol{U_i'}\} \\
&= \begin{pmatrix} E(\boldsymbol{U}_{i1}\boldsymbol{U}_{i1}') & E(\boldsymbol{U}_{i1}\boldsymbol{U}_{i2}') & E(\boldsymbol{U}_{i1}\boldsymbol{U}_{i3}') \\ E(\boldsymbol{U}_{i2}\boldsymbol{U}_{i1}') & E(\boldsymbol{U}_{i2}\boldsymbol{U}_{i2}') & E(\boldsymbol{U}_{i2}\boldsymbol{U}_{i3}') \\ E(\boldsymbol{U}_{i3}\boldsymbol{U}_{i1}') & E(\boldsymbol{U}_{i3}\boldsymbol{U}_{i2}') & E(\boldsymbol{U}_{i3}\boldsymbol{U}_{i3}') \end{pmatrix} \\
&\triangleq \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix}.
\end{aligned} \tag{3.11}
$$

Matrices A and B can be estimated by

$$
\widehat{A} = \left. \begin{pmatrix} -\frac{1}{n}\sum_{i=1}^{n} \partial \boldsymbol{U}_{i1}/\partial \boldsymbol{\theta}^{*\prime} & -\frac{1}{n}\sum_{i=1}^{n} \partial \boldsymbol{U}_{i1}/\partial \boldsymbol{\alpha}' & -\frac{1}{n}\sum_{i=1}^{n} \partial \boldsymbol{U}_{i1}/\partial \boldsymbol{\zeta}' \\ \mathbf{0} & -\frac{1}{n}\sum_{i=1}^{n} \partial \boldsymbol{U}_{i2}/\partial \boldsymbol{\alpha}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\frac{1}{n}\sum_{i=1}^{n} \partial \boldsymbol{U}_{i3}/\partial \boldsymbol{\zeta}' \end{pmatrix} \right|_{(\widehat{\boldsymbol{\theta}}^*, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\zeta}})},
$$

and $\widehat{B} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{U_i}(\widehat{\boldsymbol{\theta}}^*, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\zeta}})\ \boldsymbol{U_i}(\widehat{\boldsymbol{\theta}}^*, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\zeta}})'$, respectively, where n is the sample size. Details about the estimation of V_{θ^*} will be discussed under specific models in simulation studies.

3.2 Simulation Study

Now, we will empirically study the finite sample performance of the proposed doubly weighted estimator. In Section 2.3, we investigated the IIV weighted estimator in a parametric proportional hazards outcome model. In this section, as an alternative, we consider a structural accelerated failure time (AFT) model, e.g. model (1.8) which was introduced in Section 1.3.4. We assume a Log-normal distribution for failure times T_i, $i = 1, ..., n$. That is, the logarithm of T_i has a Normal location-scale distribution:

$$ln(T_i) = \theta_0 + \theta_1 A_i + \theta_2 L_i + \theta_3 V_i + \sigma W_i, \qquad W_i \sim N(0,1), \tag{3.12}$$

where all the regressors A_i, L_i and V_i are time-fixed. We set $\theta_0 = 0.5$, $\theta_1 = 1$, $\theta_2 = -0.3$, $\theta_3 = -0.3$ and $\sigma = 1.5$ so that approximately $90\% - 95\%$ of the T_i are interval-censored and the rest are right-censored at the last visit before the administrative end of follow-up time $\tau_i = 5$. Factors V_i and L_i are assumed to be associated with shorter failure times, while treatment A_i is supposed to prolong failure times. Here, V_i is a confounder between treatment A_i and outcome T_i, and it is assumed to follow a Normal distribution $N(\mu_v, \sigma_v)$, where $\mu_v = 3$ and $\sigma_v = 1$. Treatment A_i is a binary random variable from $BIN\left(1, \frac{\exp(\zeta_0+\zeta_1 V_i)}{1+\exp(\zeta_0+\zeta_1 V_i)}\right)$. Note that when $\zeta_1 = 0$, treatment A_i is independent of V_i. Otherwise, it is confounded by V_i. We consider two scenarios: $\boldsymbol{\zeta} = (\zeta_0, \zeta_1)' = (0, 0)'$ represents a randomized experiment, and $\boldsymbol{\zeta} = (-6, 2)'$ represents an observational study where treatment is confounded. In either case, the probability of being treated, i.e. $A_i = 1$, is approximately 0.5. Additionally, there is an intermediate variable L_i that affects visit times, with $L_i|A_i \sim N(\beta_0 + \beta_1 A_i, \sigma_l)$, where $\beta_0 = 4$, $\beta_1 = -2$, and $\sigma_l = 1$. Therefore, the

distribution of (A_i, L_i, V_i) in model (3.12) is designed to have:

$$\begin{aligned} f(A_i, L_i, V_i) =& f_L(L_i|A_i, V_i) f_A(A_i|V_i) f_V(V_i) \\ =& f_L(L_i|A_i) f_A(A_i|V_i) f_V(V_i), \end{aligned}$$

where f_L, f_A and f_V are the probability density or mass functions of the relevant random variables.

It can be shown that the (marginal) structural AFT model of T_i^a is given by

$$ln(T_i^a) = \theta_0^* + \theta_1^* a + \sigma^* W, \quad W \sim N(0,1), \tag{3.13}$$

where $\theta_0^* = \theta_0 + \theta_2\beta_0 + \theta_3\mu_v$, $\theta_1^* = \theta_1 + \theta_2\beta_1$ and $\sigma^* = \sqrt{\sigma^2 + \theta_2^2\sigma_l^2 + \theta_3^2\sigma_v^2}$. This can be shown as below by marginalizing over L_i and V_i, assuming A_i is randomized, i.e. $A_i \amalg V_i$:

$$\begin{aligned} E[ln(T_i^a)] =& E[ln(T_i^a)|A_i = a] \\ =& E\{E[ln(T_i)|A_i = a, L_i, V_i]|A_i = a\} \\ =& \theta_0 + \theta_1 a + \theta_2 E[L_i|A_i = a] + \theta_3 E[V_i|A_i = a] \\ =& \theta_0 + \theta_1 a + \theta_2 E[L_i|A_i = a] + \theta_3 E[V_i] \\ =& \theta_0 + \theta_2\beta_0 + \theta_3\mu_v + (\theta_1 + \theta_2\beta_1)a. \end{aligned}$$

Parameter $\boldsymbol{\theta^*} = (\theta_0^*, \theta_1^*, \sigma^*)'$ in (3.13) is the parameter of interest.

To imitate intermittent and dependent visit times, we assume a discrete time visit process, as discussed in Section 2.2.4. Here, we consider the case: failure occurrence terminates the visit process (i.e. CASE II). That is, $dN_i(t)$ is defined at $0 < a_1 < ... < a_M = \tau = 5$, with a time increment of 0.01 and $M = 500$, and the visit process is generated

by

$$
\begin{aligned}
&Pr[dN_{i,m+1}=1|A_i,L_i,V_i,\bar{P}_{im}^{obs}] \\
&=\begin{cases} \exp(\gamma_0+\gamma_1 A_i+\gamma_2 L_i+\gamma_3 V_i), & \text{if } \bar{P}_{im}^{obs}=1 \\ 1, & \text{if } \bar{P}_{im}^{obs}=0 \end{cases} \quad m=0,1,...,M-1, \text{ and } i=1,...,n.
\end{aligned} \tag{3.14}
$$

where $\gamma_0=-5$, $\gamma_1=-0.2$, $\gamma_2=0.2$, and $\gamma_3=0.1$, so that the median length of visit gaps is about $0.45\sim0.50$ for the untreated group, and $0.80\sim0.85$ for the treated group. Therefore, variable $\boldsymbol{Z_i}(t^-)$ in the assumption (B1*) is actually defined by $\{A_i,L_i,V_i,\bar{P}_i^{obs}(t^-)\}$. We assume that there is no random drop-out for simplicity. Similar to the simulation studies in Section 2.3.1, IIV weights $w_i(t)$ are estimated by fitting a semi-Markov proportional hazards model with a piecewise constant baseline for visit gap times, which is defined by

$$
\begin{aligned}
&Pr[dN_{i,m+1}=1|A_i,L_i,V_i,\bar{P}_{im}^{obs};\boldsymbol{\alpha}] \\
&=\begin{cases} \sum_{k=1}^{K}\alpha_k I_k[B(a_{m+1})]\exp(\alpha_{K+1}A_i+\alpha_{K+2}L_i+\alpha_{K+3}V_i), & \text{if } \bar{P}_{im}^{obs}=1 \\ 1, & \text{if } \bar{P}_{im}^{obs}=0 \end{cases},
\end{aligned} \tag{3.15}
$$

where $B(t)$ is the gap time between the most recent past visit prior to t and t; $I_k[B(t)]=I\{B(t)\in(d_{k-1},d_k]\}$, $0=d_0<d_1<,...,<d_K=+\infty$ are the cut-points for the piecewise constant baseline and they are set as $(0,0.40,0.75,1.0,+\infty)$ when sample size is 500 and as $(0,0.25,0.50,0.75,+\infty)$ when sample size equals 1000 or 2000. Parameter $\boldsymbol{\alpha}=(\alpha_1,...,\alpha_7)'$ can be approximately estimated by the R function *phreg* or by solving the following estimating function:

$$
\begin{aligned}
\boldsymbol{U_{i2}}(\boldsymbol{\alpha})=\sum_{j=1}^{m_i^*}\left\{\frac{\partial ln\{\lambda[B(t_{ij})]\}}{\partial\boldsymbol{\alpha}}-\int_0^{B(t_{ij})}\frac{\partial\lambda(s)}{\partial\boldsymbol{\alpha}}ds\right\} \\
-I(\bar{P}_i^{obs}(t_{im_i^*})=1)\int_0^{\tau-t_{im_i^*}}\frac{\partial\lambda(t)}{\partial\boldsymbol{\alpha}}dt,
\end{aligned} \tag{3.16}
$$

where $0 < t_{i1} < ... < t_{ij} < ... < t_{im_i^*} < \tau$ are the actual visits of subject i before a failure is observed, and $\lambda(s) = \sum_{k=1}^{K} \alpha_k I_k(s) \exp(\alpha_{K+1} A_i + \alpha_{K+2} L_i + \alpha_{K+3} V_i)$, $K = 4$. The IIV weights are computed by (2.14) when $\bar{P}_{im}^{obs} = 1$ and are set to be 1 when $\bar{P}_{im}^{obs} = 0$.

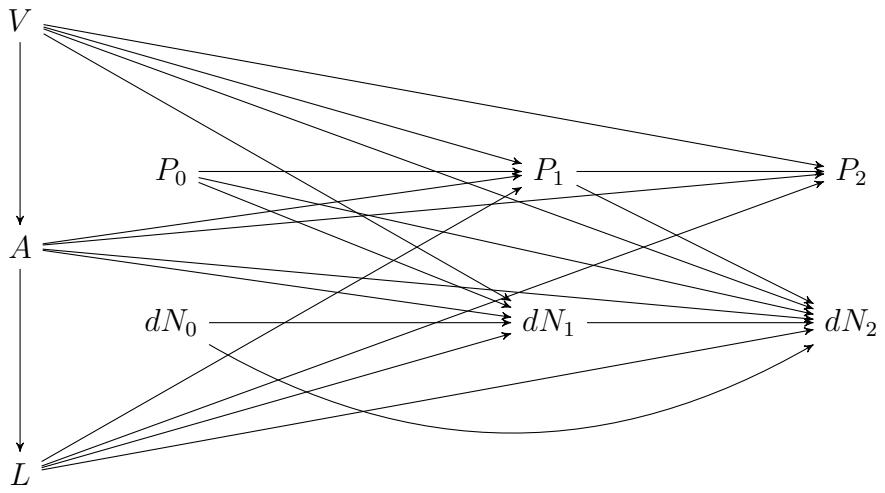


Figure 3.1: DAG for the simulation setting with risk factors A (treatment), L (intermediate variable), and V (confounder), when individuals are assumed to stop visiting after failure occurrence, where $P_m = I(T > a_m)$, with $P_0 = 1$, and dN_m indicates a clinic visit at a_m, $m = 0, 1, 2$.

A DAG is displayed in Figure 3.1 to demonstrate the simulation scenario, where we can see L is an intermediate variable on the pathway from treatment A to outcome $P(t)$, and V is a confounder between A and $P(t)$. Furthermore, the visit process $dN(t)$ and outcome process $P(t)$ share common risk factors, A, L and V. The visit process $dN(t)$ is also associated with $dN(t^-)$ and $P(t^-)$. Therefore, this example corresponds to a longitudinal study with dependent follow-up times and a confounded treatment.

Then, the doubly weighted estimating function of the outcome model for subject i is defined by

$$\boldsymbol{U_{i1}}(\boldsymbol{\theta^*}, \boldsymbol{\alpha}, \boldsymbol{\zeta}) = \sum_{a=0}^{1} \sum_{j=1}^{m_i} w_i^*(\boldsymbol{\zeta}) w_i(t_{ij}; \boldsymbol{\alpha}) I(A_i = a) \boldsymbol{c}(t_{ij}|a; \boldsymbol{\theta^*})[P_i(t_{ij}) - S_{T^a}(t_{ij}; \boldsymbol{\theta^*})], \quad (3.17)$$

where t_{ij}'s include the m_i^* actual visits and $m_i - m_i^*$ pseudo-visits after failure occurred and the survival function of the counterfactual outcome T_i^a based on model (3.13) is given as

$$S_{T^a}(t_{ij}; \boldsymbol{\theta^*}) = 1 - \Phi\left(\frac{ln(t_{ij}) - \theta_0^* - \theta_1^* a}{\sigma^*}\right),$$

where $\Phi(t)$ is the CDF of a standard normal distribution at t. Regarding the IPT weight, parameter $\boldsymbol{\zeta} = (\zeta_0, \zeta_1)'$ in the logistic regression (3.4) can be estimated by software for generalized linear models or by solving the estimating equation $\sum_{i=1}^{n} \boldsymbol{U_{i3}}(\boldsymbol{\zeta}) = 0$ given in (3.5).

Three estimators, (i) the MLE based on likelihood (2.1), (ii) the IIV weighted estimator based on estimating function (2.38), and (iii) the proposed doubly weighted estimator based on estimating function (3.17), will be examined and compared with each other. Function $\boldsymbol{c}(\cdot)$ needed in (ii) and (iii) is given in (2.4), so (ii) and (iii) are weighted GEE estimators with independent working covariance matrices. The MLEs are obtained by R function *survreg* which can handle interval-censored survival data for AFT models and yield robust sandwich variance estimates. The GEE estimates with independent working covariance matrices are obtained by R function *glm* with $\eta = \Phi^{-1}(1 - \mu)$ for binary responses $P_i^a(t)$ and with relevant weights applied, where $\eta = b_0 + b_1 a + b_2 ln(t)$ is the linear predictor and $\mu = E[P_i^a(t)] = S_{T^a}(t; \boldsymbol{\theta^*})$ is the mean function. The reparameterization is $\theta_0^* = -b_0/b_2$, $\theta_1^* = -b_1/b_2$, and $\sigma^* = 1/b_2$.

For each estimator, Bias, ASE, ESE, MSE and CP are summarized in Tables 3.1-3.6 for 500 simulation replicates; we considered three sample sizes, $n = 500$, 1000, and 2000.

The asymptotic variance V_{θ^*} of the proposed doubly weighted GEE estimator is estimated by the sandwich form variance estimator $\widehat{A}^{-1}\widehat{B}\ (\widehat{A}^{-1})'$, where the estimators of $\widehat{A}_{11}$ and $\widehat{A}_{12}$ in (3.10) can be derived similarly as for Section 2.3.1 by (2.28) and (2.31), respectively, and A_{13} can be estimated by

$$\widehat{A}_{13} = \frac{1}{n}\sum_{i=1}^{n}\sum_{a=0}^{1}\sum_{j=1}^{m_i}(2a-1)w_i(t_{ij};\widehat{\boldsymbol{\alpha}})w_i^{*2}(\widehat{\boldsymbol{\zeta}})\ \frac{\partial p(\boldsymbol{V_i}|\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}'}\bigg|_{\boldsymbol{\zeta}=\widehat{\boldsymbol{\zeta}}}\boldsymbol{c}(t_{ij}|a;\widehat{\boldsymbol{\theta}^*})[P_i(t_{ij}) - S_{T^a}(t_{ij};\widehat{\boldsymbol{\theta}^*})]. \tag{3.18}$$

Since $\boldsymbol{U_{i2}}$ and $\boldsymbol{U_{i3}}$ are score functions, $A_{22} = B_{22}$ and $A_{33} = B_{33}$ can be estimated by the observed Fisher information matrices provided by *phreg* and *glm*, respectively.

From Tables 3.1-3.6, we see that MLEs are biased both when treatment is randomized and confounded. Especially, when A is confounded by V, i.e. in Tables 3.2, 3.4 and 3.6 when $\boldsymbol{\zeta} = (-6, 2)'$, the bias produced by MLEs is very large because likelihood (2.1) fails to take the informative inspection times as well as the confounded treatment into account. Tables 3.1, 3.3 and 3.5 show that the IIV weighted estimator results in negligible bias and good coverage probabilities around the 95% nominal level. However, in Tables 3.2, 3.4 and 3.6, the IIV weighted estimates which do not adjust for confounding have some selection bias caused by the non-randomized treatment. For the doubly weighted estimator, when treatment is randomized, from Tables 3.1, 3.3 and 3.5, we see that bias is negligible and coverage is good; when treatment is not randomized, from Tables 3.2, 3.4 and 3.6, bias is mush less than that for the IIV weighted estimator, and coverage becomes close to the nominal level when sample size grows. Slightly low coverage of the doubly weighted estimates in Tables 3.2, 3.4 and 3.6 reflects the fact that ASEs are a bit smaller than the ESEs, but when sample size increases the difference becomes smaller and coverage is closer to 95%. Robins et al. (2000) commented that large variability in the weights can result in weighted estimators with large variances. In our study, there are some extremely large

weights, especially when $w_i(t)$ and w_i^* are combined for the doubly weighted estimator. How to deal with large variability in weights and improve the weighted estimators is discussed in Chapter 6.

In this chapter, we applied the double weighting method to estimating functions for parametric models. In fact, it can be applied to non- or semi-parametric estimation based on estimating equations as well. We will illustrate that in the analyses of the PsA data in Chapter 5.

Table 3.1: Bias, average of asymptotic standard errors (ASE), empirical standard error (ESE), mean squared error (MSE) and coverage probability (CP) for the case where A_i is randomized, i.e. $\boldsymbol{\zeta} = (0, 0)'$. Sample size: $n = 500$. Number of replicates: $N = 500$.

		TRUE VALUE	BIAS	ASE	ESE	MSE	CP
MLE	θ_0^*	-1.600	0.109	0.108	0.097	0.021	0.85
	θ_1^*	1.600	-0.124	0.152	0.141	0.035	0.88
	$ln(\sigma^*)$	0.444	-0.079	0.040	0.039	0.008	0.48
IIV Weighted	θ_0^*	-1.600	0.005	0.150	0.141	0.020	0.96
	θ_1^*	1.600	-0.034	0.189	0.178	0.033	0.95
	$ln(\sigma^*)$	0.444	-0.014	0.062	0.065	0.004	0.93
Doubly Weighted	θ_0^*	-1.600	0.003	0.150	0.140	0.020	0.96
	θ_1^*	1.600	-0.028	0.187	0.178	0.032	0.95
	$ln(\sigma^*)$	0.444	-0.014	0.062	0.065	0.004	0.94

Table 3.2: Bias, average of asymptotic standard errors (ASE), empirical standard error (ESE), mean squared error (MSE) and coverage probability (CP) for the case where A_i is confounded by V_i, i.e. $\boldsymbol{\zeta} = (-6, 2)'$. Sample size: $n = 500$. Number of replicates: $N = 500$.

		TRUE VALUE	BIAS	ASE	ESE	MSE	CP
MLE	θ_0^*	-1.600	0.261	0.108	0.108	0.080	0.33
	θ_1^*	1.600	-0.440	0.151	0.155	0.218	0.18
	$ln(\sigma^*)$	0.444	-0.073	0.039	0.039	0.007	0.52
IIV Weighted	θ_0^*	-1.600	0.176	0.144	0.155	0.055	0.73
	θ_1^*	1.600	-0.371	0.184	0.197	0.176	0.46
	$ln(\sigma^*)$	0.444	-0.017	0.061	0.066	0.005	0.92
Doubly Weighted	θ_0^*	-1.600	0.015	0.227	0.270	0.073	0.91
	θ_1^*	1.600	-0.061	0.295	0.360	0.133	0.91
	$ln(\sigma^*)$	0.444	-0.015	0.090	0.110	0.012	0.90

Table 3.3: Bias, average of asymptotic standard errors (ASE), empirical standard error (ESE), mean squared error (MSE) and coverage probability (CP) for the case where A_i is randomized, i.e. $\boldsymbol{\zeta} = (0, 0)'$. Sample size: $n = 1000$. Number of replicates: $N = 500$.

		TRUE VALUE	BIAS	ASE	ESE	MSE	CP
MLE	θ_0^*	-1.600	0.104	0.077	0.075	0.016	0.73
	θ_1^*	1.600	-0.117	0.107	0.103	0.024	0.81
	$ln(\sigma^*)$	0.444	-0.076	0.028	0.028	0.007	0.22
IIV Weighted	θ_0^*	-1.600	-0.005	0.107	0.111	0.012	0.93
	θ_1^*	1.600	-0.020	0.134	0.132	0.018	0.94
	$ln(\sigma^*)$	0.444	-0.009	0.044	0.045	0.002	0.95
Doubly Weighted	θ_0^*	-1.600	-0.006	0.106	0.110	0.012	0.94
	θ_1^*	1.600	-0.018	0.133	0.130	0.017	0.95
	$ln(\sigma^*)$	0.444	-0.009	0.044	0.045	0.002	0.95

Table 3.4: Bias, average of asymptotic standard errors (ASE), empirical standard error (ESE), mean squared error (MSE) and coverage probability (CP) for the case where A_i is confounded by V_i, i.e. $\boldsymbol{\zeta} = (-6, 2)'$. Sample size: $n = 1000$. Number of replicates: $N = 500$.

		TRUE VALUE	BIAS	ASE	ESE	MSE	CP
MLE	θ_0^*	-1.600	0.263	0.076	0.071	0.074	0.06
	θ_1^*	1.600	-0.446	0.107	0.100	0.209	0.01
	$ln(\sigma^*)$	0.444	-0.073	0.027	0.027	0.006	0.23
IIV Weighted	θ_0^*	-1.600	0.181	0.102	0.099	0.043	0.54
	θ_1^*	1.600	-0.380	0.130	0.126	0.161	0.16
	$ln(\sigma^*)$	0.444	-0.017	0.043	0.043	0.002	0.94
Doubly Weighted	θ_0^*	-1.600	0.005	0.171	0.204	0.041	0.93
	θ_1^*	1.600	-0.026	0.225	0.259	0.068	0.93
	$ln(\sigma^*)$	0.444	-0.012	0.069	0.085	0.007	0.91

Table 3.5: Bias, average of asymptotic standard errors (ASE), empirical standard error (ESE), mean squared error (MSE) and coverage probability (CP) for the case where A_i is randomized, i.e. $\boldsymbol{\zeta} = (0, 0)'$. Sample size: $n = 2000$. Number of replicates: $N = 500$.

		TRUE VALUE	BIAS	ASE	ESE	MSE	CP
MLE	θ_0^*	-1.600	0.111	0.054	0.050	0.015	0.47
	θ_1^*	1.600	-0.124	0.076	0.071	0.020	0.64
	$ln(\sigma^*)$	0.444	-0.074	0.020	0.018	0.006	0.02
IIV Weighted	θ_0^*	-1.600	0.010	0.075	0.073	0.005	0.95
	θ_1^*	1.600	-0.036	0.095	0.090	0.009	0.94
	$ln(\sigma^*)$	0.444	-0.006	0.031	0.031	0.001	0.94
Doubly Weighted	θ_0^*	-1.600	0.008	0.075	0.073	0.005	0.95
	θ_1^*	1.600	-0.033	0.094	0.090	0.009	0.94
	$ln(\sigma^*)$	0.444	-0.006	0.031	0.031	0.001	0.94

Table 3.6: Bias, average of asymptotic standard errors (ASE), empirical standard error (ESE), mean squared error (MSE) and coverage probability (CP) for the case where A_i is confounded by V_i, i.e. $\boldsymbol{\zeta} = (-6, 2)'$. Sample size: $n = 2000$. Number of replicates: $N = 500$.

