INTERVAL-CENSORED DATA WITH INACCURATE TIME BOUNDS AND TIME-DEPENDENT COVARIATES

Bing Chen

University of Rhode Island, bchen1@uri.edu

Follow this and additional works at: https://digitalcommons.uri.edu/theses

Recommended Citation

Chen, Bing. "INTERVAL-CENSORED DATA WITH INACCURATE TIME BOUNDS AND TIME-DEPENDENT COVARIATES" (2021). Open Access Master's Theses. Paper 1987.

https://digitalcommons.uri.edu/theses/1987

This Thesis is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Master's Theses by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.
INTERVAL-CENSORED DATA WITH INACCURATE TIME BOUNDS AND TIME-DEPENDENT COVARIATES

BY

BING CHEN

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
STATISTICS

UNIVERSITY OF RHODE ISLAND
2021
MASTER OF SCIENCE THESIS

OF

BING CHEN

APPROVED:

Thesis Committee:

Major Professor    Yichi Zhang
                  Jing Wu
                  Gavino Puggioni
                  Xiaowei Xu
                  Brenton DeBoef

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2021
ABSTRACT

A key characteristic that distinguishes survival analysis from other statistics fields is that survival data are usually censored or incomplete in some way. The event time is interval-censored when the exact event time is unknown and the event occurs within some interval of time.

The purposes of this study included: developed an easy-to-use code for interval-censored survival data with both fixed and time-dependent covariates; conducted extensive simulations to investigate the robustness of the interval-censored survival analysis with inaccurate time bounds and time-dependent covariates, particularly under noninformative censoring and informative censoring; conducted a real data analysis using the ACTG 175 data to investigate the robustness of findings under inaccurate time bounds.

The likelihood approach was used to draw inferences about the unknown parameters. The parameter estimates and standard errors were obtained; confidence intervals were constructed. The findings for simulations demonstrated that for both noninformative censoring and informative censoring, parameter estimates and coverage probability were more robust against deviations from the true time bounds for regression coefficients of the fixed and time-dependent covariates than for general hazard function parameters. The findings for real data analysis demonstrated that for both noninformative censoring and informative censoring, the estimates and $p$-values for the fixed and time-dependent covariates were robust against deviations from the true time bounds.
This study was the first to develop the code for interval-censored survival data with both fixed and time-dependent covariates, and was the first to investigate the robustness of the interval-censored survival analysis under inaccurate time bounds.
ACKNOWLEDGMENTS

I would like to first thank my major professor, Dr. Yichi Zhang, and my co-major professor, Dr. Jing Wu, for their invaluable advice and continuous support during my master’s study. They have helped me a lot with my thesis and other research projects. They always answered my questions quickly and provided me with helpful information. They also gave me good advice about my study and future career.

I also want to thank my other thesis core committee members, Dr. Gavino Puggioni and Dr. Xiaowei Xu. They have provided me with useful suggestions and ideas for my thesis. Moreover, I would like to thank Dr. Joseph Rossi for being my thesis defense chair and devoting time to attending my defense and reviewing my thesis. In addition, I would like to thank my classmates at the statistics master’s program of University of Rhode Island (URI). It is their kind help and friendship that have made my study and life at URI a wonderful time.

Finally, I want to express my gratitude to my parents. It would be impossible for me to complete my study without their encouragement and support.
PREFACE

This thesis was prepared in standard format.
## TABLE OF CONTENTS

ABSTRACT .................................................................................................................. ii  
ACKNOWLEDGMENTS .............................................................................................. iv  
PREFACE ..................................................................................................................... v  
TABLE OF CONTENTS ............................................................................................ vi  
LIST OF TABLES ...................................................................................................... vii  
LIST OF FIGURES .................................................................................................... ix  
CHAPTER 1 ............................................................................................................... 1  
INTRODUCTION ....................................................................................................... 1  
CHAPTER 2 ............................................................................................................... 12  
METHODOLOGY ..................................................................................................... 12  
CHAPTER 3 ............................................................................................................... 19  
FINDINGS .................................................................................................................. 19  
CHAPTER 4 ............................................................................................................... 44  
CONCLUSION .......................................................................................................... 44  
BIBLIOGRAPHY ...................................................................................................... 50
LIST OF TABLES

| TABLE | PAGE |
|-------|------|
| Table 1. Bias across 20 $M$ Values for Each Parameter for Noninformative Censoring of Scenario 2 | 26 |
| Table 2. Standard Error across 20 $M$ Values for Each Parameter for Noninformative Censoring of Scenario 2 | 27 |
| Table 3. Mean Squared Error across 20 $M$ Values for Each Parameter for Noninformative Censoring of Scenario 2 | 28 |
| Table 4. Coverage Probability of 95% Confidence Interval across 20 $M$ Values for Each Parameter for Noninformative Censoring of Scenario 2 | 29 |
| Table 5. Bias across 20 $M$ Values for Each Parameter for Informative Censoring of Scenario 2 | 30 |
| Table 6. Standard Error across 20 $M$ Values for Each Parameter for Informative Censoring of Scenario 2 | 31 |
| Table 7. Mean Squared Error across 20 $M$ Values for Each Parameter for Informative Censoring of Scenario 2 | 32 |
| Table 8. Coverage Probability of 95% Confidence Interval across 20 $M$ Values for Each Parameter for Informative Censoring of Scenario 2 | 33 |
| Table 9. Estimates, Standard Errors, Test Statistics, and P-values across 20 $M$ Values for Treat and CD4 Variables in ACTG 175 for Noninformative Censoring | 34 |
| Table 10. Estimates, Standard Errors, Test Statistics, and P-values across 20 $M$ Values for Treat and CD4 Variables in ACTG 175 for Informative Censoring | 35 |
| Table 11. Estimates and Standard Errors across 20 $M$ Values for Karnof, Symptom, |
Treat and CD4 Variables in ACTG 175 for Noninformative Censoring

Table 12. Test Statistics and P-values across 20 $M$ Values for Karnof, Symptom, Treat and CD4 Variables in ACTG 175 for Noninformative Censoring

Table 13. Estimates and Standard Errors across 20 $M$ Values for Karnof, Symptom, Treat and CD4 Variables in ACTG 175 for