		TRUE VALUE	BIAS	ASE	ESE	MSE	CP
MLE	θ_0^*	-1.600	0.264	0.054	0.051	0.072	0.00
	θ_1^*	1.600	-0.448	0.076	0.074	0.206	0.00
	$ln(\sigma^*)$	0.444	-0.069	0.019	0.019	0.005	0.03
IIV Weighted	θ_0^*	-1.600	0.184	0.072	0.071	0.039	0.30
	θ_1^*	1.600	-0.383	0.092	0.092	0.155	0.01
	$ln(\sigma^*)$	0.444	-0.012	0.031	0.030	0.001	0.94
Doubly Weighted	θ_0^*	-1.600	0.013	0.121	0.134	0.018	0.96
	θ_1^*	1.600	-0.044	0.160	0.166	0.029	0.92
	$ln(\sigma^*)$	0.444	-0.010	0.049	0.053	0.003	0.94

Chapter 4

NON- AND SEMI-PARAMETRIC ESTIMATION FOR INTERVAL-CENSORED FAILURE TIME DATA WITH DEPENDENT INSPECTION TIMES

In Chapter 2, we discussed parametric estimation of the distributions of failure times, in the case where failure times could be dependently interval-censored due to intermittent and outcome-dependent inspection times. This chapter focuses on non- and semi-parametric estimation for interval-censored failure times with intermittent observation. The inverse-intensity-of-visit (IIV) weighted estimating function approach will be extended and applied.

4.1 IIV Weighted Non-Parametric Estimation of Distribution Functions

Non-parametric estimation of survival functions or distribution functions is important in the analysis of lifetime or failure time data. First, a graphical display based on non-parametric estimation of a distribution function or survival function can help one choose models for fitting data or check certain model assumptions. For example, the proportional hazards assumption which is needed for a Cox model can be graphically checked by the $ln(-ln)$ transformation of survival curves. If the assumption is satisfied, then the plot of $ln\{-ln[S_T(t)]\}$ versus time t or $ln(t)$ should show roughly vertical parallel curves for individuals grouped according to covariate values (Lawless, 2003). Non-parametric estimates of survival or distribution functions can also be employed for robustly estimating distribution quantities, e.g. median or percentiles, and for multi-sample comparison with respect to a particular risk factor without model fitting.

As discussed in the preceding chapters, the assessment of failure time distributions for the whole population or subgroups stratified by some fixed variable is often of interest, e.g. the marginal CDF, $F_T(t)$, or $F_T(t|A)$ for $t > 0$, of failure time T. To validly estimate the CDF, $F_T(t)$, or survival function, $S_T(t)$, by standard methods such as the Kaplan-Meier estimator for right-censored data, we need censoring times to be independent of failure times. For interval-censored data, when inspection times are independent of failure times, failure times are independently interval-censored. Then, Turnbull (1976)'s non-parametric estimator of $F_T(t)$ is obtained by maximizing likelihood (2.1). The resulting estimate $\widehat{F}_T(t)$ has positive support in specified intervals according to the observed interval-censored observations but may be undefined over some intervals (Lawless, 2003). The validity of this estimator relies on inspection times being independent of failure times. When inspection

times t_{il} and t_{ir} in likelihood (1.3), or t_{ij} and $t_{i,j+1}$ in likelihood (2.1), are not marginally independent of failure times, estimators could be inconsistent, as we have shown in Section 2.3.1 in the simulations for parametric models.

Non-parametric estimation for dependently right-censored failure time data has been heavily discussed in the literature via a variety of techniques: e.g. inverse probability weighting by Robins (1993); Wang and Wells (1998); Robins and Finkelstein (2000); Satten et al. (2001); Hajducek and Lawless (2013), and EM algorithm by Finkelstein et al. (2002); Zhang et al. (2007); Chen et al. (2010). However, dependent interval censoring has received limited attention so far. van der Laan and Robins (1998) considered this for current status data, which involves a single observation time per individual. van der Laan and Hubbard (1997) also considered current status data but did not develop estimation methods for more general interval-censored data. Their methods assume that there exists a covariate $Z(t)$ such that given its history, the visit or observation time process is ignorable. Alternatively, introducing latent variable(s) to connect the failure time process and the visit process, which was reviewed in Chapter 2, is another way to deal with dependent interval censoring. Additionally, Park et al. (2006) considered dependent censoring as a competing risk of failure and their method for nonparametric inference can be used for dependently right-censored as well as dependently interval-censored data. In the following, we will develop an IIV weighted estimating function approach for the nonparametric estimation of distribution function $F_T(t)$ so that dependent interval censoring caused by intermittent visits can be handled.

4.1.1 Monotone-Smoothed IIV Weighted Estimators of Distribution Functions

Suppose that we are interested in the inference about the distribution function $F_T(t)$ of a failure time outcome T. Estimator of the survival function $S_T(t)$ can be simply obtained by $1 - \widehat{F}_T(t)$. We define a set of finite discrete assessment times where $F_T(t)$ will firstly be estimated, denoted by $0 < s_1, ..., < s_m \leq \tau$. We assume that the administrative end of follow-up time, τ, is the same for all individuals for simplicity. Later, we will refine the crude estimates.

Let $\theta_l = F_T(s_\ell)$, where $\ell = 1, ..., m$. First, we assume that the visit process is not terminated by the occurrence of failure and that visit times are continuous. Then, motivated by (2.13), for estimating $\boldsymbol{\theta} = (\theta_1, ..., \theta_m)'$, with $\boldsymbol{c}(t; \theta_\ell) = \{\theta_\ell(1 - \theta_\ell)\}^{-1}$, we define IIV weighted estimating functions for $\boldsymbol{\theta}$ by

$$U_\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^\tau \frac{K_b(s_\ell - t) w_i(t)}{\theta_\ell(1 - \theta_\ell)} [Y_i(t) - \theta_\ell] dN_i^*(t), \qquad \ell = 1, ..., m, \tag{4.1}$$

where $Y_i(t) = I(T_i \leq t)$, $dN_i^*(t) = dN_i(t)C_i(t)$, $dN_i(t)$ is the indicator of a visit at t, $C_i(t) = I(C_i > t)$ and C_i denotes a random drop-out time for subject i. The function $K_b(u) = K(u/b_m)/b_m$, where $K(\cdot)$ is a bounded kernel function defined to be zero outside $[-1, 1]$, e.g. uniform (rectangular), triangular, or Epanechnikov (quadratic), and $\{b_m\}$ denotes a positive bandwidth sequence converging to 0 when $\sum m_i \to +\infty$, where m_i denotes the total number of visits for subject i.

Note that (4.1) is not an unbiased estimating function, even given the conditionally independent observation assumption (B1) is true. That is, the solution to $E(U_{i\ell}) = 0$ gives $\theta_\ell^* = \frac{\int_0^\tau K_b(s_\ell - t) F_T(t) S_C(t) dt}{\int_0^\tau K_b(s_\ell - t) S_C(t) dt} \neq \theta_\ell$, provided that (B1) holds and the random drop-out is independent (i.e. $Y_i(t)$ II $C_i(t)$, $\forall t > 0$), where $S_C(t) = Pr(C_i > t) = E[C_i(t)]$.

Nevertheless, when the number of visits increases, i.e. $\sum m_i \to +\infty$, $b_m \downarrow 0$, and the estimates solved from (4.1) approach θ_ℓ, i.e. $F_T(s_\ell)$. Small bandwidths result in small bias at the cost of large variance since fewer visits lie in $[s_\ell - b_m, s_\ell + b_m]$ and contribute to the estimation of $F_T(s_\ell)$. In practice, people widely use data-driven bandwidth algorithms to do bandwidth selection, e.g. least-squares cross-validation (Hall et al., 2004). A wise selection of bandwidth can lead to good precision of estimation. In the simulation studies in Section 4.1.2 and the real data analysis in Chapter 5, we simply chose bandwidths to have about $50 - 100$ visits fall in the window $[s_\ell - b_m, s_\ell + b_m]$ so that the variances of estimates are moderate and normal approximations are accurate.

When failure occurrence terminates the visit process, motivated by (2.23) in Section 2.2.4, the estimating function for $\boldsymbol{\theta}$ with a discrete time visit process can be written by

$$U_\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{k=1}^{M} \frac{K_b(s_\ell - a_k) w_i(a_k)}{\theta_\ell(1-\theta_\ell)} [Y_i(a_k) - \theta_\ell] dN_i^*(a_k), \qquad \ell = 1, ..., m, \tag{4.2}$$

where $\{dN_i(a_k) : \ k = 1, ..., M\}$ is a discrete time visit process with $0 < a_1 < ... < a_M = \tau$. Also, we have $dN_i(a_k) = 1$ and $w_i(a_k) = 1$, when $\bar{Y}_i^{obs}(a_{k-1}) = 1$ since responses $Y_i(a_k)$'s are all known as 1 after the visit following failure occurrence. Let m_i^* denote the number of actual visits when $\bar{Y}_i^{obs}(a_{k-1}) = 0$ and $m_i - m_i^*$ denote the number of pseudo visits when $\bar{Y}_i^{obs}(a_{k-1}) = 1$ for subject i.

The estimating equation $U_\ell(\boldsymbol{\theta}) = 0$ given in (4.1) or (4.2) yields a closed form estimate for $F_T(s_\ell), \ell = 1, \ldots, m$, i.e.

$$\bar{Y}_\ell = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m_i} K(\frac{s_\ell - t_{ij}}{b_m}) w_i(t_{ij}) Y_i(t_{ij})}{\sum_{i=1}^{n} \sum_{j=1}^{m_i} K(\frac{s_\ell - t_{ij}}{b_m}) w_i(t_{ij})}, \qquad \ell = 1, \ldots, m, \tag{4.3}$$

where $\{t_{i1}, ..., t_{im_i}\} = \{t > 0 : dN_i^*(t) = 1\}$, which denote the m_i actual visit times for subject i in (4.1) and include both the m_i^* actual visit times and the $m_i - m_i^*$ pseudo visit

times which follow failure occurrence in (4.2). From (4.3), we see only visits which lie in $[s_\ell - b_m, s_\ell + b_m]$ contribute to the estimation of $F_T(s_\ell)$ and only the s_ℓ with at least one visit contained in $[s_\ell - b_m, s_\ell + b_m]$ can be estimated. In an extreme case where bandwidth equals zero and t_{ij}'s are all distinct, each solution $\bar{Y}_\ell$ to $U_\ell(\boldsymbol{\theta}) = 0$ is either 0 or 1.

Note that the estimates given in (4.3) are non-monotone in general, so we call them crude or raw estimates, and techniques for monotone smoothing will now be adopted. Isotonic regression is one of the most commonly used methods for achieving monotonicity; it yields non-decreasing estimates by minimizing a weighted sum of squares under a non-decreasing constraint. However, the fitted values provided by isotonic regression are generally step functions, so we consider the combination of an isotonic regression followed by a kernel non-parametric regression, which was proposed by Mukerjee (1988) and recently applied by Datta and Sundaram (2006) to multistage models with current status data. As discussed in He and Shi (1998), theoretically, monotone smoothing can be implemented by combining isotonic regression with any smoothing tools, e.g. kernel or spline, and smoothing can be done either before or after isotonic regression.

Analogous to Sun (2006) and Zhang and Sun (2010) where isotonic regression is applied to estimate distribution functions or survival functions with current status data, the isotonic regression problem in our context is to minimize

$$Q_w(\boldsymbol{\theta}) = \sum_{\ell=1}^{m} \sum_{i=1}^{n} \sum_{j=1}^{m_i} K[(s_\ell - t_{ij})/b_m] w_i(t_{ij}) \{Y_i(t_{ij}) - \theta_\ell\}^2$$

subject to $\theta_1 \leq ... \leq \theta_m$. This is equivalent to minimizing

$$Q_w^*(\boldsymbol{\theta}) = \sum_{\ell=1}^{m} w_+(s_\ell) \{\bar{Y}_\ell - \theta_\ell\}^2, \qquad \text{for } \ell = 1, ..., m, \tag{4.4}$$

subject to the non-decreasing constraint, where

$$\bar{Y}_\ell = \sum_{i=1}^{n} \sum_{j=1}^{m_i} K[(s_\ell - t_{ij})/b_m] w_i(t_{ij}) Y_i(t_{ij}) / w_+(s_l),$$

and

$$w_+(s_\ell) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} K[(s_\ell - t_{ij})/b_m] w_i(t_{ij}).$$

The above equation can be solved by the max-min formula (Barlow et al., 1972; Robertson et al., 1988) for isotonic regression:

$$\tilde{F}_T(s_\ell) = \max_{r \le \ell} \min_{u \ge \ell} \left\{ \frac{\sum_{v=r}^{u} w_+(s_v) \bar{Y}_v}{\sum_{v=r}^{u} w_+(s_v)} \right\}, \tag{4.5}$$

or by software based on the well known Pool-Adjacent-Violators Algorithm (PAVA) (Miles, 1959). More details about isotonic regression and PAVA can be found in Barlow et al. (1972), Robertson et al. (1988) and de Leeuw et al. (2009).

The next step is to implement kernel non-parametric regression so that the estimated survival curve of T is smooth and the estimates of $F_T(t)$ can be obtained at any t rather than only at s_ℓ's. We consider the local-constant estimator also known as the Nadaraya-Watson estimator (Nadaraya, 1964; Watson, 1964), which is given below, with a log-concave kernel function $K^*(\cdot)$ to smooth the estimates yielded by isotonic regression, i.e. $\tilde{F}_T(a_\ell)$, $\ell = 1, ..., m$.

$$\widehat{F}_T(t) = \frac{\sum_{\ell=1}^{m} \tilde{F}_T(s_\ell) K^*(\frac{s_\ell - t}{h_m})}{\sum_{\ell=1}^{m} K^*(\frac{s_\ell - t}{h_m})}, \qquad \forall t > 0, \tag{4.6}$$

where kernel $K^*(\cdot)$ is a log-concave density so that the monotonicity of (4.6) is retained (Mukerjee, 1988) and $\{h_m\}$ is a positive bandwidth sequence converging to 0 when $\sum m_i \to +\infty$. It is found that bias is mainly caused by a poor choice of b_m in (4.3), since (4.1) or (4.2) is not an unbiased estimating function for θ_ℓ or $F_T(t)$, though they are asymptotically unbiased when $b_m \downarrow 0$. The selection of of h_m in (4.6) is not sensitive to the final estimated survival curves, so one can select $\{h_m\}$ to make the final estimates of survival curves have reasonable curvature. Asymptotics are much harder here than in standard settings, e.g.

Mukerjee (1988), involving a single kernel estimator. In the proposed two-stage estimation procedure, we employ two kernels; especially in the first stage, the crude estimate $\bar{Y}_\ell$ solved from (4.1) is not a consistent estimator of θ_ℓ but its limiting value θ_ℓ^* approaches θ_ℓ when bandwidth b_m goes to 0. Then, in the second stage we apply isotonic regression and kernel again to achieve a monotone smoothing, which is extended from Mukerjee (1988) and Datta and Sundaram (2006) to the present setting. Thus, the whole procedure demands non-standard asymptotics for the final estimate $\widehat{F}_T(t)$. Instead, we consider the bootstrap to estimate the standard error of $\widehat{F}_T(t)$ in simulation and real data analysis.

4.1.2 Simulation Study

To demonstrate the proposed method and examine the finite sample performance of resulting estimates, a simulation study is conducted for the non-parametric estimation of $F_T(t|A)$ for two treatment groups separately, $A = 0$ and $A = 1$. The simulation design is basically the same as in Section 2.3 for parametric estimation, and we will discuss two cases as usual: failure occurrence does not terminate visits (CASE I) and failure occurrence terminates visits (CASE II). Since bootstrap is very computationally intensive, the sample size is set as 200 in total (about 100 for each treatment group) and the number of simulation replicates is 500; 100 bootstrap samples are used to estimate a standard error.

At $t = 0$, treatment variable A is generated from $BIN(1, p)$ with $p = 0.5$. The administrative end of follow-up time, τ, is set to be 5. For technical simplicity, time is discretized in this simulation study for both CASE I and CASE II, with an increment of 0.01, so (4.2) will be used for crude estimates and M in (4.2) equals 500. Then for any a_k, $k = 0, ..., M$ with $a_0 = 0$ and $a_M = 5$, an ancillary variable L_k is generated by

$$L_k|A, \bar{L}_{k-1} \sim N(\beta_1 A, \sigma_l), \tag{4.7}$$

with $\beta_1 = -0.5$ and $\sigma_l = 1$ so that treatment A is an inhibitor for L_k. Then, response Y_{k+1} is generated by

$$Pr[Y_{k+1} = 1|\bar{L}_k, A, Y_k = 0] = \Phi(\eta_0 + \eta_1 L_k + \eta_2 A), \tag{4.8}$$

where Φ denotes the CDF of a standard normal distribution $N(0,1)$ and $Y_k = I(T \leq a_k)$. Parameters η_0, η_1 and η_2 are given by -2.8, 0.5 and -0.1, respectively, so that treatment A has a negative effect on the risk of failure and L_k is an intermediate variable between A and failure time which has a positive strong effect on the risk of failure but is inhibited by treatment. As a result, 98% of untreated individuals ($A = 0$) and 80% of treated individuals ($A = 1$) fail before τ. After marginalizing over L_k, the marginal outcome model can be presented as

$$Pr[Y_{k+1} = 0|A, Y_k = 1] = e^{\psi_0} \exp(\psi_1 A), \tag{4.9}$$

where $c = 1/\sqrt{1 + \eta_1^2 \sigma_l^2}$, $\psi_0 = ln\left\{\Phi(c \cdot \eta_0)\right\} = -5.09$, $\psi_1 = ln\left\{\frac{\Phi[c\cdot(\eta_0+\eta_2+\eta_1\beta_1)]}{\Phi(c\cdot\eta_0)}\right\} = -0.93$, so there is about a 60% risk deduction for $A = 1$ versus $A = 0$.

For the case where visits continue after failure occurs (CASE I), the discrete time visit process, $\{dN_k : \ k = 1, ..., M\}$, is generated by a Markov proportional intensities model with a constant baseline intensity:

$$\begin{aligned} Pr[dN_{k+1} = 1|A, \bar{L}_k] &= \exp(\gamma_0 + \gamma_1 L_k + \gamma_2 A) \\ &= e^{\gamma_0} \exp(\gamma_1 L_k + \gamma_2 A), \qquad k = 0, 1, ..., M-1, \end{aligned} \tag{4.10}$$

where $\gamma_0 = -3.5$, $\gamma_1 = 0.5$, and $\gamma_2 = -0.2$ so that the average visit gap time is approximately 0.30 for the untreated group ($A = 0$) and about 0.44 for the treated group ($A = 1$). From model (4.10), we see that the visit process is outcome-dependent via A as well as $\bar{L}_k$. For the case where failure occurrence stops visits (CASE II), $\{dN_k : k = 1, ..., M\}$ are

generated by

$$
\begin{aligned}
&Pr[dN_{k+1}=1|A,\bar{L}_k,\bar{Y}_k^{obs}] \\
&=\begin{cases} \exp(\gamma_0+\gamma_1 L_k+\gamma_2 A), & \text{if } \bar{Y}_k^{obs}=0 \\ 1, & \text{if } \bar{Y}_k^{obs}=1 \end{cases}, \qquad k=0,1,...,M-1. \qquad (4.11)
\end{aligned}
$$

To estimate $F_T(t|A)$ for a given A, we have proposed a two-stage estimation procedure in the preceding subsection. First, we estimate the inverse-intensity-of-visit (IIV) weights $w_i(t)$ for the crude estimator given in (4.3), which can be done by fitting a piecewise constant proportional hazards semi-Markov model given in (2.39) for the gap times of "actual" visits. Here, we use R function *phreg* with $(0.40, 0.75, 1.0)$ as cut-points to estimate the IIV weights. In addition, we let the kernel function in (4.3), $K(\cdot)$, be the Epanechnikov (EP) kernel, i.e. $k(x)=\frac{3}{4}(1-x^2)I(|x|\leq 1)$, which is bounded and smooth. The bandwidth b_m in (4.3) is selected to be 0.125, so $m=20$. Then, there are about 70 actual visits contained in interval $[s_\ell - b_m, s_\ell + b_m]$ for crudely estimating $F_T(s_\ell)$ in CASE I. Second, the isotonic regression of crude estimates $(\bar{Y}_1,...,\bar{Y}_m)$ with weights $(w_+(s_1),...,w_+(s_m))$ is implemented by the R function *monoreg* in package *fdrtool*. Finally, kernel smoothing for the monotone estimates $\tilde{F}_T(s_\ell)$'s is done with a standard normal (Gaussian) kernel given for $K^*(\cdot)$ and a bandwidth h_m specified as 0.2 for the CASE I and 0.3 for the CASE II, which can be implemented by the R function *npreg* in the package *np*. Simulation results are summarized and the plots of estimated distribution function curves are shown for $A=1$ versus $A=0$ in the following.

CASE I: Failure occurrence does not terminate visits

Figure 4.1 shows the mean raw estimate $\bar{Y}_\ell$ at each s_ℓ, $s_\ell=1,...,m$, in 500 simulation replicates, for different m or b_m. We can see bias is negligible in each plot, so estimation is

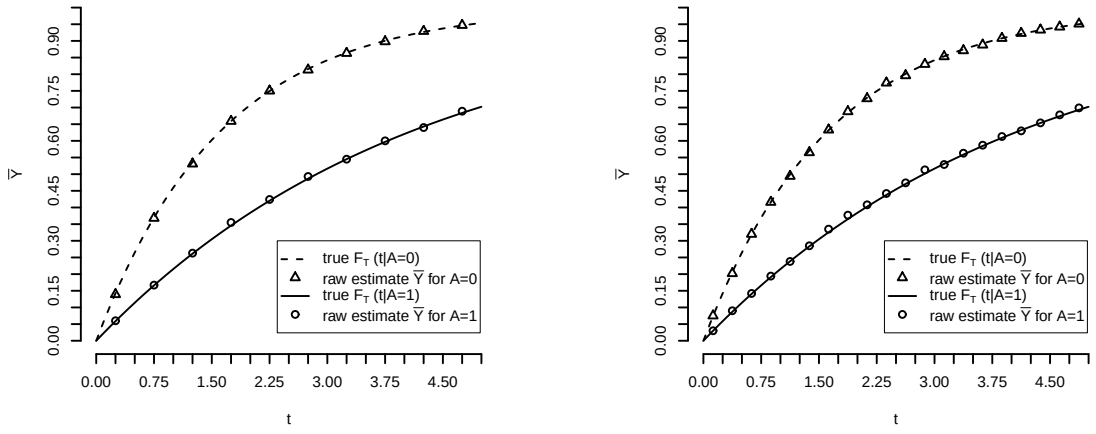
not very sensitive to bandwidth selection, when the number of visits around s_ℓ is sufficient. Therefore, we let $m = 20$, i.e. $b_m = 0.125$ in the following. Figure 4.2 and Figure 4.3 demonstrate the crude estimates $\bar{Y}_\ell$, isotonic estimates $\tilde{F}_T(s_\ell)$ and kernel-smoothed isotonic estimates $\widehat{F}_T(t)$ for $A = 0$ and 1 in one simulation sample, respectively. We see that isotonic regression monotonizes the raw estimates and then the kernel smoothing with a bandwidth of 0.2 produces smooth distribution function curves. It is seen that the raw estimates in Figure 4.2 have relatively smaller variability than those in Figure 4.3, because untreated individuals have more visits than the treated ones by design. Also, the raw estimates for $A = 0$ in Figure 4.2 are close to monotone, while isotonic regression corrects relatively more the raw estimates for $A = 1$ in Figure 4.3. For either group, the kernel-smoothed isotonic estimate of distribution function, denoted by a solid curve agrees quite well with the true distribution function curve denoted by a dashed curve. Figure 4.4 displays the estimated and true distribution functions of the two groups as a comparison, where we can see that for sample size $n = 200$ and the visit frequency for the treated ($A = 1$) group, estimate of $F_T(t|A = 1)$ will not be really smooth. Additionally, note that Figures 4.2-4.4 display the performance of estimates in one simulation sample to demonstrate the two-stage estimation procedure.

Table 4.1 summarizes the kernel-smoothed isotonic estimates of $F_T(t|A)$ by (4.6) at $t = 0.5, 1.0, ..., 4.5, 5.0$, for the two groups across the 500 simulation samples and provides 95% pointwise empirical coverage probabilities (ECPs) to draw statistical conclusions. Standard errors are estimated by non-parametric bootstraps, which can be implemented by resampling with replacement using R function *sample*. The bootstrap sample size is set to be 100. All the true values, estimates and standard errors are reported for $F_T(t|A)$, while coverage probabilities and Z tests are presented after a $ln(-ln(F_T(t|A)))$ transformation. If we let $ln(-ln(\widehat{F}_T(t|A)))$ be $\widehat{\vartheta}(t)$, then an empirical confidence interval is computed by

$\widehat{\vartheta}(t) \pm 1.96\ se(\widehat{\vartheta}(t))$ where $se(\widehat{\vartheta}(t))$ is the bootstrap estimated standard error of $\widehat{\vartheta}(t)$. Approximate normality is checked in Figure 4.5 by histograms and QQ-plots of the pooled Z tests across time, which are defined by $z(t) = \frac{\widehat{\vartheta}(t)-\vartheta(t)}{se(\widehat{\vartheta}(t))}$ for all $t = 0.5, 1.0, ..., 4.5, 5.0$. We see that estimates in either group have good coverage, except for the regions close to $t = 0$ or 5. The mean of bootstrap standard errors (BSEs) underestimate the empirical standard errors (ESEs) slightly, which may be explainable by the small sample size (i.e. 200) and should improve when sample size n increases. Poor coverage for $t = 0.5$ and $t = 5.0$ is caused by the greater underestimation of ESE. For example, at $t = 5.0$, approximately 98% of individuals fail in the untreated group, i.e. $F_T(t|A = 0)$ is close to 1. Also, we know that kernel smoothing usually does not perform perfectly close to bounds, so underestimation of standard errors and low coverage could result. Poor normal approximation close to $t = 0$ and $t = 5$ can be found in Figure 4.5 as well. Figure 4.6 shows the truncated histograms and QQ-plots for $t = 1.0-4.0$, and we see that estimates appear to be normally distributed over this range.

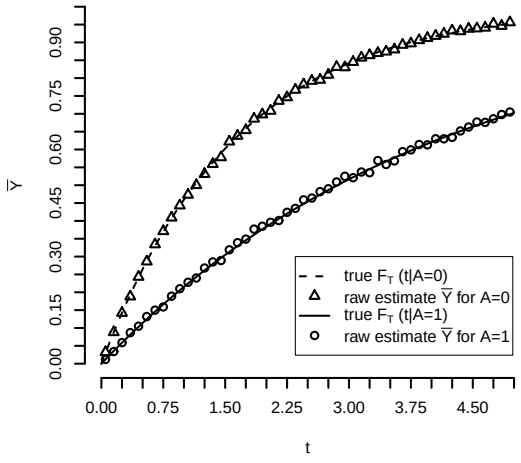
CASE II: Failure occurrence terminates visits

For the case where failure occurrence discontinues the visit process, Figure 4.7 shows the mean raw estimate $\bar{Y}_\ell$ for different bandwidth b_m selections in the 500 simulations. Each plot shows that the raw estimates are essentially unbiased at each t. In addition, Figure 4.8 and Figure 4.9 show the raw estimates, isotonic estimates and kernel-smoothed isotonic estimates (final estimates) for the untreated group ($A = 0$) and the treated group ($A = 1$) in one simulation sample, respectively. Additionally, Figure 4.10 compares the estimated and true distribution functions of the two groups. Overall, estimates $\widehat{F}_T(t|A = 0)$ and $\widehat{F}_T(t|A = 1)$ shown in Figure 4.10 agree well with the true distribution functions, though because of sampling variation they naturally fall above or below the true functions.



(a)

(b)



(c)

Figure 4.1: Plot of the mean of raw estimates $\bar{Y}_\ell$ at s_ℓ, $\ell = 1, ..., m$, for $m = 10,\ 20,\ 50$, in 500 simulations, compared with the true $F_T(t|A)$ curves for two treatment groups in CASE I. (a) is for $m = 10$, i.e. $b_m = 0.25$, (b) is for $m = 20$, i.e. $b_m = 0.125$, (c) is for $m = 50$, i.e. $b_m = 0.05$. Sample size is 200.

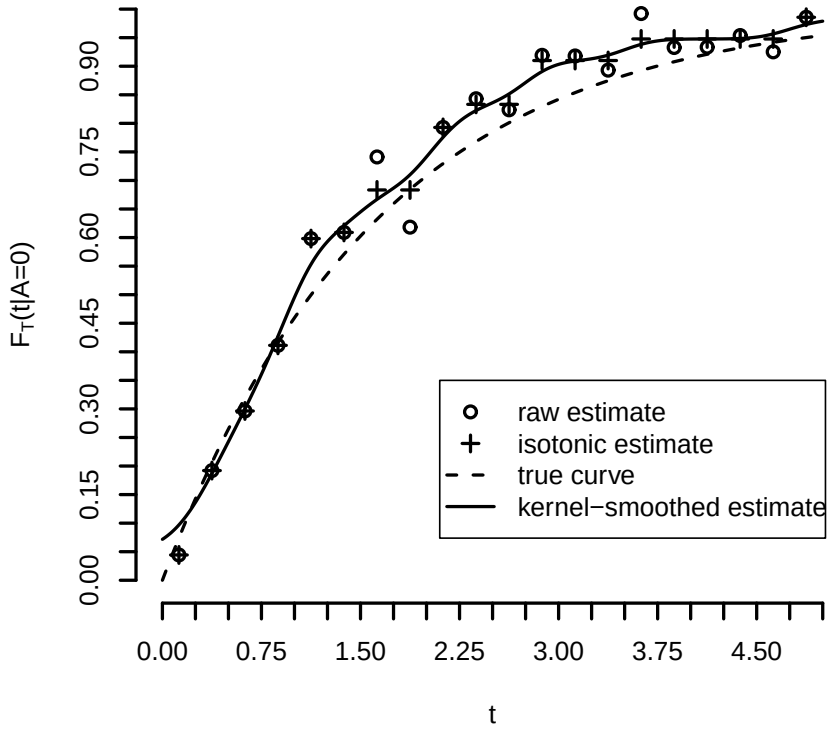


Figure 4.2: The plot of raw estimates, isotonic estimates, kernel-smoothed isotonic estimates and true $F_T(t|A = 0)$ of one simulation sample in CASE I. Sample size is about 100, $m = 20$, and bandwidth h_m in (4.6) is 0.2.

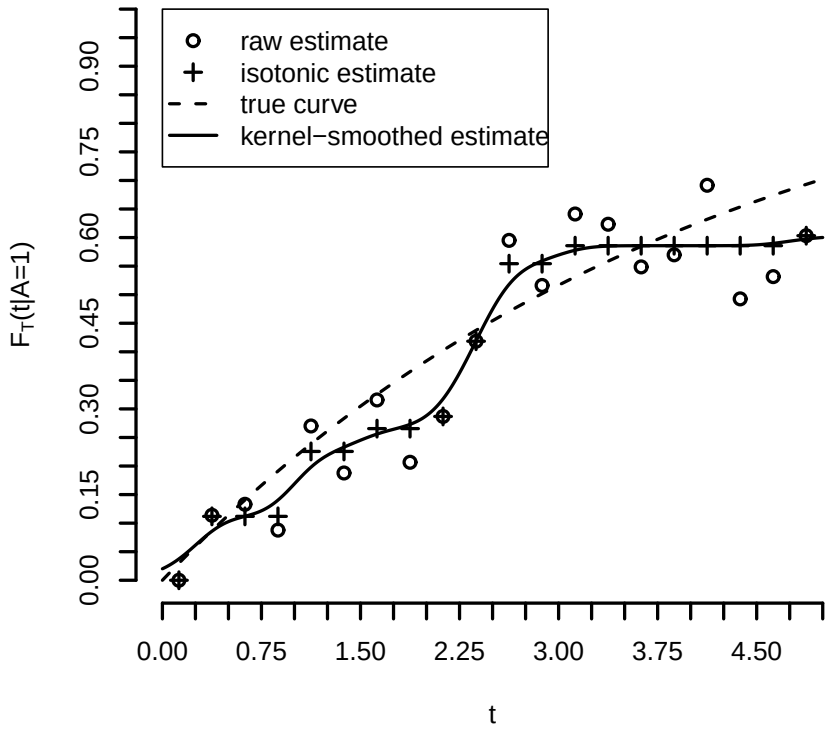


Figure 4.3: The plot of raw estimates, isotonic estimates, kernel-smoothed isotonic estimates and true $F_T(t|A = 1)$ of one simulation sample in CASE I. Sample size is about 100, $m = 20$, and bandwidth h_m in (4.6) is 0.2.

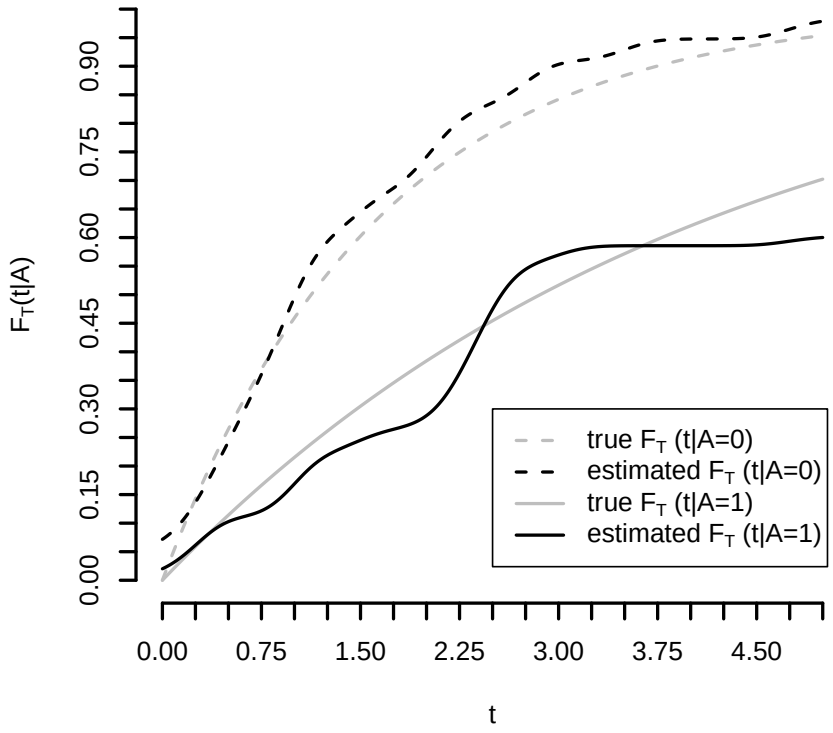


Figure 4.4: The kernel-smoothed isotonic estimate by (4.6) and true $F_T(t|A)$, $A = 1$ versus $A = 0$ of one simulation sample in CASE I. Sample size is about 100 for each treatment group, $m = 20$, and bandwidth h_m in (4.6) is 0.2.

Table 4.1: Bias, empirical standard error (ESE), mean of bootstrap estimated standard errors (BSE), and empirical coverage probability (ECP) of the kernel-smoothed isotonic estimate at time t by (4.6) for CASE I. Sample size is 200 (about 100 for each treatment group), number of simulation replicates is 500, with $m = 20$, i.e. $b_m = 0.125$, and $h_m = 0.2$. Note that ECPs are reported based on $ln(-ln(\widehat{F}_T(t|A)))$, while other quantities are reported based on $\widehat{F}_T(t|A)$.

	A=0					A=1				
t	True Value	BIAS	ESE	BSE	ECP	True Value	BIAS	ESE	BSE	ECP
0.5	0.265	-0.007	0.050	0.048	0.93	0.114	-0.002	0.041	0.036	0.90
1.0	0.459	-0.009	0.056	0.057	0.95	0.215	-0.002	0.052	0.049	0.92
1.5	0.603	-0.007	0.059	0.056	0.92	0.305	0.002	0.057	0.055	0.93
2.0	0.708	-0.002	0.053	0.051	0.93	0.384	0.004	0.058	0.059	0.95
2.5	0.785	-0.002	0.049	0.046	0.94	0.454	0.002	0.062	0.059	0.93
3.0	0.842	-0.003	0.042	0.041	0.93	0.517	0.002	0.063	0.059	0.93
3.5	0.884	-0.003	0.038	0.035	0.93	0.572	0.001	0.064	0.058	0.92
4.0	0.915	0.000	0.032	0.030	0.93	0.621	0.002	0.061	0.057	0.92
4.5	0.937	0.003	0.027	0.025	0.93	0.664	0.009	0.060	0.057	0.93
5.0	0.954	0.003	0.025	0.021	0.86	0.702	0.013	0.070	0.063	0.92

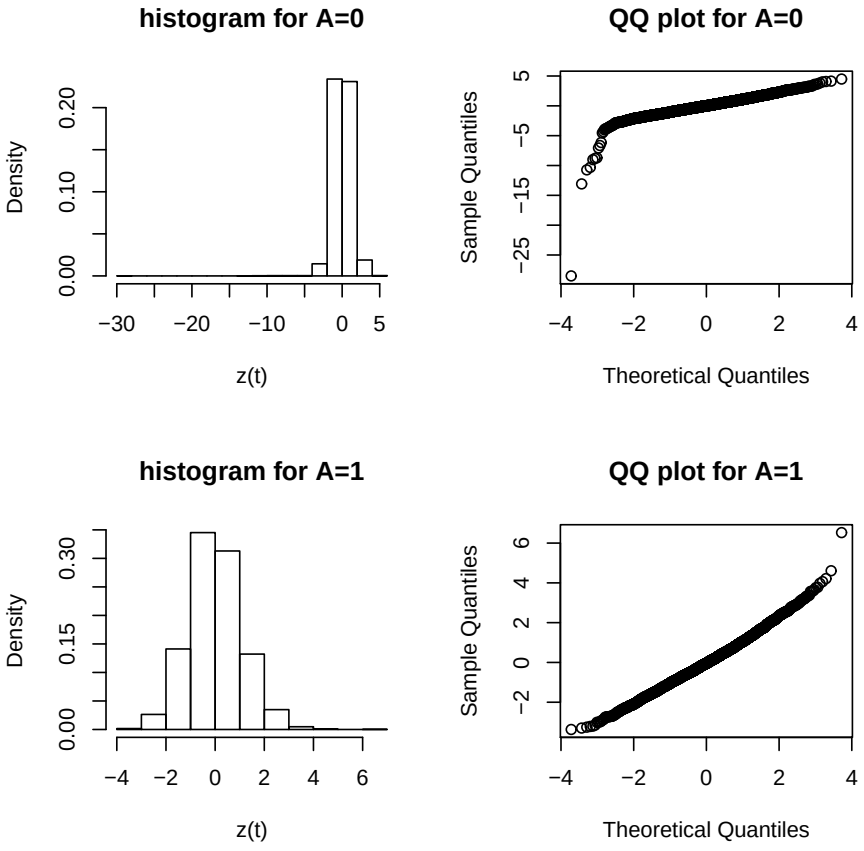


Figure 4.5: Histograms and QQ-plots of $z(t)$'s based on the transformed estimates, $ln(-ln(\widehat{F}_T(t|A)))$, corresponding to Table 4.1 for CASE I.

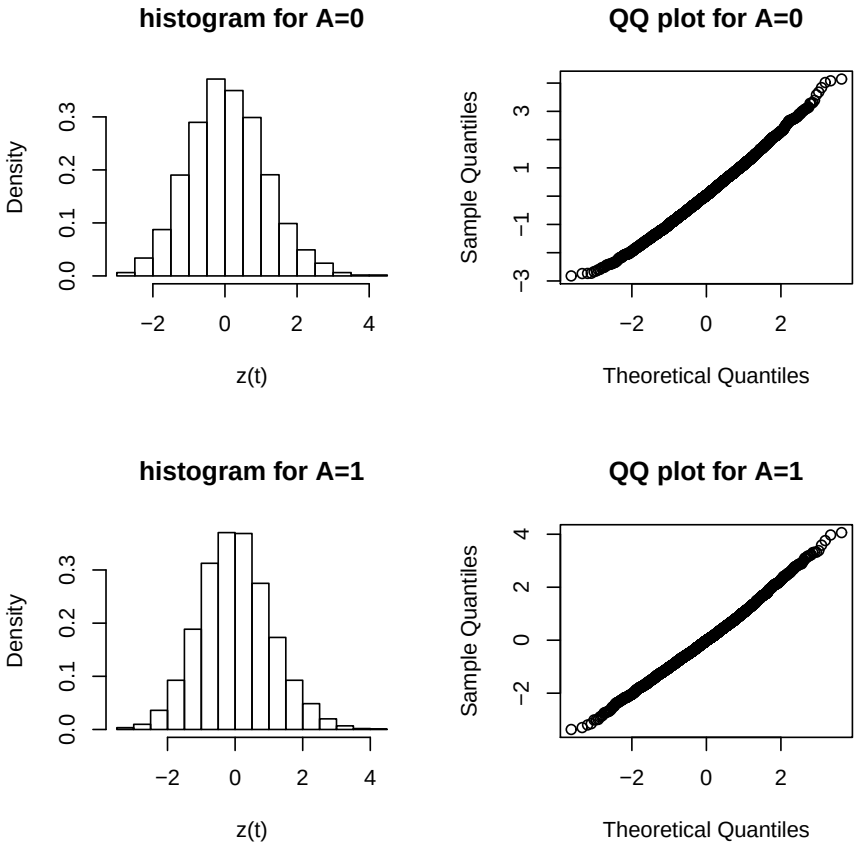


Figure 4.6: Histograms and QQ-plots of $z(t)$'s based on the transformed estimates, $ln(-ln(\widehat{F}_T(t|A)))$, where $t = 1.0 - 4.0$, corresponding to Table 4.1 for CASE I.

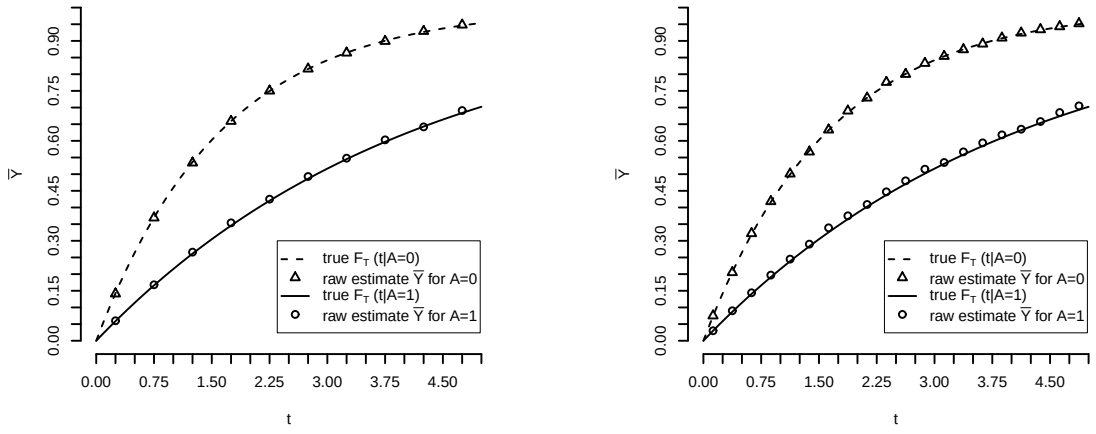
The pointwise kernel-smoothed estimates of $F_T(t|A)$ at $t = 0.5, 1.0, ..., 4.5, 5.0$ are summarized in Table 4.2, for the two groups separately. As in Table 4.1, true values, bias, standard errors are reported in the original form, while coverage probabilities are reported with a $ln(-ln)$ transformation. Figure 4.11 shows the histograms and QQ-plots of pooled $z(t)$ values and Figure 4.12 shows the truncated histograms and QQ-plots for $t = 1.0 - 4.0$. Similar to CASE I, it can be concluded that except when t is close to 0 or 5, estimates have negligible bias and empirical coverage is close to the 95% nominal level.

4.2 IIV Weighted Semi-Parametric Estimation Based on Additive Hazards Models

In the preceding section, we proposed an IIV weighted non-parametric estimator for $F_T(t)$ of dependently interval-censored failure times due to intermittent and outcome-dependent inspection times. In the present section, we will introduce the IIV weighted semi-parametric estimation under additive hazards models (Aalen, 1980, 1989; Lin and Ying, 1994), which have the form

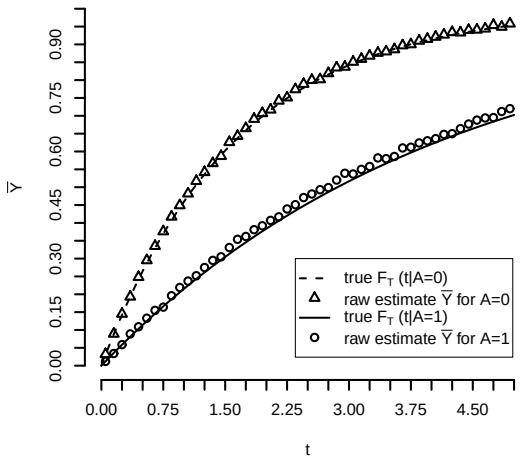
$$\lambda_T(t|\boldsymbol{A}_i(t^-)) = \lambda_{T,0}(t) + \boldsymbol{\beta}'\boldsymbol{A}_i(t^-), \qquad i = 1, .., n, \tag{4.12}$$

where $\lambda_{T,0}(t)$ is an unspecified baseline hazard function, and $\boldsymbol{A}_i(t^-)$ is a vector of time-varying covariates which are usually assumed to be external covariates (Kalbfleisch and Prentice, 2002). Semi-parametric analysis of model (4.12) with left-truncated and right-censored data was studied by Lin and Ying (1994), and Lin et al. (1998) proposed a semi-parametric estimation method for "case-I" interval-censored data (current status data), i.e. when there is only one inspection time per subject. The appealing idea of Lin et al. (1998) is to reduce the problem to ordinary Cox models so that the partial likelihood principle can



(a)

(b)



(c)

Figure 4.7: Plot of the mean of raw estimates $\bar{Y}_\ell$ at s_ℓ, $\ell = 1, ..., m$, for $m = 10,\ 20,\ 50$, in 500 simulations, compared with the true $F_T(t|A)$ curves for two treatment groups in CASE II. (a) is for $m = 10$, i.e. $b_m = 0.25$, (b) is for $m = 20$, i.e. $b_m = 0.125$, (c) is for $m = 50$, i.e. $b_m = 0.05$. Sample size is 200.

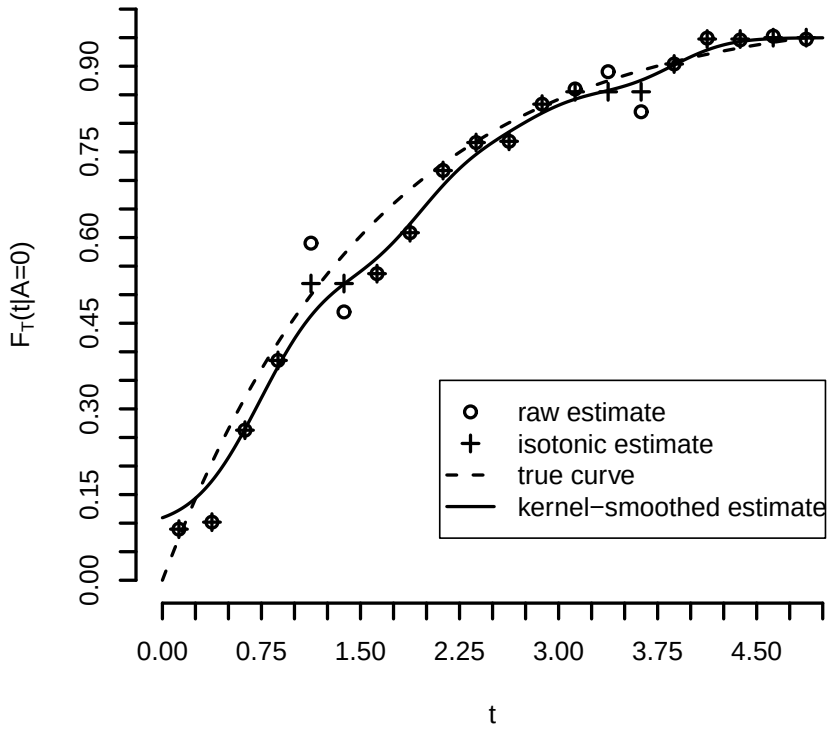


Figure 4.8: The plot of raw estimates, isotonic estimates, kernel-smoothed isotonic estimates and true $F_T(t|A = 0)$ of one simulation sample in CASE II. Sample size is about 100, $m = 20$, and bandwidth h_m in (4.6) is 0.3.

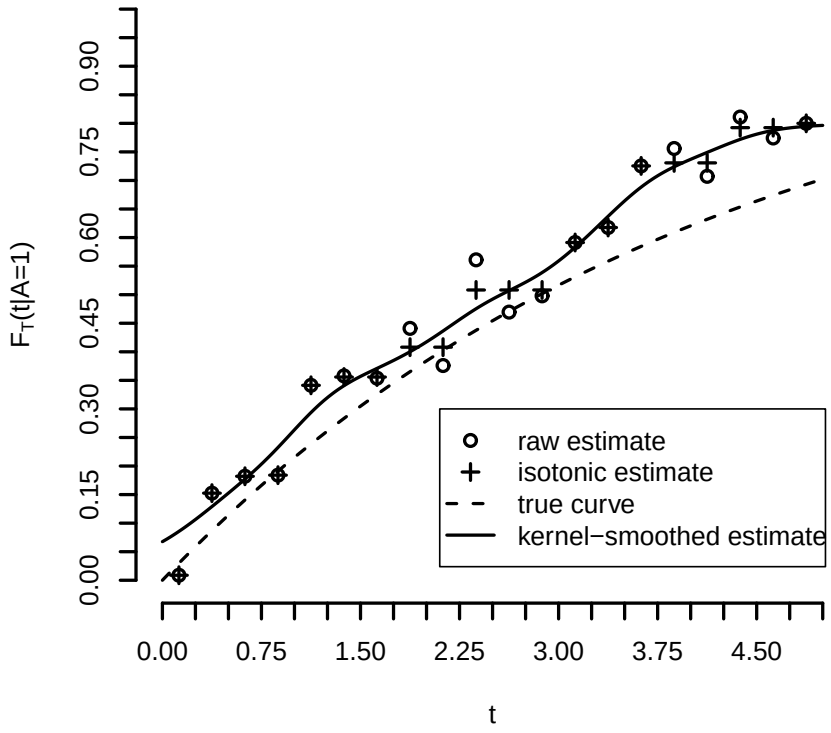


Figure 4.9: The plot of raw estimates, isotonic estimates, kernel-smoothed isotonic estimates and true $F_T(t|A = 1)$ of one simulation sample in CASE II. Sample size is about 100, $m = 20$, and bandwidth h_m in (4.6) is 0.3.

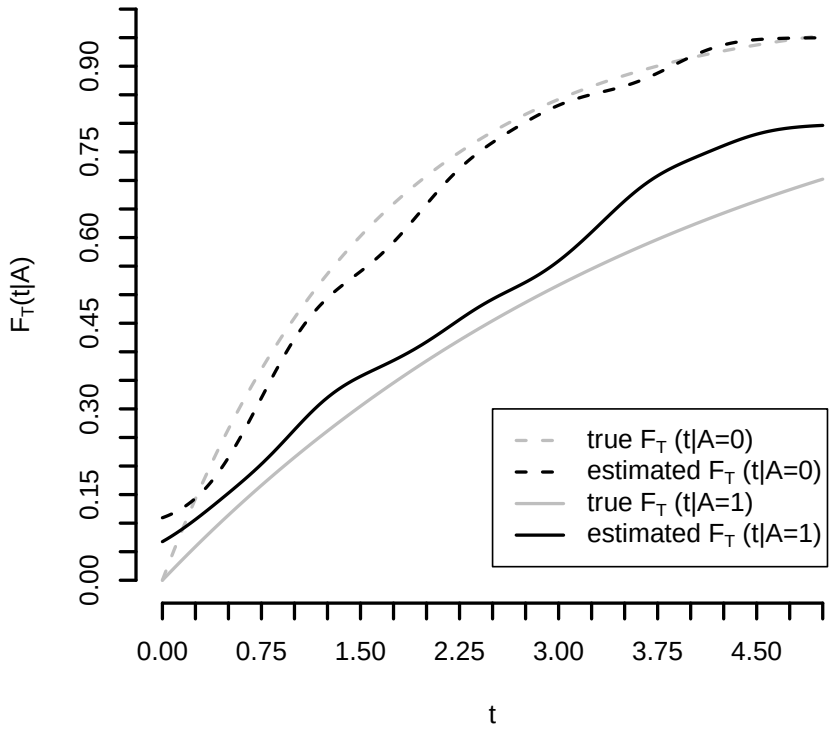


Figure 4.10: The kernel-smoothed isotonic estimate by (4.6) and true $F_T(t|A)$ for $A = 0$ versus $A = 1$ of one simulation sample in CASE II. Sample size is about 100 for each treatment group, $m = 20$, and bandwidth h_m in (4.6) is 0.3.

Table 4.2: Bias, empirical standard error (ESE), mean of bootstrap estimated standard errors (BSE), and empirical coverage probability (ECP) of the kernel-smoothed isotonic estimate at time t by (4.6) for CASE II. Sample size is 200 (about 100 for each group), number of simulation replicates is 500. Let $m = 20$, i.e. $b_m = 0.125$, and $h_m = 0.3$. Note that ECPs are reported based on $ln(-ln(\widehat{F}_T(t|A)))$, while other quantities are reported based on $\widehat{F}_T(t|A)$.

	A=0					A=1				
time	true value	BIAS	ESE	BSE	ECP	true value	BIAS	ESE	BSE	ECP
0.5	0.265	0.004	0.041	0.041	0.94	0.114	0.006	0.034	0.032	0.92
1.0	0.459	-0.010	0.049	0.048	0.94	0.215	0.002	0.043	0.042	0.95
1.5	0.603	-0.009	0.051	0.049	0.93	0.305	0.004	0.047	0.048	0.96
2.0	0.708	-0.005	0.048	0.046	0.93	0.384	0.003	0.051	0.051	0.96
2.5	0.785	-0.003	0.045	0.042	0.94	0.454	0.003	0.055	0.052	0.95
3.0	0.842	-0.003	0.040	0.038	0.93	0.517	0.003	0.055	0.053	0.94
3.5	0.884	-0.003	0.035	0.033	0.93	0.572	0.002	0.055	0.053	0.93
4.0	0.915	0.000	0.030	0.028	0.93	0.621	0.003	0.054	0.052	0.94
4.5	0.937	0.002	0.026	0.024	0.93	0.664	0.007	0.055	0.052	0.93
5.0	0.954	-0.001	0.025	0.021	0.88	0.702	0.003	0.062	0.056	0.91

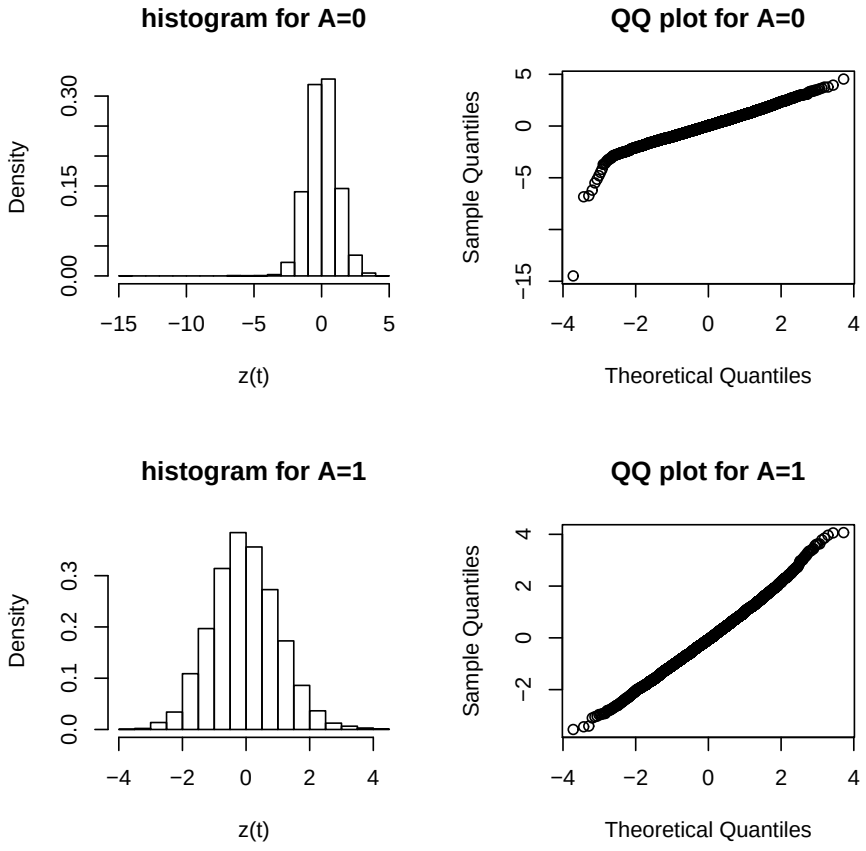


Figure 4.11: Histograms and QQ-plots of $z(t)$'s based on the transformed estimates, $ln(-ln(\widehat{F}_T(t|A)))$, corresponding to Table 4.2 for CASE II.

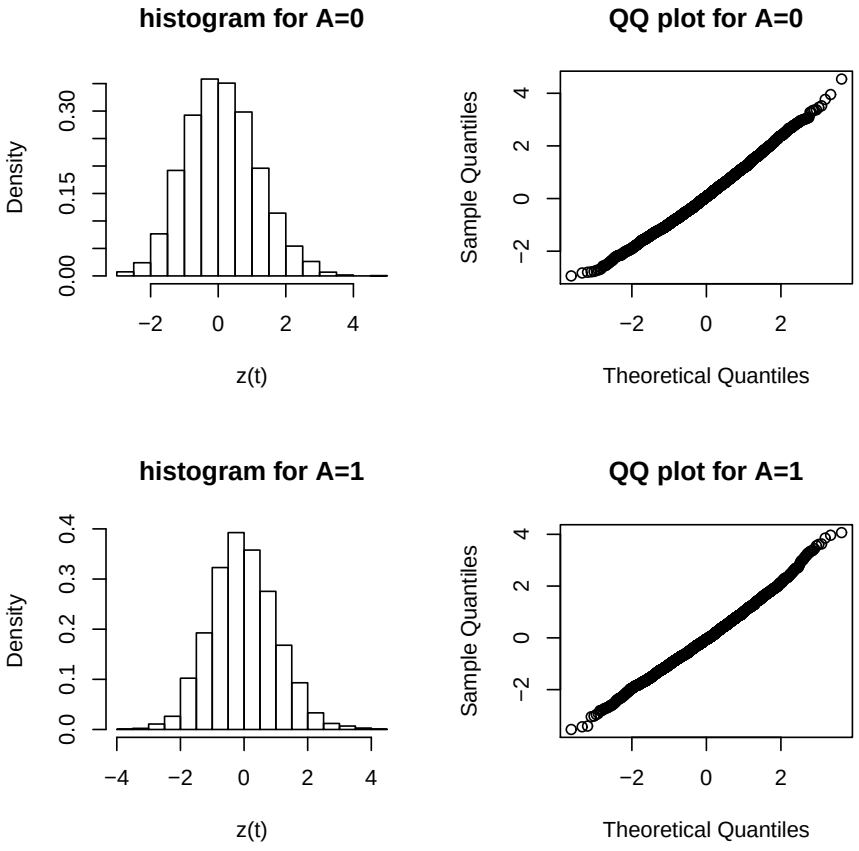


Figure 4.12: Histograms and QQ-plots of $z(t)$'s based on the transformed estimates, $ln(-ln(\widehat{F}_T(t|A)))$, where $t = 1.0 - 4.0$, corresponding to Table 4.2 for CASE II.

be applied, with restricting inspection times under proportional hazards models. Later, Ghosh (2001) and Martinussen and Scheike (2002) followed up their paper by discussing and improving efficiency. For general interval-censored data, i.e. "case-II", Zhao and Hsu (2005) studied the problem via empirical likelihood and Zeng et al. (2006) proposed a maximum likelihood estimation approach. However, most literature on semi-parametric analysis for case-II interval-censored data under model (4.12) assumes that inspection times or monitoring times, e.g. t_{il} and t_{ir} in (1.3), are independent of failure time T_i given $\boldsymbol{A}_i(t^-)$. Work on the setting where inspection times are related to failure times is limited. Recently, Wang et al. (2010) accommodated informative interval censoring by introducing an unobservable random process to characterize the dependency between inspection times and failure times, and Zhao et al. (2015) utilized copula models to model the correlation. In this section, we will generalize the IIV weighted estimating function approach to the case of semi-parametric estimation based on additive hazards models so that dependently case-II interval-censored data can be consistently analysed. The procedure that will be introduced below can be applied to proportional hazards models as well, and we will discuss that at the end of this section.

4.2.1 An Iterative Two-Stage IIV Weighted Semi-Parametric Estimation Procedure

Given model (4.12) and the fact that $S_T(t) = \exp\{-\Lambda_T(t)\}$ for continuous failure times T_i where $\Lambda_T(t) = \int_0^t \lambda_T(s)ds$ is the cumulative hazard function, the survival function is

$$S_T(t) = S_{T,0}(t)\, \exp\left\{ -\int_0^t \boldsymbol{\beta}' \boldsymbol{A_i}(s)ds \right\}, \tag{4.13}$$

where $S_{T,0}(t) = \exp(-\Lambda_{T,0}(t))$ is the baseline survival function. We now proceed in two stages.

First, we estimate the baseline survival function $S_{T,0}(t)$ for a given value of the regression parameter $\boldsymbol{\beta}$. Similar to Section 4.1.1, we define a set of finite time points where $S_{T,0}(t)$ is estimated, i.e. $0 < s_1 <, ..., < s_m \leq \tau$ and let $\theta_\ell = \theta_\ell(\boldsymbol{\beta})$ be $S_{T,0}(t;\boldsymbol{\beta})$ at s_ℓ, $\ell = 1, ..., m$. Then, given $\boldsymbol{\beta} = \boldsymbol{\beta_0}$, an IIV weighted estimating function for $\boldsymbol{\theta} = (\theta_1, ..., \theta_m)'$ can be defined by

$$U_{1\ell}(\boldsymbol{\theta}(\boldsymbol{\beta_0})) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} K_b(s_\ell - t_{ij}) w_i(t_{ij}) \exp\left\{-\int_0^{t_{ij}} \boldsymbol{\beta_0'} \boldsymbol{A_i}(s) ds\right\} \times \left\{P_i(t_{ij}) - \theta_\ell \exp\left[-\int_0^{t_{ij}} \boldsymbol{\beta_0'} \boldsymbol{A_i}(s) ds\right]\right\}, \tag{4.14}$$

for $\ell = 1, ..., m$, where $K_b(u) = K(u/b_m)/b_m$; $K(\cdot)$ is a bounded kernel function; $0 < b_m \downarrow 0$ is a bandwidth sequence; $P_i(t) = I(T_i > t)$ is the response at t; $t_{i1}, ..., t_{im_i}$ denote the actual visits in CASE I (failure does not terminate visits) but include the actual visits and pseudo visits in CASE II (failure terminates visits), and in general we define $\{t_{i1}, ..., t_{im_i}\} = \{t > 0 : dN_i^*(t) = 1\}$. Following the convention defined before, for CASE I we have $\{dN_i(t) : t > 0\}$ as a continuous visit process; in CASE II, we consider the visit process as a discrete time process, i.e. $\{dN_i(a_k) : k = 1, ..., M\}$. After the visit following failure occurrence in CASE II, i.e. when $\bar{P}_i^{obs}(t^-) = 0$, we set the IIV weight $w_i(t)$ to be 1, provided that assumption (B1*) is true. Only the actual visits in either CASE I or CASE II are used for estimating the IIV weights $w_i(t) = 1/\lambda_N(t|\boldsymbol{Z_i}(t^-))$, where $\lambda_N(t)$ is the visit intensity at t and $\boldsymbol{Z_i}(t^-)$ includes features in the observed history prior to t. We have proposed earlier to estimate $w_i(t)$ by fitting a semi-Markov proportional hazards model for visit gap times, like (2.19). Solving $U_{1\ell}(\boldsymbol{\theta}) = 0$ gives a closed form estimate of θ_ℓ, i.e. $S_{T,0}(s_\ell)$, as

$$\widehat{\theta}_\ell = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_b(s_\ell - t_{ij}) w_i(t_{ij}) \exp\left\{-\int_0^{t_{ij}} \boldsymbol{\beta_0'} \boldsymbol{A_i}(s) ds\right\} P_i(t_{ij})}{\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_b(s_\ell - t_{ij}) w_i(t_{ij}) \exp\left\{-2\int_0^{t_{ij}} \boldsymbol{\beta_0'} \boldsymbol{A_i}(s) ds\right\}}. \tag{4.15}$$

Note that only visits that happened within window $[s_\ell - b_m, s_\ell + b_m]$ contribute to the estimation of θ_ℓ and the selection of bandwidth b_m depends on the number of visits that

fall in $[s_\ell - b_m, s_\ell + b_m]$, as well as concern for the variability and smoothness of final estimate of $S_{T,0}(t)$.

If we define $w_+(s_\ell) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} K_b(s_\ell - t_{ij})w_i(t_{ij})\ \exp\left\{-2\int_0^{t_{ij}} \boldsymbol{\beta_0'A_i}(s)ds\right\}$, then (4.15) can be rewritten as

$$\widehat{\theta}_\ell = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_b(s_\ell - t_{ij})w_i(t_{ij})\ \exp\left\{-\int_0^{t_{ij}} \boldsymbol{\beta_0'A_i}(s)ds\right\} P_i(t_{ij})}{w_+(s_\ell)}.$$

To monotonize the raw estimates, $\widehat{\theta}_\ell$, based on the constraint $\theta_1 \geq \theta_2 \geq ... \geq \theta_m$, we can adopt an antitonic (monotonically non-increasing) regression of $(\widehat{\theta}_1, ..., \widehat{\theta}_m)$ with weights $(w_+(s_1), ..., w_+(s_m))$ to produce non-increasing estimates of $S_{T,0}(s_\ell)$'s which are denoted by $\tilde{\theta}_\ell$, $\ell = 1, ..., m$. Following that, a kernel-smoothed estimate of baseline survival function $S_{T,0}(t)$ is defined by

$$\widehat{S}_{T,0}(t; \boldsymbol{\beta_0}) = \frac{\sum_{\ell=1}^{m} \tilde{\theta}_\ell K^*\left(\frac{s_\ell - t}{h_m}\right)}{\sum_{\ell=1}^{m} K^*\left(\frac{s_\ell - t}{h_m}\right)}, \qquad \text{for all } t > 0, \tag{4.16}$$

where $K^*(\cdot)$ is a log-concave density and $\{h_m\}$ is a positive bandwidth sequence.

Second, we construct an IIV weighted profile GEE with the estimated baseline survival $\widehat{S}_{T,0}(t; \boldsymbol{\beta_0})$ plugged in to estimate the regression parameter $\boldsymbol{\beta}$, defined by

$$\boldsymbol{U_2}(\boldsymbol{\beta}|\boldsymbol{\beta_0}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{w_i(t_{ij})\frac{\partial e(\bar{\boldsymbol{A}}_{\boldsymbol{i}}(t_{ij});\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\left[P_i(t_{ij}) - \widehat{S}_{T,0}(t_{ij}; \boldsymbol{\beta_0})\ e(\bar{\boldsymbol{A}}_{\boldsymbol{i}}(t_{ij}); \boldsymbol{\beta})\right]}{e(\bar{\boldsymbol{A}}_{\boldsymbol{i}}(t_{ij}); \boldsymbol{\beta})\left[1 - \widehat{S}_{T,0}(t_{ij}; \boldsymbol{\beta_0})\ e(\bar{\boldsymbol{A}}_{\boldsymbol{i}}(t_{ij}); \boldsymbol{\beta})\right]}, \tag{4.17}$$

where $e(\bar{\boldsymbol{A}}_{\boldsymbol{i}}(t_{ij}); \boldsymbol{\beta}) = \exp\left\{-\int_0^{t_{ij}} \boldsymbol{\beta' A_i}(s)ds\right\}$ and $\widehat{S}_{T,0}(t_{ij}; \boldsymbol{\beta_0})$ is obtained from (4.16) given $t = t_{ij}$. Solving $\boldsymbol{U_2}(\boldsymbol{\beta}|\boldsymbol{\beta_0}) = \boldsymbol{0}$ gives an estimate of $\boldsymbol{\beta}$ with $\boldsymbol{\beta_0}$ specified as an initial value.

The iterative algorithm for semi-parametric estimation of regression parameters $\boldsymbol{\beta}$ and baseline survival function $S_{T,0}(t)$ in model (4.13) can be summarized by the following steps, starting with $l = 0$:

- Step 1: specify an initial value for $\boldsymbol{\beta}$, denoted by $\widehat{\boldsymbol{\beta}}_{\boldsymbol{l}}$, which can be given by a naive estimate based on model (4.12) with informative censoring not adjusted for.

- Step 2: given $\widehat{\boldsymbol{\beta}}_{\boldsymbol{l}}$, solve the estimating equation $U_{1\ell}(\boldsymbol{\theta}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{l}})) = 0$ from (4.14) and then monotonize and smooth the raw estimates, $\widehat{\theta}_\ell$, $\ell = 1, ..., m$, to obtain an estimate of the baseline survival function $\widehat{S}_{T,0}(t; \widehat{\boldsymbol{\beta}}_{\boldsymbol{l}})$.

- Step 3: substitute $\widehat{S}_{T,0}(t; \widehat{\boldsymbol{\beta}}_{\boldsymbol{l}})$ in the estimating function $\boldsymbol{U_2}(\boldsymbol{\beta}|\widehat{\boldsymbol{\beta}}_{\boldsymbol{l}})$ given in (4.17) to obtain a new estimate of $\boldsymbol{\beta}$, denoted by $\widehat{\boldsymbol{\beta}}_{\boldsymbol{l+1}}$.

- Step 4: use the new estimate of $\boldsymbol{\beta}$ as the initial value in Step 1 and repeat Step 2 and Step 3 until the estimates converge, i.e. $|\widehat{\boldsymbol{\beta}}_{\boldsymbol{l+1}} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{l}}| \doteq 0$.

As we obtain the estimate of $\boldsymbol{\beta}$, then the baseline survival function $S_{T,0}(t)$ can be estimated by repeating the Step 2 with $\widehat{\boldsymbol{\beta}}$ plugged in.

4.2.2 Simulation Study

In this subsection, a simulation will be conducted to study the finite sample performance of the proposed semi-parametric estimator of $\boldsymbol{\beta}$ based on an additive hazards model (4.12). The Bootstrap will be utilized to estimate the standard errors of resulting estimates.

As mentioned before, Martinussen and Vansteelandt (2013) have shown the collapsibility of additive hazards models, and the model we use to generate failure times will be based on that. Explicitly, the treatment variable of interest, A_i, is assumed to have a $BIN(1, p)$ distribution with $p = 0.5$, and an ancillary variable, L_i, is generated from $L_i|A_i \sim N(0, 1)$, so L_i and A_i are assumed to be independent. Failure time T_i is generated from an Exponential distribution with a hazard rate given by $\lambda_T(t|L_i, A_i) = \beta_0 + \beta_1 L_i + \beta_2 A_i$, where

$\beta_0 = 1$, $\beta_1 = 0.2$, and $\beta_2 = -0.4$, so that larger L_i is associated with higher risk of failure and treatment lowers the risk. Following the convention of simulations in this thesis, we define an administrative end of follow-up time as $\tau = 5$ and discretize time with a grid of 100 per unit, i.e. $0 = a_0 < a_1 <, ..., < a_M = \tau$ with a time increment of 0.01 and $M = 500$. As a result, almost all untreated subjects have failed by τ and about $85\% - 90\%$ of treated subjects have failed by τ. As shown in Martinussen and Vansteelandt (2013), the model of T_i conditional on A_i alone is still an additive hazards model with hazard function of the form

$$\lambda_T(t|A_i) = \lambda_{T,0}(t) + \beta_2 A_i, \tag{4.18}$$

where $\lambda_{T,0}(t) = \beta_0 + \beta_1 \frac{E(e^{-\beta_1 t L_i} L_i)}{E(e^{-\beta_1 t L_i})}$.

Visit times are generated similarly to Section 4.1.2. That is, for the case where failure does not discontinue the visit process (CASE I), visit indicators dN_{k+1}'s are generated based on model (4.10) with L_k replaced by L and with the subscript for subjects suppressed. For the case where failure occurrence terminates visits (CASE II), dN_{k+1}'s are generated based on model (4.11) with L_k replaced by L and $\bar{Y}_k^{obs} = x$ replaced by $\bar{P}_k^{obs} = 1 - x$, where $x = 0$ or 1, since we defined $Y(t) = I(T \leq t) = 1 - P(t)$. Corresponding parameters in model (4.10) and model (4.11) are given by $\gamma_0 = -4$, $\gamma_1 = 0.5$, and $\gamma_2 = -0.2$, and as a result, the average visit gap time is about 0.41 for the group with $A = 0$ and is about 0.50 for the group with $A = 1$ in CASE I, and is about 0.46 when $A = 0$ and about 0.59 when $A = 1$ in CASE II. The IIV weights $w_i(t)$ are estimated by fitting the semi-Markov proportional hazards visit gap time model (2.19) with $\boldsymbol{Z_i}(t^-) = (L_i, A_i)'$ for CASE I and $\boldsymbol{Z_i}(t^-) = (L_i, A_i, \bar{P}_i^{obs}(t^-))'$ for CASE II, which is implemented by R function *phreg*. Cut-points for the piecewise-constant baseline hazard are set as $(0.42, 0.69, 0.92)$ for CASE I and CASE II, by the principle of comparison with the non-parametric estimate of baseline hazard produced by *coxreg*. Remember that in CASE II, when $\bar{P}_i^{obs}(t^-) = 0$ we fix $w_i(t)$

as 1.

The naive semi-parametric estimate of β_2 is provided by the R function *aalen* in the package *timereg*. If it is known that $T_i \in (t_{il}, t_{ir}]$, we consider $T_i = (t_{ir} + t_{il})/2$, i.e. we use the mid-point to approximate an interval-censored failure time, so failure times are completely observed or right-censored. Other than the mid-point approximation, informative inspection times are not adjusted for in the naive estimation, so naive estimates are expected to have some bias. They are used as initial values for β_2 in the proposed iterative two-stage semi-parametric estimation and will be compared with the estimates yielded by our proposed approach. The number of time points where raw estimates of $S_{T,0}(t)$ are computed, i.e. m, is set to be 26, so in estimating function (4.14) $b_m = 0.1$ so that for example in CASE I, there are about 69 and 47 actual visits included in $[s_\ell - b_m, s_\ell + b_m]$ for $A = 0$ and $A = 1$, respectively. In addition, we let h_m in the non-parametric kernel regression (4.16) be 0.1. The antitonic regression for monotonizing the raw estimates of baseline survival probabilities is implemented by *monoreg*, kernel smoothing is still implemented by *npreg*, and the estimating function (4.17) is solved by *nleqslv*. Kernel function $K(\cdot)$ in (4.15) is selected to be the EP kernel and $K^*(\cdot)$ in (4.16) is the Gaussian (standard normal) kernel. Sample size is 300 and 200 simulations are conducted for each case. Standard errors are estimated for a given sample by a non-parametric bootstrap with 100 replicates. Estimates from the proposed approach, compared with the naive estimates, are summarized in Table 4.3. A convergence is declared for the IIV weighted semi-parametric estimator of β_2 in model (4.18), when $|\widehat{\beta}_{2,l+1} - \widehat{\beta}_{2,l}| < 10^{-3}$ where $\widehat{\beta}_{2,l}$ denotes the estimate of β_2 from the lth iteration. Empirical coverage probabilities (ECPs) are computed based on the 95% empirical confidence intervals constructed by $\widehat{\beta}_2 \pm 1.96\ se(\widehat{\beta}_2)$, where se is the bootstrap estimated standard error. From Table 4.3, it can be seen that for either CASE, the bias of the IIV weighted estimates is negligible and empirical coverage probabilities are

close to the nominal level, even though sample size is only 300. On the other hand, the coverage probabilities of naive estimates are much lower because of large bias.

Table 4.3: Bias, mean of bootstrap estimated standard errors (BSE), empirical standard error (ESE), mean squared error (MSE) and empirical coverage probability (ECP) of the proposed estimates and naive estimates for β_2 in model (4.18) for two cases: CASE I and CASE II. True value of β_2 is -0.4. Sample size is 300. Number of simulation replicates is 200. Bootstrap sample size is 100.

	CASE I				
	BIAS	BSE	ESE	MSE	ECP
naive	0.096	0.087	0.088	0.017	0.79
IIV	0.005	0.123	0.117	0.014	0.94
	CASE II				
	BIAS	BSE	ESE	MSE	ECP
naive	0.095	0.087	0.088	0.017	0.79
IIV	0.001	0.122	0.120	0.014	0.94

4.2.3 Discussion on the Extension to the Cox Proportional Hazards Models

For the Cox proportional hazards models, which have a form of

$$\Lambda_T(t|\boldsymbol{A_i}) = \Lambda_{T,0}(t)\ \exp(\boldsymbol{\beta'A_i}), \tag{4.19}$$

or written as

$$S_T(t|\boldsymbol{A_i}) = S_{T,0}(t)^{\exp(\boldsymbol{\beta'A_i})}. \tag{4.20}$$

Huang (1996) proposed a two-step maximum profile likelihood function procedure to estimate $(\Lambda_{T,0}, \boldsymbol{\beta})$ based on model (4.19) for independently case-I interval-censored data, i.e. current status data. Our proposed iterative two-stage weighted profile GEE approach also can be applied to the Cox models.

Explicitly, the mean of $P_i(t_{ij})$ in (4.14) and (4.17) should be changed to be $S_{T,0}(t_{ij})^{\exp(\boldsymbol{\beta}'\boldsymbol{A_i})}$. With this form, equation solving might be more difficult. For example, for a fixed $\boldsymbol{\beta_0}$, the estimating function (4.14) for baseline survival function $\theta_\ell = S_{T,0}(s_\ell)$ can be modified here as

$$U^*_{1\ell}(\boldsymbol{\theta}(\boldsymbol{\beta_0})) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{K_b(s_\ell - t_{ij})w_i(t_{ij})\exp(\boldsymbol{\beta_0'A_i})\left\{P_i(t_{ij}) - \theta_\ell^{\exp(\boldsymbol{\beta_0'A_i})}\right\}}{\theta_\ell\left[1-\theta_\ell^{\exp(\boldsymbol{\beta_0'A_i})}\right]}, \quad \ell = 1, ..., m. \tag{4.21}$$

Unlike (4.14), which leads to a closed form estimate of θ_ℓ, $\widehat{\theta}_\ell$ here has to be solved numerically from $U^*_{1\ell}(\boldsymbol{\theta}) = 0$. A similar change is made to the second estimating function (4.17) to accommodate a proportional hazards model. Alternatively, piecewise-constant baseline hazard can be assumed to make the additive hazards model (4.12) or the proportional hazards model (4.19) be "flexibly" parametric, e.g. the piecewise-constant baseline proportional hazards model (2.5), and then the methodologies introduced in Chapter 2 can be applied.

So far, we have introduced parametric, non-parametric and semi-parametric estimation of failure time distributions based on estimating functions weighted by a so-called inverse-intensity-of-visit (IIV) weight to adjust for the intermittent and informative inspection times which cause failure times dependently interval-censored. In addition, we also considered confounded treatments in observational studies in Chapter 3 and introduced an inverse-probability-of-treatment (IPT) weighting method which is helpful to adjust for measured confounders. In the next chapter, we will apply these methodologies to the PsA

Toronto cohort to study treatment with biologics and joint damage as an illustration.

Chapter 5

APPLICATIONS TO PSA COHORT

In Chapter 2 and Chapter 4, we introduced an inverse-intensity-of-visit (IIV) weighted estimating function approach to estimate marginal association measure(s) for interval-censored failure time outcomes. When the visit process is intermittent and informative, interval censoring may not be independent of failure times. From the simulation studies shown in the preceding chapters, we have seen that the proposed IIV weighting method adjusts for outcome-dependent inspection times and eliminates the resulting selection bias. In this chapter, we will apply this method to the PsA Toronto Cohort Study as an illustration. We saw in Section 1.4.3 that visit times for individuals in this cohort were related to prior disease status and other factors.

Biologics are presently widely used to reduce signs and symptoms of active arthritis and to slow the progression of joint destruction in patients with PsA, especially in patients who have had an inadequate response to one or more DMARDs such as Methotrexate (MTX). One interesting question is to evaluate the effect of biologics on inhibiting the progression of joint damage. That is, assess the association between treatment with biologics and time

to an increase in the number of damaged joints after enrolment. Biologics are recently used for treating rheumatoid arthritis, and a variety of biological agents have been developed and licensed during the past few decades. It is noted that patients could be treated with distinct biological agents at different chronological times. Therefore, our analysis set is selected by a 1:1 matching of patients in terms of their status of using biologics by calendar day. Although it is more reasonable to look at one particular drug to assess a treatment effect, a limited number of one specific drug users makes it hardly feasible. Thus, we hope that the matching could roughly adjust for the drug variety into account in the analysis. Explicitly, we first take a patient who began an initial course of biologics treatment on some calendar day after clinic enrolment and then randomly choose another patient who had never taken biologics as of the same calendar day. As a result, 414 (207 treated and 207 untreated) patients are included in this analysis set. The calendar years when the 207 treated patients received biologics and the 207 untreated patients were randomly selected for matching range from 1981 to 2012 with a median of 2006 (1st quartile = 2002, 3rd quartile = 2009). A histogram of the frequency of calendar years when treated and untreated individuals were matched is displayed in Figure 5.1. Most patients (approximately 85%) were matched between 2000 and 2010. In the present chapter, we define t_{i0} as the time when subject i was matched. Let $A_i = 1$, or 0, be the indicator of initial biologics status at t_{i0}. Patients' treatment status may change later, e.g. biologics may be terminated, paused or switched for inefficacy, adverse effects or other reasons. Here, we consider two studies: one is to investigate the treatment intention with biologics and the other is to evaluate the treatment effect of biologics. For the former, since treatment intention is of interest, patients' treatment status is allowed to vary after matching. For the latter, the change of treatment status is considered as a competing risk of the joint damage event, and estimation of cumulative incidence functions (CIFs) will be the goal.

We define a failure time outcome, T_i, as the time to an increase in the number of damaged joints from t_{i0}. Patients are followed up to the last visit prior to the administrative end of follow-up which is Nov. 2013, and there are no random drop-outs in this case. However, we consider patients who do not visit for longer than 3 years as lost to follow-up, since the protocol suggests that patients should be assessed about every six to 12 months. That is, we artificially censor visit gap times longer than 3 years. In addition, since joint damage is inspected only at clinic visits and visits are intermittent, failure times T_i are subject to interval censoring, or right censoring at the last visit. In this study, data is recorded in days, and the occurrence of failure does not prohibit visits, so this is an example of CASE I defined in preceding simulations. Since visit gap times are relatively shorter than joint damage event times T_i, in this chapter, analyses of visit gap times are reported in days, while analyses of T_i are reported in years. Additionally, the kernel function $K(\cdot)$ in (4.3) and (4.15) is selected to be the Epanechnikov (EP) kernel, i.e. $K(u) = \frac{3}{4}(1-u^2)I(|u| \leq 1)$, and the $K^*(\cdot)$ in (4.6) and (4.16) is selected to be the Gaussian kernel, i.e. $K^*(u) = \frac{1}{\sqrt{2\pi}}e^{-u^2/2}$ all through this chapter. The convergence tolerance of the regression parameter estimates in the iterative two-stage semi-parametric estimation procedure is set as $< 10^{-6}$.

5.1 Association between Joint Damage and Intended Biologics Treatment

In this study, A represents intended biologics treatment, and we will study its association with the time T to an increase in the number of damaged joints. Among the 207 treated patients, 105 have interval-censored failure times and 102 have right-censored failure times. For the 207 untreated patients, 84 failure times are interval-censored and 123 failure times

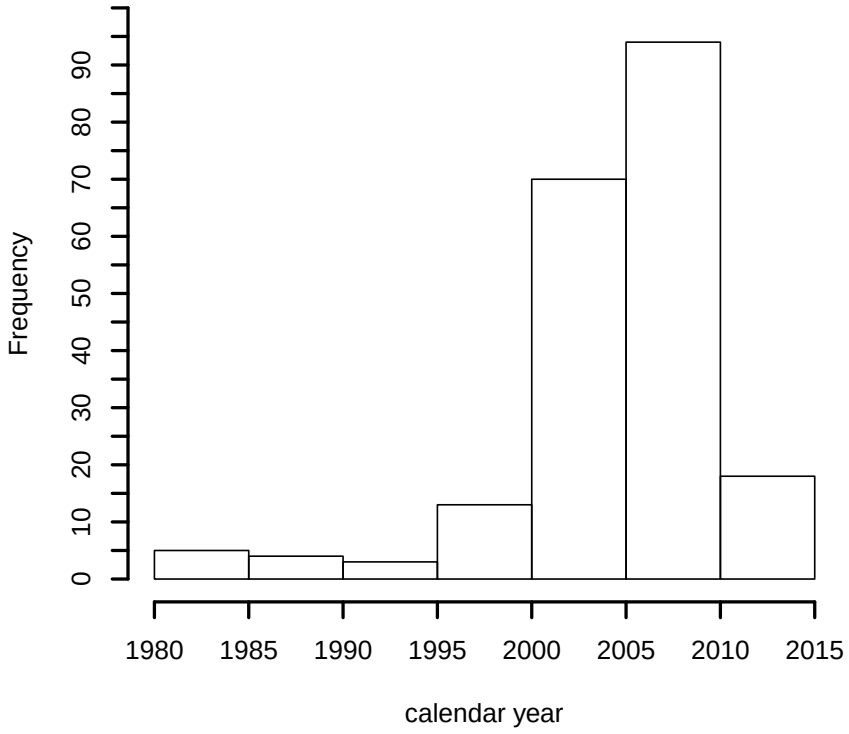


Figure 5.1: Histogram of frequency of the calendar year of matching time, t_{i0}, for the 207 treated patients.

are right-censored. On average, treated patients have 9 visits and untreated patients have 7 visits. The gap times between consecutive visits for the treated group have an average of 7.2 months, a median of 6.3 months and a standard error of 3.3 months. For the untreated group, the average is 7.4 months, the median is 6.3 months, while the standard error is 4.0 months. Although visits are planned to be every 6 – 12 months in this cohort, and the median of visit gap times is about 6 months in each treatment group, considerable variability of the frequency and timing of visits is seen. It was mentioned earlier that visit gap times longer than 3 years are artificially censored so as to avoid extremely large values of IIV weights. As a result, 38 individuals have visit gap times artificially censored, and 115 visits are deleted in total.

5.1.1 Analysis of Visit Times and Estimation of IIV Weights

To investigate the visit process and to estimate IIV weights, a semi-Markov proportional hazards model with a piecewise constant baseline, like (2.25), is used to analyse visit gap times. Models of visit times are fitted in days. The range of calendar dates of visits in this analysis set is from 1980-01-11 to 2013-03-25. To satisfy the proportional hazards assumption, visit gap times are analysed within subgroups which are defined by the calendar decade of the previous visit. They are $[1980, 2000)$, $[2000, 2010)$ and $[2010, 2013]$. For the lth subgroup, where $l = 1$, 2, or 3, the model for the intensity of visit gap times has the form

$$\lambda_N(t|\boldsymbol{Z_i}(t^-); \boldsymbol{\theta_l}) = \lambda^{\dagger}_{N0}(B(t); \boldsymbol{\pi_l}) \exp(\boldsymbol{\alpha}_{\boldsymbol{l}}' \boldsymbol{Z_i}(t^-)), \tag{5.1}$$

where $\lambda^{\dagger}_{N0}[B(t); \boldsymbol{\pi_l}] = \sum_{k=1}^{K} \pi_{l,k} I_{l,k}(t)$, $\boldsymbol{\pi_l} = (\pi_{l,1}, ..., \pi_{l,K})'$, $I_{l,k}(t) = I\{B(t) \in (a_{l,k-1}, a_{l,k}]\}$, $B(t)$ is the elapsed time or gap time from the previous visit and I is an indicator function which equals 1 if true and 0 otherwise; $\pi_{l,k}$, $k = 1, ..., K$, are the piecewise constant baseline

hazards for the lth subgroup, and $a_{l,k}$, $k = 1, ..., K$, are the corresponding cut-points which are selected by the comparison with a non-parametric estimate of baseline hazard; and $\boldsymbol{\theta_l} = (\boldsymbol{\pi_l'}, \boldsymbol{\alpha_l'})'$. Covariates $\boldsymbol{Z_i}(t^-)$ may include the interaction of functions of the visit gap time $B(t)$ and some regressors which have time-varying coefficients, e.g. ESR measures and the median length of past visit gap times. Proportional hazards assumptions have been tested by R function *cox.zph* for a semi-parametric version of model (5.1), which are accepted for all subgroups. Thus, R function *phreg* is used for fitting the piecewise constant hazard semi-Markov model (5.1) with the baseline hazard estimated parametrically. All regression summaries are given in Tables 5.1-5.3. The attributes of all regressors used in Tables 5.1-5.3 are listed in Table A.1 in Appendix A.1 at the end of the thesis.

From Table 5.1 for visit times before 2000, we see that there is evidence that visit intensities are significantly associated with ESR, sex, age, joint damage, at significance level of $\alpha = 0.05$. From Tables 5.2 for visit times between 2000 and 2010, factors significant at $\alpha = 0.05$ are family history of PsA, NSAIDs, DMARDs, biologics, ESR and the history of past visits. In Table 5.3, since there are limited visits between 2010 and 2013, there is not much evidence of significant associations between visit intensities and these risk factors except for DMARDs and the history of past visits. In addition, we can see that the effects of ESR and past visit history on the present visit intensities may vary over the period of gap time $B(t)$. The IIV weights used below are estimated by formula (2.20) defined earlier with all the covariates that are considered in model fits based on (5.1) which are shown in Tables 5.1-5.3. The estimated IIV weights $\widehat{w}_i(t)$ at actual visit times t_{ij}'s across subjects have a minimum of 10.48, a median of 62.50, and a maximum of 8816.

From these analyses we can see that visit times are correlated with time-fixed as well as time-varying disease-related variables and also associated with past visit history, which might account for some unknown risk factors of visit intensities.

Table 5.1: Summary for the 1st subgroup (< 2000) of visit gap times by model (5.1) for the study of biologics intention. Time is in days. Cut-points selected for this subgroup are $150, 240, 383, 414, 692$. Except for treatment (i.e. ns, dm, bg) time-varying covariates change only at visits. Variable med.gap denotes the median length of past visit gap times.

	coef	exp(coef)	se(coef)	z	$Pr(>\|z\|)$
ESR	-0.0163	0.9838	0.0052	-3.1381	0.0017
sex	-0.4012	0.6695	0.2016	-1.9902	0.0466
age	-0.0210	0.9793	0.0095	-2.2063	0.0274
PS duration	0.0061	1.0062	0.0082	0.7501	0.4532
PsA duration	0.0011	1.0011	0.0127	0.0887	0.9294
family history of PS	0.0264	1.0267	0.2249	0.1172	0.9067
family history of PsA	-0.0197	0.9805	0.4146	-0.0475	0.9621
number of active joints	0.0049	1.0049	0.0094	0.5188	0.6039
number of damaged joints	0.0414	1.0423	0.0130	3.1827	0.0015
NSAIDs	0.3177	1.3739	0.1644	1.9324	0.0533
DMARDs	0.2320	1.2611	0.2288	1.0142	0.3105
biologics	0.0845	1.0882	0.2783	0.3036	0.7615
med.gap	0.0012	1.0012	0.0008	1.4919	0.1357

Table 5.2: Summary for the 2nd subgroup $(2000-2010)$ of visit gap times by model (5.1) for the study of biologics intention. Time is in days. Cut-points selected for this subgroup are $170, 247, 375, 450, 1030$. Variable med.gap denotes the median length of past visit gap times.

	coef	exp(coef)	se(coef)	z	$Pr(>\|z\|)$
sex	0.0774	1.0805	0.0477	1.6229	0.1046
age	0.0031	1.0031	0.0022	1.4419	0.1493
PS duration	0.0043	1.0043	0.0023	1.8874	0.0591
PsA duration	-0.0008	0.9992	0.0029	-0.2864	0.7746
family history of PS	-0.0288	0.9716	0.0451	-0.6397	0.5224
family history of PsA	-0.2147	0.8068	0.0772	-2.7820	0.0054
number of active joints	-0.0027	0.9973	0.0026	-1.0528	0.2924
number of damaged joints	0.0017	1.0017	0.0018	0.8978	0.3693
NSAIDs	0.0902	1.0944	0.0453	1.9919	0.0464
DMARDs	0.1299	1.1388	0.0454	2.8616	0.0042
biologics	0.1377	1.1477	0.0473	2.9130	0.0036
ESR:$I(B(t) \leq 180)$	0.0052	1.0052	0.0020	2.5587	0.0105
ESR:$I(B(t) > 180)$	-0.0088	0.9912	0.0020	-4.3467	0.0000
ESR:$I(B(t) > 365)$	-0.0026	0.9974	0.0052	-0.5051	0.6135
med.gap:$I(B(t) \leq 180)$	-0.0068	0.9932	0.0004	-16.2028	0.0000
med.gap:$I(B(t) > 180)$	-0.0009	0.9991	0.0003	-2.8924	0.0038
med.gap:$I(B(t) > 365)$	0.0005	1.0005	0.0005	0.9889	0.3227

Table 5.3: Summary for the 3rd subgroup (≥ 2010) of visit gap times by model (5.1) for the study of biologics intention. Time is in days. Cut-points selected for this subgroup are $172, 196, 265, 307, 364$. Variable med.gap denotes the median length of past visit gap times.

	coef	exp(coef)	se(coef)	z	$Pr(>\|z\|)$
ESR	-0.0019	0.9981	0.0026	-0.7242	0.4689
sex	-0.0204	0.9798	0.0704	-0.2894	0.7723
age	0.0019	1.0019	0.0033	0.5741	0.5659
PS duration	0.0010	1.0010	0.0033	0.3038	0.7613
PsA duration	-0.0015	0.9985	0.0043	-0.3440	0.7308
family history of PS	-0.0193	0.9809	0.0679	-0.2840	0.7764
family history of PsA	-0.0193	0.9809	0.1127	-0.1712	0.8641
number of active joints	0.0008	1.0008	0.0049	0.1606	0.8724
number of damaged joints	0.0007	1.0007	0.0027	0.2376	0.8122
NSAIDs	-0.0386	0.9622	0.0650	-0.5936	0.5528
DMARDs	0.1807	1.1980	0.0677	2.6672	0.0076
biologics	0.1109	1.1173	0.0701	1.5826	0.1135
med.gap:$I(B(t) \leq 180)$	-0.0068	0.9932	0.0008	-8.4731	0.0000
med.gap:$I(B(t) > 180)$	-0.0016	0.9984	0.0005	-2.9983	0.0027
med.gap:$I(B(t) > 365)$	-0.0007	0.9993	0.0016	-0.4552	0.6489

5.1.2 Estimation of the Failure Time Distribution

In this section, we will apply the IIV weighting approach to non-parametric, parametric, and semi-parametric estimation of distributions of failure time T_i (time to an increase in the number of damaged joints) to deal with the intermittent and informative clinic visit times. Note that all plots and regression parameters of failure time distributions will be presented with time in years and that there is no censoring upon a change of biologics status.

First, the non-parametric estimation of survival function or CDF of failure time T_i which was introduced in Section 4.1.1 will be illustrated for each of the two treatment groups. The bandwidths b_m and h_m in (4.1) and (4.6), respectively, are both selected to be 365 (days). Then, for the treated group, about 138 visits, and for the untreated group, about 145 visits, are contained in $[s_\ell - b_m, s_\ell + b_m]$ on average, and contribute to the crude or raw estimates of $S_T(s_\ell|A) = Pr(T_i > s_\ell|A)$ by (4.1), where $A = 0$ or 1. For each group, the crude estimates are monotonized by an isotonic regression, and then are smoothed by a non-parametric kernel regression by (4.6). The final estimate of the survival function of T, i.e. $S_T(t|A)$, $A = 0$ or $A = 1$, is shown in Figure 5.2. The grey solid and dashed curves are the pointwise empirical confidence intervals (ECIs) for the two survival curves, which are given by $\widehat{S}_T(t|A) \pm 1.96\ se(\widehat{S}_T(t|A))$, where se denotes a bootstrap estimated standard error. Bootstrap sample size is set to be 100 in this analysis. Turnbull (1976)'s estimator is utilized as a naive estimation method which does not adjust for the dependent inspection or visit times for comparison with the proposed weighted approach. In Figure 5.2, it can be seen that the Turnbull's estimate of the survival curve, $S_T(t|A = 0)$, basically agrees with the IIV weighted non-parametric estimate. However, Turnbull's estimator overestimates the survival probabilities of the treated group ($A = 1$). Additionally, there is some evidence

that the weighted estimates of $S_T(t|A)$, $A = 0$ and $A = 1$, are significantly different, since the corresponding ECIs hardly overlap. This suggests that patients who were intended to be treated with biologics are at a higher risk of an increase in the number of damaged joints. This is explainable because more severely sick patients will be prescribed biologics and these patients are usually at higher risk of joint damage. Since the initial biologics assignments are not randomized, later in this section we will apply the IPT weighting method which was introduced in Section 3.1.2 to adjust for the bias caused by potential confounding variables. However, the present analysis does not address the efficacy of treatment with biologics since some patients who did not receive biologics at t_{i0} later switched to be on it during follow-up. The analysis here is directed at whether there is an association between the intended treatment with biologics at t_{i0} and joint damage.

Next, we will apply the parametric estimation introduced in Section 2.2 to this example. A piecewise-constant proportional hazards model given treatment A is assumed for failure times T_i, i.e.

$$\lambda_T(t|A_i) = \sum_{s=1}^{S} \rho_s I_s(t) \exp(\beta A_i), \qquad i = 1, ..., n, \tag{5.2}$$

where $I_s(t) = I\{t \in (a_{s-1}, a_s]\}$, $s = 1, ..., 4$, cut-points are $a_0 = 0 < a_1 = 0.2 < a_2 = 1.5 < a_3 = 6 < a_4 = +\infty$, and ρ_s's are unknown non-negative constants. The survival function of T_i is then given by

$$S_T(t|A_i) = \exp\{-\sum_{s=1}^{S} \rho_s v_s(t) \exp(\beta A_i)\}, \qquad i = 1, ..., n, \tag{5.3}$$

where $v_s(t) = \int_0^t I_s(u)du$, $s = 1, ..., 4$. Figure 5.3 displays the non-parametric IIV weighted estimates of the baseline distribution of failure times, i.e. $A = 0$. The left plot shows the baseline cumulative hazard function, $\Lambda_T(t|A = 0)$, and the right plot shows the baseline CDF, $F_T(t|A = 0)$. We can see that a piecewise exponential distribution can be applicable

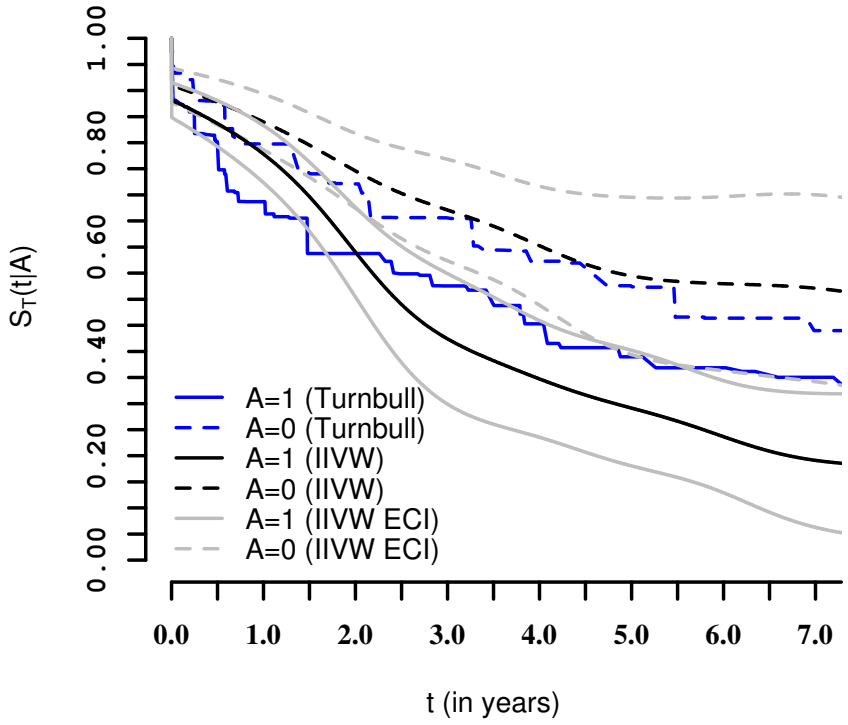


Figure 5.2: Non-parametric IIV weighted estimates of $S_T(t|A)$ where $A = 0$, or 1, where A denotes biologics intention. The number of patients who were followed up to 1, 3, 5, 7 years respectively are 181, 131, 77, 42 for $A = 0$, and are 183, 140, 102, 63 for $A = 1$.

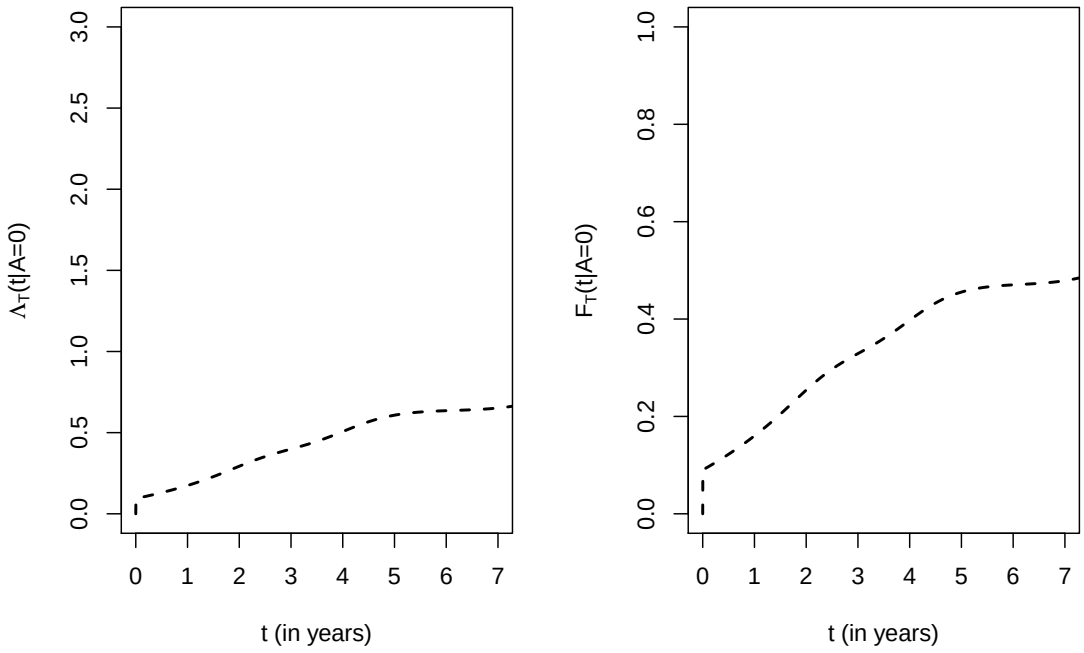


Figure 5.3: Non-parametric IIV weighted estimates of baseline cumulative hazard, $\Lambda_T(t|A=0)$, and baseline CDF, $F_T(t|A=0)$, where A denotes biologics intention.

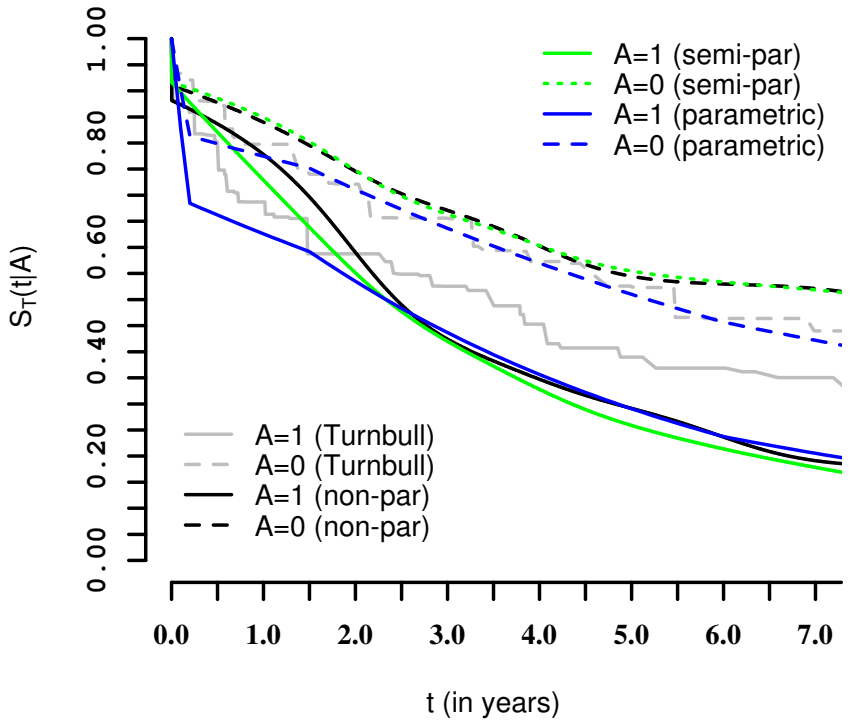


Figure 5.4: Plot of Turnbull estimates, non-parametric (denoted by non-par), parametric and semi-parametric (denoted by semi-par) IIV weighted estimates of $S_T(t|A)$, where $A = 0$, or 1, A denotes biologics intention. The number of patients who were followed up to 1, 3, 5, 7 years respectively are 181, 131, 77, 42 for $A = 0$, and are 183, 140, 102, 63 for $A = 1$.

to this example, and later we will compare the parametric estimated distributions based on model (5.2) with the non-parametric estimates we have obtained above.

A consistent estimate of β from (5.2) can be obtained by solving the weighted estimating function given in (2.13) where IIV weights are estimated as in the non-parametric estimation. Given the many parameters in the visit time model (5.1), it is computationally intensive to compute the asymptotic variance by the large sample theory introduced in Section 2.2.5, so the bootstrap is employed instead to estimate the standard errors of estimates. As a result, the estimates of ρ_1, ρ_2, ρ_3, ρ_4 and β solved by the R function *nleqslv* are: 1.033 (0.327), 0.061 (0.056), 0.111 (0.028), 0.079 (0.052), 0.607 (0.212), where the values in brackets are the bootstrap estimated standard errors. The estimated survival functions, $S_T(t|A)$, are displayed in Figure 5.4, denoted by a blue solid curve and a blue dashed curve for the treated group ($A = 1$) and untreated group ($A = 0$), respectively. We can see that the IIV weighted parametric estimates agree well with the IIV weighted non-parametric estimates which are denoted by black solid and dashed curves, though for $A = 0$ and when $t > 5$ years the parametric estimate is apart from the non-parametric estimate since the fully parametric model (5.2) shapes the estimated curve. The estimated hazard ratio $e^{\widehat{\beta}}$ indicates that patients initially treated with biologics have about 1.83 times the risk of failure of the untreated patients. Also, there is some evidence that two treatment groups have different risk of joint damage: the 95% empirical confidence interval for β computed by $\widehat{\beta} \pm 1.96\ se(\widehat{\beta})$ is $(0.192, 1.022)$, which indicates that the difference is significant at $\alpha = 0.05$.

Next, we will illustrate the IIV weighted semi-parametric estimation based on an additive hazards model which has a form of

$$\lambda_T(t|A_i) = \lambda_{T,0}(t) + \gamma A_i, \qquad i = 1, .., n, \tag{5.4}$$

where $\lambda_{T,0}(t)$ is an unspecified non-negative function of t. The IIV weights are again estimated as in the non-parametric estimation and parametric estimation. Then, the iterative algorithm introduced in Section 4.2.1 is implemented with the naive estimate of γ given by R function *aalen* in the package *timereg* as an initial value. As a result, the estimate $\widehat{\gamma}$ in model (5.4) equals 0.152 with a standard error estimated by bootstrap of 0.043, and the corresponding 95% empirical confidence interval equals $(0.067, 0.237)$. Therefore, at a significance level of 0.05, statistical evidence indicates that the two treatment groups have a different risk of joint damage. Specifically, the intended to be treated patients are found to be at a higher risk of an increase in the number of damaged joints. The average number of iterations is about 9, so the algorithm converges well and is not very computationally intensive in this case. The baseline hazard $\lambda_{T,0}(t)$ can be estimated by solving estimating function (4.14) and then monotonizing and smoothing the resulting raw estimates. Finally, the estimated survival curves are also shown in Figure 5.4 by a green solid curve and a green dotted curve for $A = 1$ and $A = 0$, respectively. Additionally, in Figure 5.4, we can see that for each group, when the IIV weighting method is applied, non-parametric estimates, piecewise exponential parametric estimates, and additive hazards semi-parametric estimates agree well with each other, which justifies the selection of model (5.2) and model (5.4) for failure times.

So far, we have demonstrated the IIV weighted estimation for the adjustment of intermittent clinic visits. Nevertheless, it is known that biologics are usually assigned for adult patients with moderate to severe joint activity. Some biologics such as Adalimumab can be used alone or are recommended to be used in combination with Methotrexate (MTX) or other DMARDs. Therefore, joint activity, disease duration, age and other concurrent treatments could be potential confounders of the association between the initial intention to treat with biologics and the joint damage event. To eliminate the bias caused by observed

confounding variables in this observational study, we can employ the inverse-probability-of-treatment (IPT) weighting method which was introduced in Section 3.1.2 for parametric estimation. A logistic regression model is fitted for treatment intention, A, at t_{i0}, to estimate the IPT weights (3.6). The model fitting summary is presented in Table 5.4 where we see that ESR, age, joint activity and NSAIDs are found to be significantly associated with biologics assignment at $\alpha = 0.05$. Although subjects for the analyses here were chosen according to their treatment status ($A = 0$ or 1), this model, when applied to a randomly selected person in the study group, will adjust for differences in risk factors across the two treatment subgroups. For parametric model (5.2), a double weight (3.7) will replace the IIV weight $w_i(t|\boldsymbol{\alpha})$ in the estimating function (2.13), and the doubly weighted parametric estimates of survival curves are shown in Figure 5.5 and are denoted by red solid and red dashed curves. The estimates of ρ_1, ρ_2, ρ_3, ρ_4 and β are 0.821 (0.321), 0.102 (0.066), 0.104 (0.048), 0.128 (0.062), and 0.580 (0.260). Similarly, for additive hazards model (5.4), a double weight (3.7) replaces the IIV weight in the estimating function (4.14) and (4.17) and the resulting estimates are displayed in Figure 5.6. The estimate of γ in model (5.4) is 0.155 with a bootstrap estimated standard error of 0.046. The 95% empirical confidence interval for β in model (5.2) is $(0.071, 1.089)$ and that for γ in model (5.4) is $(0.064, 0.246)$. Additionally, in Figure 5.5 or Figure 5.6, it is seen that the doubly weighted estimates and the IIV weighted estimates are very close. That is, the confounders which we considered in the analysis of Table 5.4 barely bias the estimates, which makes sense in this case, since patients can change treatment status later but stay in the same group, which was defined at t_{i0}, for estimation. However, unmeasured confounders may exist, so the analysis results still could have hidden bias.

We conclude that the patients who were intended to be treated with biologics are at a relatively higher risk of an increase in the number of damaged joints than those who were

not intended to be treated. Furthermore, we note again that because some individuals in the untreated group actually switched to treatment later, we cannot make conclusions about the efficacy of biologics treatment from this analysis.

Table 5.4: Logistic regression model fitting summary for A at t_{i0} where A denotes the intention of biologics.

	coef	se(coef)	z	$Pr(>\|z\|)$	
(Intercept)	-1.6699	0.3830	-4.3600	1.30E-05	***
ESR	0.0159	0.0070	2.2770	0.0228	*
year of enrolment	0.0394	0.0237	1.6580	0.0973	.
year of the visit before t_{i0}	0.0223	0.0314	0.7100	0.478	
sex	0.3383	0.2549	1.3270	0.1845	
age	-0.0483	0.0114	-4.2300	2.34E-05	***
PS duration	-0.0007	0.0124	-0.0580	0.9536	
PsA duration	0.0019	0.0195	0.0950	0.9242	
family history of PS	0.1907	0.2487	0.7670	0.4431	
family history of PsA	-0.0333	0.3930	-0.0850	0.9324	
number of active joints	0.1261	0.0174	7.2580	3.93E-13	***
number of damaged joints	0.0212	0.0123	1.7230	0.0849	.
NSAIDs	0.7318	0.2661	2.7500	0.006	**
DMARDs	0.1722	0.2544	0.6770	0.4984	
—					
Signif. codes:	*** 0.001	** 0.01	* 0.05	. 0.1	

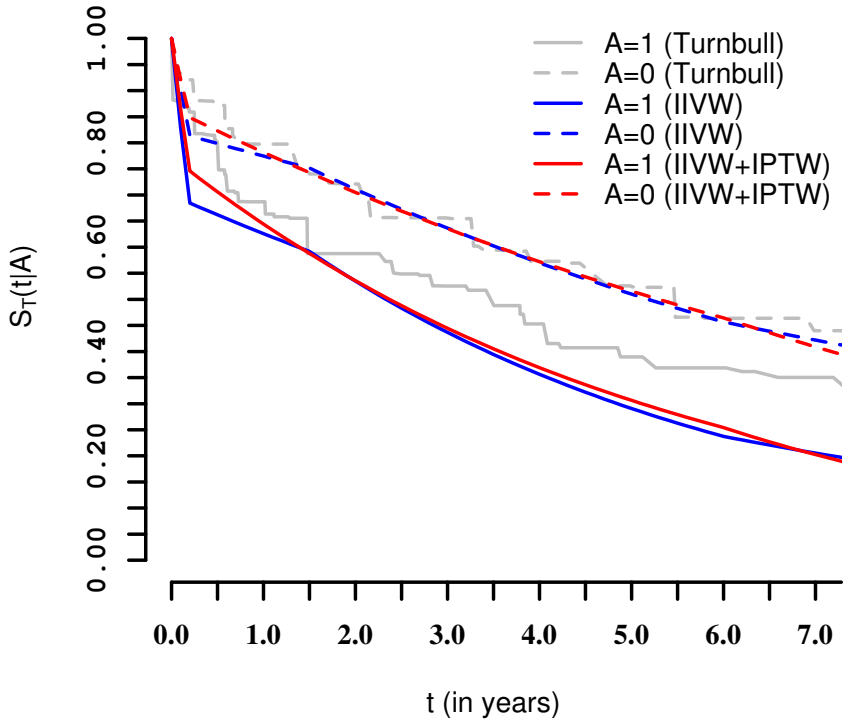


Figure 5.5: Plot of Turnbull, IIV weighted and IIV + IPT doubly weighted parametric estimates of $S_T(t|A)$ based on model (5.3) for the study of biologics intention, where $A = 0$, or 1. The number of patients who were followed up to 1, 3, 5, 7 years respectively are 181, 131, 77, 42 for $A = 0$, and are 183, 140, 102, 63 for $A = 1$.

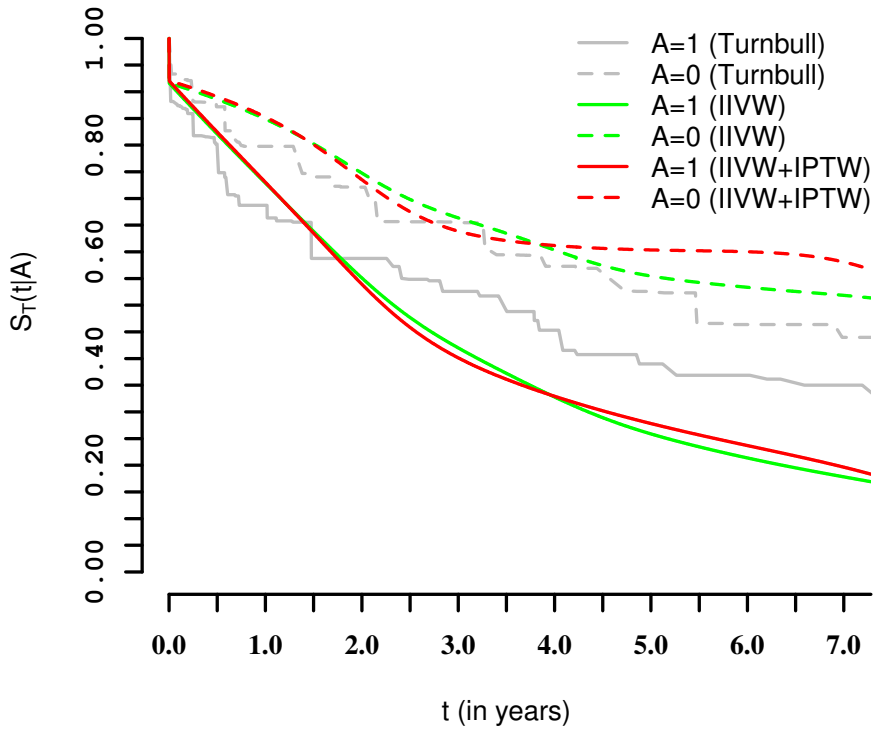


Figure 5.6: Plot of Turnbull, IIV weighted and IIV + IPT doubly weighted semi-parametric estimates of $S_T(t|A)$ based on model (5.4) for the study of biologics intention, where $A = 0$, or 1. The number of patients who were followed up to 1, 3, 5, 7 years respectively are 181, 131, 77, 42 for $A = 0$, and are 183, 140, 102, 63 for $A = 1$.

5.2 Association between Joint Damage and Biologics Treatment: Competing Risks Analysis

In this section, we want to study the efficacy of biologics treatment on time to an increase in the number of damaged joints. The analysis set includes the 177 patients who initially started using biologics after 2000 and the 177 untreated patients who were matched with the treated ones by calendar day. Comparing this analysis set with that used in Section 5.1, we exclude 60 patients who received biologics or were matched prior to 2000, since biologics, which are widely used recently, were mostly licensed around or after 2000. The range of visit dates in this analysis set is from 2000-01-17 to 2013-03-25. Among the 177 treated patients, 8.3 clinic visits were attended on average. For the 177 untreated patients, the average number of visits is 7.4. The estimation of IIV weights is conducted for two subgroups of visit gap times: $[2000, 2010)$ and $[2010, 2013]$, and the corresponding model fitting summaries based on model (5.1) are provided in Tables 5.5-5.6. We see that visit intensities are significantly associated with family history of PsA, treatment with DMARDs and biologics, ESR, and the history of past visits. The estimated IIV weights $\widehat{w}_i(t)$ at visit times t_{ij}'s across subjects have a minimum of 11.22, a median of 61.54, and a maximum of 8373. In this analysis, 14 individuals have long visit gap times artificially censored at 3 years, and 17 visits are deleted in total.

To study the time to joint damage increase under treatment with biologics, i.e. A_i, we consider the joint damage event and time-to-treatment switch as a pair of competing risks (CRs). Let T_{i1} be the time to joint damage increase under the initial treatment fixed at t_{i0} and let T_{i2} be the time to a treatment switch. Treatment history is recalled retrospectively at clinic visits and the exact start date and stop date of the usage of a specific drug are ascertained, so T_{i2} can be observed exactly, whereas T_{i1} is interval-censored. Since T_{i1}

Table 5.5: Model fit summary for the 1st subgroup of visit gap times, i.e. the previous visit lies in $[2000, 2010)$, based on model (5.1) in the competing risks analysis. Model is fitted in days. Cut-points selected for this subgroup are $170, 247, 375, 450, 1030$. Except treatment (i.e. ns, dm, bg) other time-varying covariates change only at visits. Variable med.gap denotes the median length of past visit gap times.

	coef	exp(coef)	se(coef)	z	$Pr(>\|z\|)$
sex	0.0555	1.0570	0.0512	1.0832	0.2787
age	0.0042	1.0042	0.0023	1.8184	0.0690
PS duration	0.0034	1.0034	0.0025	1.3937	0.1634
PsA duration	0.0004	1.0004	0.0034	0.1190	0.9053
family history of PS	-0.0149	0.9852	0.0482	-0.3093	0.7571
family history of PsA	-0.2433	0.7841	0.0820	-2.9673	0.0030
number of active joints	-0.0050	0.9950	0.0028	-1.7866	0.0740
number of damaged joints	-0.0008	0.9992	0.0020	-0.3969	0.6914
NSAIDs	0.0440	1.0450	0.0487	0.9053	0.3653
DMARDs	0.1038	1.1094	0.0487	2.1295	0.0332
biologics	0.1769	1.1935	0.0509	3.4721	0.0005
ESR:$I(B(t) \leq 180)$	0.0062	1.0063	0.0023	2.7659	0.0057
ESR:$I(B(t) > 180)$	-0.0114	0.9887	0.0024	-4.8263	0.0000
ESR:$I(B(t) > 365)$	-0.0044	0.9956	0.0066	-0.6661	0.5053
med.gap:$I(B(t) \leq 180)$	-0.0069	0.9932	0.0004	-15.4485	0.0000
med.gap:$I(B(t) > 180)$	-0.0007	0.9993	0.0003	-2.0712	0.0383
med.gap:$I(B(t) > 365)$	0.0009	1.0009	0.0005	1.7930	0.0730

Table 5.6: Model fit summary for the 2nd subgroup of visit gap times, i.e. the previous visit lies in $[2010, 2013]$, based on model (5.1) in the competing risks analysis. Model is fitted in days. Cut-points selected for this subgroup are $172, 196, 265, 307, 364$. Except treatment (i.e. ns, dm, bg) other time-varying covariates change only at visits. Variable med.gap denotes the median length of past visit gap times.

	coef	exp(coef)	se(coef)	z	$Pr(>\|z\|)$
ESR	-0.0019	0.9981	0.0026	-0.7393	0.4597
sex	-0.0232	0.9771	0.0725	-0.3199	0.7490
age	0.0030	1.0030	0.0033	0.8963	0.3701
PS duration	0.0003	1.0003	0.0034	0.0768	0.9388
PsA duration	-0.0012	0.9988	0.0044	-0.2718	0.7858
family history of PS	0.0202	1.0204	0.0697	0.2898	0.7720
family history of PsA	-0.0219	0.9784	0.1159	-0.1887	0.8503
number of active joints	0.0002	1.0002	0.0050	0.0356	0.9716
number of damaged joints	0.0006	1.0006	0.0029	0.1918	0.8479
NSAIDs	-0.0491	0.9521	0.0673	-0.7304	0.4652
DMARDs	0.1641	1.1784	0.0701	2.3396	0.0193
biologics	0.1173	1.1244	0.0719	1.6308	0.1029
med.gap:$I(B(t) \leq 180)$	-0.0068	0.9933	0.0008	-8.3028	0.0000
med.gap:$I(B(t) > 180)$	-0.0015	0.9985	0.0005	-2.7119	0.0067
med.gap:$I(B(t) > 365)$	-0.0008	0.9992	0.0017	-0.4944	0.6211

and T_{i2} are not independent given the observed history up to the most recent past visit, considering treatment switch as the censoring of T_{i1} and applying the inverse-probability-of-censoring weighting (IPCW) is not helpful. Following the convention of CR studies, we define the observed failure time as $T_i = \min(T_{i1}, T_{i2}, C_i)$, where we know there is no random drop-out in this example, so $C_i = \tau_i$. We let ε_i be the cause of failure, where $\varepsilon_i = 1$ if $T_i = T_{i1}$, $\varepsilon_i = 2$ if $T_i = T_{i2}$, and $\varepsilon_i = 0$ otherwise. First, the additive hazards model (4.12) is assumed for the subdistribution hazard of cause 1, i.e. $\lambda_1(t|A_i)$, which was defined by Fine and Gray (1999). One can consider $\lambda_1(t|A_i)$ as the hazard function of the improper failure time variable $T_{i1} = I(\varepsilon_i = 1)T_i + I(\varepsilon_i \neq 1)\infty$. Then, we still have the survival function model (4.13) which can be written as

$$S_1(t|A_i) = S_{1,0}(t) \exp\{-\gamma A_i t\}, \tag{5.5}$$

where $S_1(t|A_i) = 1 - F_1(t|A_i)$ and $F_1(t|A_i) = Pr(T_{i1} \leq t|A_i) = Pr(T_i \leq t, \varepsilon_i = 1|A_i)$ is referred to as the cumulative incidence function (CIF) for failure from cause 1 (i.e. joint damage increase) given biologics status A_i, and $S_{1,0}(t) = 1 - F_1(t|A_i = 0)$ is unspecified. In this section, we will investigate the effect of biologics on time to an increase in the number of damaged joints via the estimation of $F_1(t|A_i)$ non-parametrically or semi-parametrically.

The responses in (4.1) for non-parametric estimation are defined as $Y_i(t) = I(T_i \leq t, \varepsilon = 1)$, so the corresponding means are $\theta_\ell = F_1(\ell)$, for $\ell = 1, ..., m$. For semi-parametric estimation, the responses in (4.14) and (4.17) are defined as $P_i(t) = 1 - Y_i(t)$ with means $E\{P_i(t)|A_i\} = S_1(t|A_i) = 1 - F_1(t|A_i)$ that are modelled by (5.5). The bandwidths, b_m and h_m, in the non-parametric estimation and the semi-parametric estimation are selected to be 180 days, so that for either group m equals 12 and there are $100 - 120$ visits lying in $[s_\ell - b_m, s_\ell + b_m]$.

We know that treatment is non-randomized at t_{i0}, so to study the efficacy of biologics

on a joint damage event, we apply the IPT weights estimated by (3.6) via fitting a logistic regression like (3.4). The model fitting summary is presented in Table 5.7 where we see treatment assignment is significantly associated with joint activity, age, ESR, and concurrent treatment with NSAIDs at $\alpha = 0.05$. To adjust for these confounders, we substitute the IIV weight $w_i(t)$ in estimating functions (4.1), (4.14), and (4.17) with the double weight (3.7) where w_i^* is the IPT weight.

Table 5.7: Logistic regression model fitting summary for A at t_{i0} where A denotes biologics treatment status in the competing risks analysis.

	coef	se(coef)	z	$Pr(>\|z\|)$	
(Intercept)	-1.9146	0.4924	-3.8880	0.0001	***
ESR	0.0166	0.0077	2.1580	0.0309	*
year of enrolment	0.0340	0.0256	1.3270	0.1844	
year of the visit before t_{i0}	0.0651	0.0465	1.4010	0.1613	
sex	0.3094	0.2732	1.1330	0.2574	
age	-0.0478	0.0122	-3.9130	9.13E-05	***
PS duration	0.0038	0.0130	0.2930	0.7694	
PsA duration	-0.0083	0.0217	-0.3800	0.7043	
family history of PS	0.4130	0.2683	1.5390	0.1237	
family history of PsA	-0.0448	0.4322	-0.1040	0.9175	
number of active joints	0.1256	0.0186	6.7580	1.40E-11	***
number of damaged joints	0.0254	0.0130	1.9500	0.0512	.
NSAIDs	0.6965	0.2932	2.3760	0.0175	*
DMARDs	0.0254	0.2750	0.0920	0.9264	

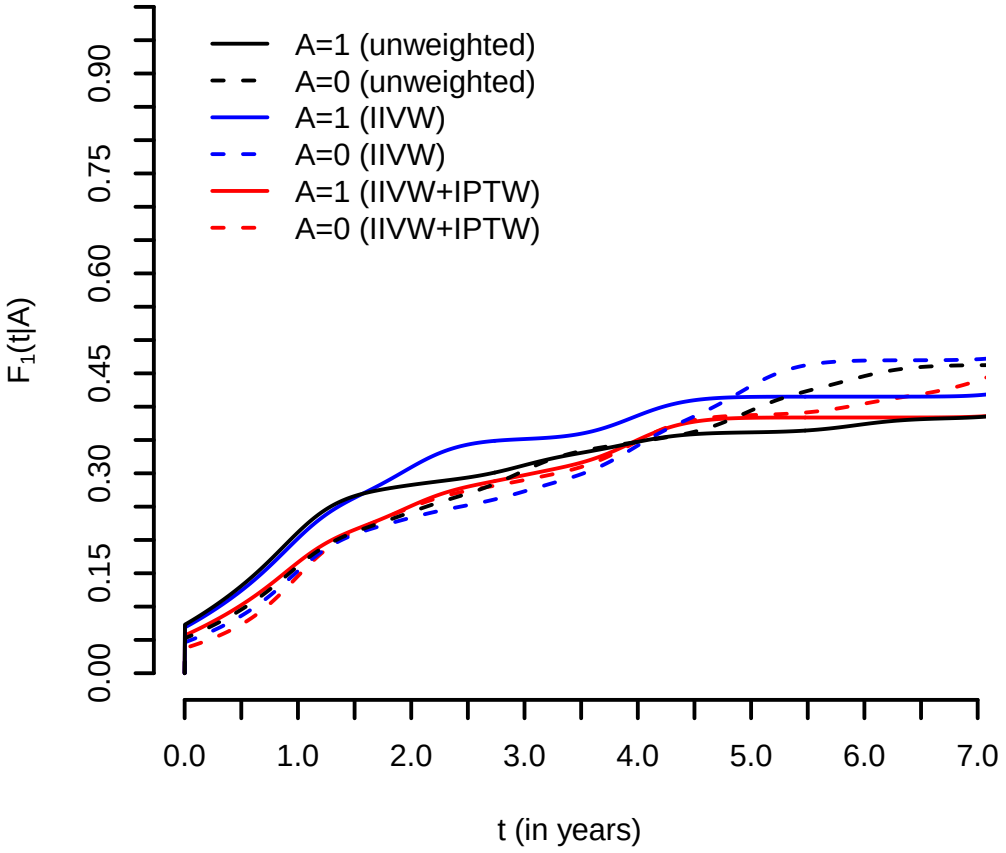


Figure 5.7: Plot of unweighted, IIV weighted and IIV + IPT doubly weighted non-parametric estimates of $F_1(t|A)$, where $A = 0$, or 1, A denotes biologics status in the competing risks analysis. The number of patients who were followed up to 1, 3, 5, 7 years respectively are 155, 109, 70, 38 for $A = 0$, and are, 153, 113, 82, 46 for $A = 1$.

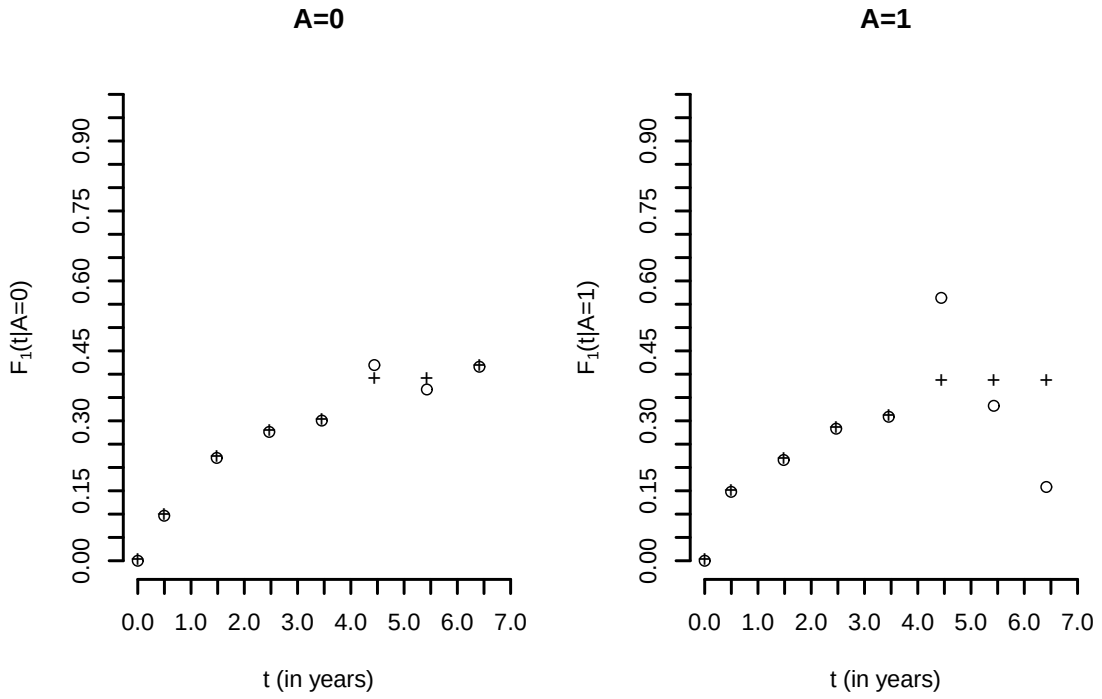


Figure 5.8: Doubly weighted non-parametric crude estimate, denoted by a circle, and isotonic estimate, denoted by a plus, of $F_1(s_\ell|A)$, where $\ell = 1, ..., m$ and $A = 0$, or 1. Variable A denotes biologics status in the competing risks analysis. The number of patients who were followed up to 1, 3, 5, 7 years respectively are 155, 109, 70, 38 for $A = 0$, and are, 153, 113, 82, 46 for $A = 1$.

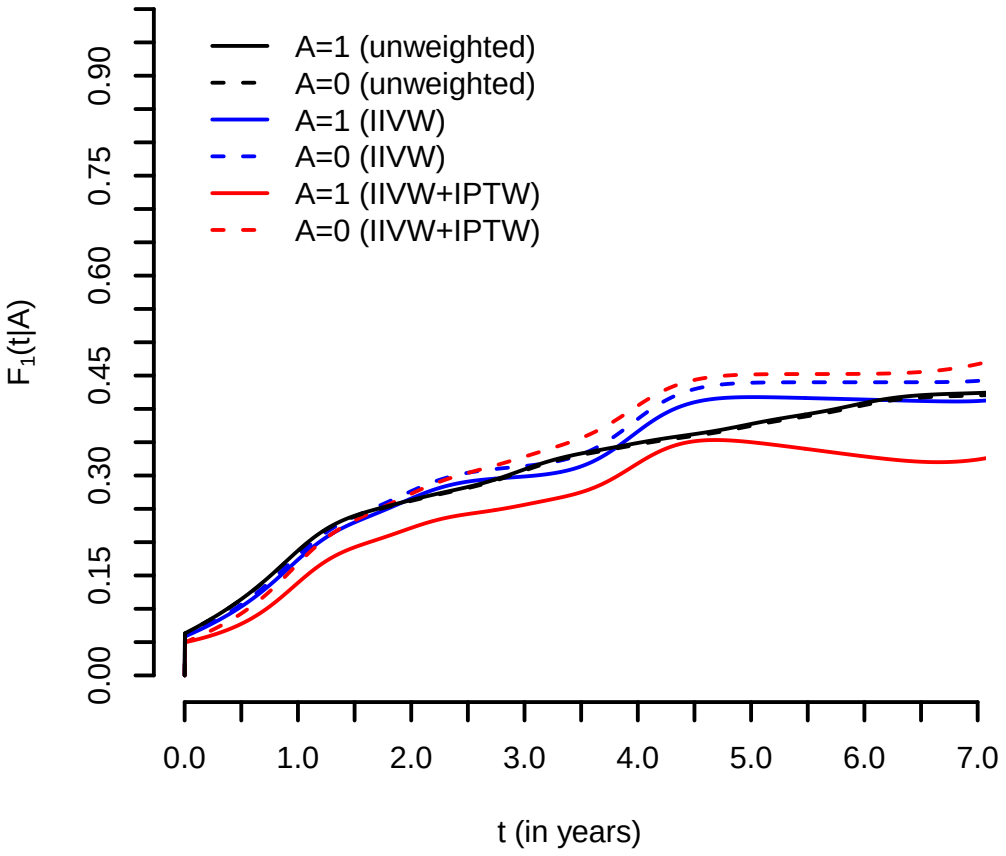


Figure 5.9: Plot of unweighted, IIV weighted and IIV + IPT doubly weighted semi-parametric estimates of $F_1(t|A)$ based on (5.5), where $A = 0$, or 1, A denotes biologics status in the competing risks analysis. The number of patients who were followed up to 1, 3, 5, 7 years respectively are 155, 109, 70, 38 for $A = 0$, and are, 153, 113, 82, 46 for $A = 1$.

Figure 5.7 shows the IIV weighted and the doubly weighted non-parametric estimates of $F_1(t|A)$ for $A = 0$ and $A = 1$. Unweighted estimates obtained by setting $w_i(t)$ (IIV weight) and w_i^* (IPT weight) as 1 are considered as naive estimates and are also displayed in Figure 5.7 for comparison. The IIV weighted estimates and naive estimates shown in Figure 5.7 indicate that treated patients have relatively higher cumulative incidence of an increase in the number of damaged joints than untreated patients during the early years, which is not expected and suggests that there may exist bias. Then, after the IPT weight is further applied, it has been corrected a bit by the measured confounders considered in the analysis of Table 5.7, and the doubly weighted estimates (red solid and red dashed curves) in Figure 5.7 indicate that two treatment groups have no difference in terms of cumulative incidence of joint damage increase. However, there may exist unknown factors which are associated with joint damage as well as visit times and unmeasured confounders related to treatment and disease progression that fail to be adjusted for, due to the limitation of information or data; we will further discuss this later in this chapter and in the next chapter. In addition to Figure 5.7, raw estimates $\bar{Y}_\ell$, given in (4.3), and isotonic estimates for each group when double weights are applied are shown in Figure 5.8.

Figure 5.9 displays unweighted, IIV weighted and doubly weighted semi-parametric estimates of $F_1(t|A)$ versus A, based on the additive hazards model (5.5). We can see that treated patients and untreated patients have no difference in terms of the incidence of joint damage increase from the unweighted estimates of $F_1(t|A)$. When the IIV weight is applied, it is found that the two treatment groups have slightly different cumulative incidences, but when the IPT weight is further applied, it is easily seen that treated patients have a lower cumulative incidence than untreated patients. The unweighted estimate of γ is 0.001 with a bootstrap estimated standard error (se) of 0.024 and the 95% empirical CI is $(-0.047, 0.049)$, which indicates there is no evidence that biologics have any effect on

the risk of joint damage increase at $\alpha = 0.05$. In this analysis, 500 bootstrap samples are used to estimate standard errors. The IIV weighted estimate of γ equals -0.008 with a standard error of 0.035 and the 95% empirical CI is $(-0.076, 0.061)$, so there is once again no evidence that biologics have an effect on decreasing the risk of joint damage increase. Then, when the IPT weight is applied, the estimated γ equals -0.034 with a standard error of 0.037 and a 95% empirical CI of $(-0.107, 0.039)$.

Based on the results here, there is not sufficient evidence that biologics have an effect on reducing the risk of an increase in the number of damaged joints, though Figure 5.9 based on the additive hazards model (5.5) suggests that treated patients have relatively lower cumulative incidence of a joint damage increase than untreated patients when the double weighting method is employed. However, we have to carefully draw conclusions on these analysis results, because there are several concerns in our data set. Sample size is small, which gives large standard errors and wide confidence intervals. Also, there are actually eight specific drugs in the category of biologics and patients may use more than one drug simultaneously. Here, we consider the class of all biologic drugs as one treatment since sample size will be even smaller if we focus on one particular drug. Moreover, due to unmeasured confounders, the bias induced by confounding may not be sufficiently corrected by the IPT weights. This will be further discussed in the next section. In addition, the doubly weighted semi-parametric estimates of $F_1(t|A)$ shown in Figure 5.9 do not agree well with the non-parametric estimates shown in Figure 5.7, which may suggest a lack of fit for the additive hazards model (5.4). By (5.5), we see that the additive hazards model forces the proportion $S_1(t|A = 1)/S_1(t|A = 0)$ to grow exponentially over time with a negative estimate of γ. Model assumption could be relaxed by allowing a time-varying coefficient γ in model (5.4). This demands further investigation, and it will be considered as future work.

5.3 Concluding Remarks on the Analyses

To conclude, intermittent inspection times are common in longitudinal cohort studies, especially when the protocol is designed to be history-dependent, physician or patient-driven. Since how often and when patients come to visit the clinic could be associated with disease status or disease-related variables, naive analysis of the observed data may lead to misleading conclusions due to the dependent interval censoring. From the above analyses, we see that the IIV weighted estimates effectively adjust for the informative visit times and lead to more plausible conclusions.

Like many weighting adjustment methods, the proposed IIV weighted estimating function approach requires some crucial assumptions. In addition to the condition (B1) about conditionally independent observation scheme which was introduced in Section 2.2.1 and the assumption of independent drop-out, we also need the visit time model and the failure time model to be correctly specified, so careful selection of variables and models is essential. It is noted that there may exist unknown or unobserved factors and their history between the outcome process and the visit process in the PsA analyses; that is, the assumption (B1) could be violated. Thus, the IIV weighting method may not sufficiently adjusts for the selection bias caused by dependent follow-up. Sensitivity analysis tools, e.g. Scharfstein et al. (1999), can be helpful to check the impacts of a non-ignorable observation scheme.

In observational studies, direct comparison of two treatment groups could lead to biased results because of potential confounders, since treatment is not randomized, as we showed and discussed at the ends of Section 5.1 and Section 5.2. The IPT weighting method can help with eliminating the bias caused by confounding. However, to draw reliable causal conclusions the assumption (C1) which assumes that there are no unmeasured confounders has to be satisfied. In this data set, due to the missing or limited information on poten-

tial confounding variables between treatment (biologics) and failure times, rigorous causal conclusions cannot be drawn, but we see that the estimates which are adjusted by the IPT weights are more plausible and presumably correct some bias induced by confounding. We know that sicker patients are more likely to receive and adhere to biologics which is known as a second-line treatment. In addition, economic status may be an important factor because of the high expenses of biologics; we do not have information about this. Therefore, biomarkers that reveal disease severity, efficacy of other treatments, economic status, patients' preferences, and information about side effects could be confounders between treatment and failure times. Although we have adjusted for potential confounders such as ESR, joint activity, and joint damage measured at or prior to t_{i0}, other variables like Health Assessment Questionnaire (HAQ) score, Psoriasis Area Severity Index (PASI), employment status or education status may be good candidates for confounders but they have a lot of missingness in our database. Also, we don't have information available to check the positivity condition for treatment which was introduced in (C1). Due to the limitation of data, this setting is not ideal for causal inference, because the assumption (C1) is likely not satisfied. However, the analyses here illustrate the proposed estimation methods. Moreover, the problems that we discuss are present in the majority of observational studies, where the assumption of no unmeasured confounders is rarely plausible.

Chapter 6

CONCLUSION, DISCUSSION AND FUTURE RESEARCH

In practice, the periodic inspection times of a longitudinal cohort study are often irregular. For example, in the PsA Toronto Cohort Study, patients visit the PsA clinic at different times due to circumstances that may relate to their health status, disease status, responses to therapies, etc. Irregular inspection times may carry information about the outcomes of interest, e.g. sicker patients visit the clinic more often and are also more likely to experience disease progression or other disease-related events of interest. In this sense, when and how often the investigators observe outcomes are dependent on the values of outcomes or outcome-related variables. In other words, irregular inspection times could be outcome-dependent and may lead to a biased sample for analysis. As a result, standard analysis methods such as MLE and GEE methods based on observed data could yield biased estimates and even misleading conclusions. This was seen in the simulation studies and the analyses of the PsA cohort study in the preceding chapters. Although multivariate

regression models which include all the potential common risk factors between the visit process and the outcome process as covariates may take the outcome-dependent inspection times into account, the targets of inference are often the marginal or partially conditional effects of some primary factors on outcomes. The inverse-intensity-of-visit (IIV) weighting method which is proposed in this thesis adjusts for informative inspection times and results in the estimates of marginal associations or effects.

When the outcome of interest is a failure time or an event time, intermittent visits and irregular inspection times could cause failure times to be dependently interval-censored. Then, standard analysis methods, e.g. MLE based on likelihood (2.1) for interval-censored data, may lead to biased estimates, because most standard methods and existing software for failure time data assume that the censoring is independent and non-informative. Sensitivity analysis tools can be used to check the dependence of censoring, e.g. Siannis (2004), Siannis et al. (2005) and Zhang and Heitjan (2006). It is appealing that our proposed IIV weighted estimating functions based on a class of (marginal) binary responses defined for failure times, e.g. $P_i(t) = I(T_i > t)$, can deal with dependent interval censoring and result in the estimates of marginal or partially conditional effects. Meanwhile, other problems in survival analysis such as left truncation, informative drop-out times, confounded treatment, and competing risks can be dealt by similar formulations. When failure occurrence discontinues visits, we proposed to discretize the visit process, create pseudo visits every time unit after failure occurrence and consider the visit intensities as 1. The discretization of the visit process is fine in practical studies, since data are recorded in certain time units, e.g. the PsA data is recorded in days.

The assumptions (B1) in Section 2.2.1 and (B1*) in Section 2.2.4 that a visit time and the response defined for failure time outcome is independent given the observed history of covariates, past visits and past responses. This assumption is generally untestable as other

assumptions for ignorable coarsening. If a visit time is associated with factors which are not measured, e.g. patients' preference and personalities, or the history of some variables between the previous visit and the present one, then the assumption (B1) or (B1*) could be violated. Random effect models can be considered in that case, but the estimates obtained from a random effect model usually have a lack of interpretability and the assumption regarding random effect(s) is usually untestable as well. A combination of the IIV weighted estimating function approach and random effect models might be considered so that the known factor or history can be adjusted for by weighting and unknown factors can be represented by random effects. More discussion on that can be found in Pullenayegum and Lim (2014).

We have introduced a variety of survival models in simulation studies, e.g. a proportional hazards model, a log-normal AFT model and a semi-parametric additive hazards model. Some assumptions based on recent articles are proposed to simplify the data generation and the forms of marginal outcome models, which might not always be plausible in the real world. For example, we assumed that $\bar{L}^{obs}(t^-) = \bar{L}(t^-)$ and L_m is independent of $\bar{L}_{m-1}$ given A in Section 2.3, following Young and Tchetgen Tchetgen (2014). The key issue here is that there are few conditional outcome models that lead to marginal models such as proportional hazards models, so simulations in the literature make very restrictive assumptions, which do, however, allow assessment of the proposed methods of estimation.

In observational studies, treatments are non-randomized and are often confounded by known and unknown factors which are common causes of treatment assignments as well as outcomes. The inverse-probability-of-treatment (IPT) weighting method provides a useful and convenient way to adjust for measured confounders and mimic a randomized trial so that causal conclusions can be drawn. Other confounder adjustment methods via propensity scores, such as matching, also could be helpful to eliminate the selection bias induced

by confounders. However, the important assumption of strongly ignorable treatment assignment, i.e. (A1) or (C1), has to be satisfied to ensure no hidden bias is caused by unmeasured confounders. However, this assumption is generally untestable and too ideal to be true in practice. For example, in the PsA data, if patients' unstated preference about treatments also affects outcome, it could be an unmeasured confounder. Furthermore, in practical studies, even some known factors may fail to be measured properly, as we discussed in Section 5.3. Good background information may help with the selection of appropriate potential confounders to reduce selection bias; sound collection and manipulation of data is also suggested to support reliable causal inference.

In the simulations we have discussed in the preceding chapters, we see that in some cases coverage probabilities are slightly lower than the nominal level and the empirical standard errors (ESEs) are a bit greater than the average asymptotic standard errors (ASEs), especially when estimating functions are doubly weighted, e.g. in Table 3.2 and Table 3.4. In the PsA case, since some patients have extremely long visit gap times, e.g. longer than 20 years, which may lead to extremely large values of the IIV weights, $w_i(t)$, if the denominator is an estimated visit intensity of very small value, so we artificially censor visit gap times longer than 3 years and consider those patients as lost to follow-up. It is known that large variability in the weights can cause estimators with large variances and the estimators may even fail to be approximately normally distributed (Robins et al., 2000). One way to mitigate the variability caused by large weights is to stabilize weights. For example, the IIV weight $w_i(t)$ can be replaced by $sw_i(t) = a(t)/\lambda_N(t|\boldsymbol{Z_i}(t^-))$, where $a(t)$ stabilizes the weights and leaves the estimating functions unbiased (Bůžková and Lumley, 2007, 2009; Pullenayegum and Feldman, 2013; Pullenayegum and Lim, 2014). Alternatively, we could truncate extremely large weights to reduce the variances of estimates at a cost of some bias. For example, Bembom and van der Laan (2008) proposed a selection

of truncation level of the IPT weights on a basis of minimizing the expected MSE of the estimator. Later, Pullenayegum and Feldman (2013) extended their selection method to truncate the IIV weights for irregularly observed longitudinal data.

Inverse probability weighting methods usually need the model for outcomes and the models for weights correctly specified. A class of augmented inverse probability weighted (AIPW) estimators have been proposed, e.g. Robins et al. (1994); Scharfstein et al. (1999); Bang and Robins (2005); Seaman and Copas (2009); Pullenayegum and Feldman (2013). An AIPW estimator is consistent when either the outcome model or the weight model is correct, so it is also referred to as a doubly robust estimator. If applied to our IIV weighted estimator, for example, the IIV weighted estimating function (2.23) can be modified as

$$\begin{aligned} U_{ik}^{DR} =& c(\mu_T)\left\{\frac{dN_{ik}}{\lambda_N(a_{ik}|\boldsymbol{Z}_{i,k-1})}[P_{ik}-\mu_T(a_{ik})]+\left[1-\frac{dN_{ik}}{\lambda_N(a_{ik}|\boldsymbol{Z}_{i,k-1})}\right][E(P_{ik}|\boldsymbol{\mathcal{H}}_{i,k-1}^{obs})-\mu_T(a_{ik})]\right\} \\ =& c(\mu_T)\left\{\frac{dN_{ik}}{\lambda_N(a_{ik}|\boldsymbol{Z}_{i,k-1})}[P_{ik}-E(P_{ik}|\boldsymbol{\mathcal{H}}_{i,k-1}^{obs})]+[E(P_{ik}|\boldsymbol{\mathcal{H}}_{i,k-1}^{obs})-\mu_T(a_{ik})]\right\}, \end{aligned} \tag{6.1}$$

where $P_{ik} = I(T_i > a_{ik})$, $\mu_T(a_{ik}) = E(P_{ik})$, and note that $\boldsymbol{Z}_{i,k-1} = \boldsymbol{h}\{\boldsymbol{\mathcal{H}}_{i,k-1}^{obs}\}$. Here, we consider a discrete visit process as an example and assume that there is no random drop-out for simplicity. It can be shown that if either the IIV weight model $\lambda_N(a_{ik}|\boldsymbol{Z}_{i,k-1})$ or a "working" outcome model $E(P_{ik}|\boldsymbol{\mathcal{H}}_{i,k-1}^{obs})$, e.g. (2.33) given in Section 2.3.1, is correctly specified, then (6.1) is unbiased. Furthermore, the augmented inverse probability weighting method can be extended to our doubly weighted estimator so that a triply robust estimator can be obtained, i.e.

$$\begin{aligned} U_{ik}^{TR} =& c(\mu_{T^a})\left\{\left[\frac{dN_{ik}}{\lambda_N(a_{ik}|\boldsymbol{Z}_{i,k-1})}\frac{I(A_i=a)}{\pi_i(a|\boldsymbol{V_i})}-\frac{dN_{ik}}{\lambda_N(a_{ik}|\boldsymbol{Z}_{i,k-1})}-\frac{I(A_i=a)}{\pi_i(a|\boldsymbol{V_i})}\right][P_{ik}^a-E(P_{ik}^a|\boldsymbol{\mathcal{H}}_{i,k-1}^{obs})]\right. \\ & \left.-[E(P_{ik}^a|\boldsymbol{\mathcal{H}}_{i,k-1}^{obs})-\mu_{T^a}(a_{ik})]\right\}, \end{aligned} \tag{6.2}$$

where $P_{ik}^a = I(T_i^a > a_{ik})$; $\mu_{T^a}(a_{ik}) = E(P_{ik}^a) = Pr(T_i^a > a_{ik})$; T_i^a is the counterfactual outcome under treatment a; and we defined the IPT weight $w_i^* = 1/\pi_i(a|\boldsymbol{V_i}) = 1/Pr(A_i = a|\boldsymbol{V_i})$. Here, we assume $E(P_{ik}^a|dN_{ik}, \boldsymbol{\mathcal{H}}_{i,k-1}^{obs}, A_i, \boldsymbol{V_i}) = E(P_{ik}^a|\boldsymbol{\mathcal{H}}_{i,k-1}^{obs})$ for a "working" outcome model, $E(dN_{ik}|P_{ik}^a, \boldsymbol{\mathcal{H}}_{i,k-1}^{obs}, A_i, \boldsymbol{V_i}) = E(dN_{ik}|\boldsymbol{Z}_{i,k-1})$ for the IIV weight model, and $E(A_i|P_{ik}^a, \boldsymbol{V_i}, \boldsymbol{\mathcal{H}}_{i,k-1}^{obs}, dN_{ik}) = E(A_i|\boldsymbol{V_i})$ for the IPT weight model. It can be shown that if

any one of the IIV weight model, the IPT weight model, and a "working" outcome model $E(P_{ik}^a|\boldsymbol{\mathcal{H}}_{i,k-1}^{obs})$, e.g. (3.12) in Section 3.2, is correct, then (6.2) is an unbiased estimating function. This is one of my future research directions.

The non- or semi-parametric estimation introduced in Chapter 4 is relatively robust to the model assumption of the outcome process. Since kernel smoothing is employed, the selections of bandwidths and kernel functions affect the final estimation results. For instance, there is a tradeoff between bias and variance when different bandwidths are chosen. Explicitly, small bandwidths result in smaller bias but larger variances, while larger bandwidths lead to less variability but more bias. An optimal bandwidth or a data driven bandwidth might be considered in the future, though a simple selection based on the number of visits contained by a window, i.e. $[t-h, t+h]$ where h denotes a bandwidth, works in our simulations and real data analyses.

This thesis research is initially inspired by the intermittent inspection times in the PsA cohort study. The impacts of irregular (and potentially informative) inspection times on the analysis of outcomes have been addressed. The IIV weighting method was proposed and studied by a few authors for irregularly observed longitudinal data. We extend this method to study time-to-event or failure time data with dependent follow-up. When inspection times are outcome-dependent, failure times are subject to dependent interval censoring. Literature on dependently interval-censored failure times is very limited, so we believe this thesis contributes to this topic significantly. Parametric estimation, non- and semi-parametric estimation of marginal failure time distributions in the presence of dependent inspection times has been comprehensively discussed. Monotone smoothing which can be implemented by existing software is introduced to conduct the IIV weighted non- and semi-parametric estimation procedures. Additionally, causal inference is also considered and the IPT weight can be easily combined with the IIV weighting method.

Ideally, regular inspection times make the measured responses be a completely random sample. Unfortunately, follow-up is often irregular for a variety of reasons in practice. We suggest that reasons of missed visits or the deviation of planned visit gap times should be investigated so that appropriate assessment and adjustment of visit times can be achieved. Then, the introduced IIV weighting method is believed a useful and convenient approach to eliminate the selection bias due to irregular inspection times.

References

Odd O Aalen. A model for nonparametric regression analysis of counting processes. In *Lecture Notes in Statistics 2*, pages 1–25. Springer, New York, 1980.

Odd O Aalen. A linear regression model for the analysis of life times. *Statistics in Medicine*, 8(8):907–925, 1989.

Odd O Aalen, Richard J Cook, and Kjetil Røysland. Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Analysis*, 21(4):579–593, 2015.

Per Kragh Andersen, John P Klein, and Susanne Rosthøj. Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, 90 (1):15–27, 2003.

Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

Richard E Barlow, David J Bartholomew, JM Bremner, and H Daniel Brunk. *Statistical Inference under Order Restrictions: the Theory and Application of Isotonic Regression*. Wiley, New York, 1972.

Oliver Bembom and Mark J van der Laan. Data-adaptive selection of the truncation level for inverse-probability-of-treatment-weighted estimators. *U.C. Berkeley Division of Biostatistics Working Paper Series*, (230), 2008. URL `http://biostats.bepress.com/ucbbiostat/`.

Petra Bůžková and Thomas Lumley. Longitudinal data analysis for generalized linear models with follow-up dependent on outcome-related variables. *Canadian Journal of Statistics*, 35(4):485–500, 2007.

Petra Bůžková and Thomas Lumley. Semiparametric modeling of repeated measurements under outcome-dependent follow-up. *Statistics in Medicine*, 28(6):987–1003, 2009.

Na Cai, Wenbin Lu, and Hao Helen Zhang. Time-varying latent effect model for longitudinal data with informative observation times. *Biometrics*, 68(4):1093–1102, 2012.

Vinod Chandran, Richard J Cook, Jonathan Edwin, Hua Shen, Fawnda J Pellett, Sutharshini Shanmugarajah, Cheryl F Rosen, and Dafna D Gladman. Soluble biomarkers differentiate patients with psoriatic arthritis from those with psoriasis without arthritis. *Rheumatology*, 49(7):1399–1405, 2010.

Baojiang Chen, Grace Y Yi, and Richard J Cook. Analysis of interval-censored disease progression data via multi-state models under a nonignorable inspection process. *Statistics in Medicine*, 29(11):1175–1189, 2010.

Ling Chen and Jianguo Sun. A multiple imputation approach to the analysis of interval-censored failure time data with the additive hazards model. *Computational Statistics & Data Analysis*, 54(4):1109–1116, 2010.

William G Cochran. The planning of observational studies of human populations. *Journal of the Royal Statistical Society. Series A (General)*, 128:234–255, 1965.

Richard J Cook and Jerald F Lawless. *The Statistical Analysis of Recurrent Events*. Springer, New York, 2007.

Richard J Cook and Jerald F Lawless. Statistical issues in modeling chronic disease in cohort studies. *Statistics in Biosciences*, 6(1):127–161, 2014.

Somnath Datta and Rajeshwari Sundaram. Nonparametric estimation of stage occupation probabilities in a multistage model with current status data. *Biometrics*, 62(3):829–837, 2006.

Jan de Leeuw, Kurt Hornik, and Patrick Mair. Isotone optimization in R: pool-adjacent-violators algorithm (PAVA) and active set methods. *Journal of Statistical Software*, 32 (5):1–24, 2009.

Peter Diggle and Michael G Kenward. Informative drop-out in longitudinal data analysis. *Applied Statistics*, 43:49–73, 1994.

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

Dianne M Finkelstein, William B Goggins, and David A Schoenfeld. Analysis of failure time data with dependent interval censoring. *Biometrics*, 58(2):298–304, 2002.

D Ghosh. Efficiency considerations in the additive hazards model with current status data. *Statistica Neerlandica*, 55(3):367–376, 2001.

Dafna D Gladman. Psoriatic arthritis. *Rheumatic Disease Clinics of North America*, 24 (4):829–844, 1998.

Sander Greenland, James M Robins, and Judea Pearl. Confounding and collapsibility in causal inference. *Statistical Science*, 14(1):29–46, 1999.

Dagmar M Hajducek and Jerald F Lawless. Estimation of finite population duration distributions from longitudinal survey panels with intermittent followup. *Lifetime Data Analysis*, 19(3):371–392, 2013.

Peter Hall, Jeff Racine, and Qi Li. Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 99(2):1015–1026, 2004.

WG Havercroft and V Didelez. Simulating from marginal structural models with time-dependent confounding. *Statistics in Medicine*, 31(30):4190–4206, 2012.

Wenqing He and Jerald F Lawless. Flexible maximum likelihood methods for bivariate proportional hazards models. *Biometrics*, 59(4):837–848, 2003.

Xuming He and Peide Shi. Monotone B-spline smoothing. *Journal of the American Statistical Association*, 93(442):643–650, 1998.

Daniel F Heitjan and Donald B Rubin. Ignorability and coarse data. *The Annals of Statistics*, 19(4):2244–2253, 1991.

Miguel A Hernán and James M Robins. Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health*, 60(7):578–586, 2006.

Miguel A Hernán and James M Robins. *Causal Inference*. Forthcoming by Chapman & Hall/CRC, New York, 2016.

Miguel A Hernán, Babette Brumback, and James M Robins. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 11(5):561–570, 2000.

Miguel A Hernán, Stephen R Cole, Joseph Margolick, Mardge Cohen, and James M Robins. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiology and Drug Safety*, 14(7):477–491, 2005.

Joseph W Hogan, Jason Roy, and Christina Korkontzelou. Tutorial in biostatistics. handling drop-out in longitudinal studies. *Statistics in Medicine*, 23(9):1455–1497, 2004.

Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47 (260):663–685, 1952.

Chiu-Hsieh Hsu, Jeremy MG Taylor, Susan Murray, and Daniel Commenges. Multiple imputation for interval censored data with auxiliary variables. *Statistics in Medicine*, 26 (4):769–781, 2007.

Jian Huang. Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics*, 24(2):540–568, 1996.

Jian Huang and AJ Rossini. Sieve estimation for the proportional-odds failure-time regression model with interval censoring. *Journal of the American Statistical Association*, 92 (439):960–967, 1997.

Martin Jacobsen and Niels Keiding. Coarsening at random in general sample spaces and random censoring in continuous time. *The Annals of Statistics*, 23(3):774–786, 1995.

Holly Janes, Francesca Dominici, and Scott Zeger. On quantifying the magnitude of confounding. *Biostatistics*, 11(3):572–582, 2010.

Nicholas P Jewell and Stephen C Shiboski. Statistical analysis of HIV infectivity based on partner studies. *Biometrics*, 46(4):1133–1150, 1990.

John D Kalbfleisch and Jerald F Lawless. Likelihood analysis of multi-state models for disease incidence and mortality. *Statistics in Medicine*, 7:149–160, 1988.

John D Kalbfleisch and Ross L Prentice. *The Statistical Analysis of Failure Time Data, 2nd Ed.* John Wiley & Sons, Inc., New York, 2002.

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

John P Klein and Per Kragh Andersen. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, 61(1):223–229, 2005.

John P Klein and Melvin L Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media, New York, 2003.

R. G. Langley, G. G. Krueger, and C. E. Griffiths. Psoriasis: epidemiology, clinical features, and quality of life. *Annals of the Rheumatic Diseases*, 64(suppl 2):ii18–ii23, 2005.

Jerald F Lawless. *Statistical Models and Methods for Lifetime Data, 2nd Ed.* John Wiley & Sons, Inc., Hoboken, New Jersey, 2003.

Jerald F Lawless and Min Zhan. Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canadian Journal of Statistics*, 26(4):549–565, 1998.

Kung-Yee Liang and Scott L Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

Yu Liang, Wenbin Lu, and Zhiliang Ying. Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics*, 65(2):377–384, 2009.

D. Y. Lin, David Oakes, and Zhiliang Ying. Additive hazards regression with current status data. *Biometrika*, 85(2):289–298, 1998.

DY Lin and Zhiliang Ying. Semiparametric analysis of the additive risk model. *Biometrika*, 81(1):61–71, 1994.

Haiqun Lin, Daniel O Scharfstein, and Robert A Rosenheck. Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):791–813, 2004.

Stuart R Lipsitz, Garrett M Fitzmaurice, Joseph G Ibrahim, Richard Gelber, and Steven Lipshultz. Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics*, 58(3):621–630, 2002.

Charles F Manski and Steven R Lerman. The estimation of choice probabilities from choice based samples. *Econometrica*, 45(8):1977–1988, 1977.

Torben Martinussen and Thomas H Scheike. Efficient estimation in additive hazards regression with current status data. *Biometrika*, 89(3):649–658, 2002.

Torben Martinussen and Stijn Vansteelandt. On collapsibility and confounding bias in Cox and Aalen regression models. *Lifetime Data Analysis*, 19(3):279–296, 2013.

Peter McCullagh. Quasi-likelihood functions. *The Annals of Statistics*, 11:59–67, 1983.

Philip Mease and Bernard S Goffe. Diagnosis and treatment of psoriatic arthritis. *Journal of the American Academy of Dermatology*, 52(1):1–19, 2005.

R. E. Miles. The complete amalgamation into blocks, by weighted means, of a finite set of real numbers. *Biometrika*, 46:317–327, 1959.

Hari Mukerjee. Monotone nonparametric regression. *The Annals of Statistics*, 16(2):741–750, 1988.

Elizbar A Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

Aidan G O'Keeffe, Brian DM Tom, and Vernon T Farewell. A case-study in the clinical epidemiology of psoriatic arthritis: multistate models and causal arguments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(5):675–699, 2011.

Wei Pan. A multiple imputation approach to Cox regression with interval-censored data. *Biometrics*, 56(1):199–203, 2000.

Yuhyun Park, Lu Tian, and LJ Wei. One-and two-sample nonparametric inference procedures in the presence of a mixture of independent and dependent censoring. *Biostatistics*, 7(2):252–267, 2006.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

Judea Pearl. Causality: models, reasoning, and inference. *Econometric Theory*, 19:675–685, 2003.

Eleanor M Pullenayegum and Brian M Feldman. Doubly robust estimation, optimally truncated inverse-intensity weighting and increment-based methods for the analysis of irregularly observed longitudinal data. *Statistics in Medicine*, 32(6):1054–1072, 2013.

Eleanor M Pullenayegum and Lily SH Lim. Longitudinal data subject to irregular observation: A review of methods with a focus on visit processes, assumptions, and study design. *Statistical Methods in Medical Research*, 0962280214536537, 2014.

Tim Robertson, F. T. Wright, Richard L Dykstra, and T. Robertson. *Order Restricted Statistical Inference*, volume 229. John Wiley & Sons, New York, 1988.

James M Robins. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proceedings of the Biopharmaceutical Section, American Statistical Association*, pages 24–33, 1993.

James M Robins. Association, causation, and marginal structural models. *Synthese*, 121 (1):151–179, 1999.

James M Robins and Dianne M Finkelstein. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*, 56(3):779–788, 2000.

James M Robins and Miguel A Hernán. Estimation of the causal effects of time-varying exposures. *Longitudinal Data Analysis*, pages 553–599, 2009.

James M Robins and Andrea Rotnitzky. Recovery of information and adjustment for dependent censoring using surrogate markers. In N Jewell, K Dietz, and V Farewell, editors, *AIDS Epidemiology: Methodological Issues*, pages 24–33. Boston: Birkhäuse, 1992.

James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994.

James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429):106–121, 1995.

James M Robins, Miguel A Hernán, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

Donald B Rubin. Inference and missing data (with discussion). *Biometrika*, 63(3):581–592, 1976.

Glen A Satten, Somnath Datta, and James Robins. Estimating the marginal survival function in the presence of time dependent covariates. *Statistics & Probability Letters*, 54(4):397–403, 2001.

Daniel O Scharfstein, Andrea Rotnitzky, and James M Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120, 1999.

Shaun Seaman and Andrew Copas. Doubly robust generalized estimating equations for longitudinal data. *Statistics in Medicine*, 28(6):937–955, 2009.

Stephen C Shiboski and Nicholas P Jewell. Statistical analysis of the time dependence of HIV infectivity based on partner study data. *Journal of the American Statistical Association*, 87(418):360–372, 1992.

Fotios Siannis. Applications of a parametric model for informative censoring. *Biometrics*, 60(3):704–714, 2004.

Fotios Siannis, John Copas, and Guobing Lu. Sensitivity analysis for informative censoring in parametric survival models. *Biostatistics*, 6(1):77–91, 2005.

Fotios Siannis, Vernon T Farewell, Richard J Cook, Catherine T Schentag, and Dafna D Gladman. Clinical and radiological damage in psoriatic arthritis. *Annals of the Rheumatic Diseases*, 65(4):478–481, 2006.

Yvonne H Sparling, Naji Younes, John M Lachin, and Oliver M Bautista. Parametric survival models for interval-censored data with time-dependent covariates. *Biostatistics*, 7(4):599–614, 2006.

Jianguo Sun. Regression analysis of interval-censored failure time data. *Statistics in Medicine*, 16(5):497–504, 1997.

Jianguo Sun. *The Statistical Analysis of Interval-Censored Failure Time Data*. Springer Science & Business Media, New York, 2006.

Jianguo Sun, Do-Hwan Park, Liuquan Sun, and Xingqiu Zhao. Semiparametric regression analysis of longitudinal data with informative observation times. *Journal of the American Statistical Association*, 100(471):882–889, 2005.

Jianguo Sun, Liuquan Sun, and Dandan Liu. Regression analysis of longitudinal data in the presence of informative observation and censoring times. *Journal of the American Statistical Association*, 102(480):1397–1406, 2007.

Liuquan Sun, Xinyuan Song, Jie Zhou, and Lei Liu. Joint analysis of longitudinal data with

informative observation times and a dependent terminal event. *Journal of the American Statistical Association*, 107(498):688–700, 2012.

Bruce W Turnbull. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(3):290–295, 1976.

Mark J van der Laan and Alan Hubbard. Estimation with interval censored data and covariates. *Lifetime Data Analysis*, 3(1):77–91, 1997.

Mark J van der Laan and James M Robins. Locally efficient estimation with current status data and time-dependent covariates. *Journal of the American Statistical Association*, 93 (442):693–701, 1998.

Lianming Wang, Jianguo Sun, and Xingwei Tong. Regression analysis of case II interval-censored failure time data with the additive hazards model. *Statistica Sinica*, 20(4):1709, 2010.

Weijing Wang and Martin T Wells. Nonparametric estimation of successive duration times under dependent censoring. *Biometrika*, 85(3):561–572, 1998.

Geoffrey S Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372, 1964.

Robert WM Wedderburn. Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika*, 61(3):439–447, 1974.

Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50 (1):1–25, 1982.

Margaret C Wu and Raymond J Carroll. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44(1):175–188, 1988.

Jun Xie and Chaofeng Liu. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Statistics in Medicine*, 24 (20):3089–3110, 2005.

Jessica G Young and Eric J Tchetgen Tchetgen. Simulation from a known Cox MSM using standard parametric models for the g-formula. *Statistics in Medicine*, 33(6):1001–1014, 2014.

Jessica G Young, Miguel A Hernán, Sally Picciotto, and James M Robins. Simulation from structural survival models under complex time-varying data structures. *JSM Proceedings, Section on Statistics in Epidemiology, Denver, CO: American Statistical Association*, 2008.

Jessica G Young, Miguel A Hernán, Sally Picciotto, and James M Robins. Relation between three classes of structural models for the effect of a time-varying exposure on survival. *Lifetime Data Analysis*, 16(1):71–84, 2010.

Donglin Zeng, Jianwen Cai, and Yu Shen. Semiparametric additive risks model for interval-censored data. *Statistica Sinica*, 16(1):287, 2006.

Jiameng Zhang and Daniel F Heitjan. A simple local sensitivity analysis tool for nonignorable coarsening: application to dependent censoring. *Biometrics*, 62(4):1260–1268, 2006.

Zhigang Zhang and Jianguo Sun. Interval censoring. *Statistical Methods in Medical Research*, 19(1):53–70, 2010.

Zhigang Zhang, Liuquan Sun, Xingqiu Zhao, and Jianguo Sun. Regression analysis of interval-censored failure time data with linear transformation models. *Canadian Journal of Statistics*, 33(1):61–70, 2005.

Zhigang Zhang, Liuquan Sun, Jianguo Sun, and Dianne M Finkelstein. Regression analysis of failure time data with informative interval censoring. *Statistics in Medicine*, 26(12): 2533–2546, 2007.

LP Zhao and S Lipsitz. Designs and analysis of two-stage studies. *Statistics in Medicine*, 11(6):769–782, 1992.

Shishun Zhao, Tao Hu, Ling Ma, Peijie Wang, and Jianguo Sun. Regression analysis of interval-censored failure time data with the additive hazards model in the presence of informative censoring. *Statistics and Its Interface*, 8(3):367–377, 2015.

Yichuan Zhao and Yu-Sheng Hsu. Semiparametric analysis for additive risk model via empirical likelihood. *Communications in Statistics, Simulation and Computation*, 34(1): 135–143, 2005.

Appendix A

A.1 List of Regressors in the Analyses of Clinic Visit Times and Treatment with Biologics

Table A.1: Descriptions and center values of the variables regressed in the analyses of Table 1.1 and Tables 5.1-5.7. Time-varying variables are measured only at visits, except treatment variables (NSAIDs, DMARDs, biologics) that can change at arbitrary times and whose full history is known. ESR denotes erythrocyte sedimentation rate (mm/hr); med.gap denotes median length of past visit gap times (in days).

Variable	Value/Level	Center	Fixed/Time-varying
sex	0: female, 1: male	—	fixed
year of enrolment	continuous	2000	fixed
year of the visit before t_{i0}	continuous	2000	fixed
family history of PS	0: No, 1: Yes	—	fixed at enrolment
family history of PsA	0: No, 1: Yes	—	fixed at enrolment
age (in years)	continuous	40	time-varying
PS duration (in years)	continuous	20	time-varying
PsA duration (in years)	continuous	5	time-varying
number of active joints	continuous	—	time-varying
number of damaged joints	continuous	—	time-varying
NSAIDs	0: No, 1: Yes	—	time-varying
DMARDs	0: No, 1: Yes	—	time-varying
biologics	0: No, 1: Yes	—	time-varying
ESR	continuous	20	time-varying
med.gap	continuous	180	time-varying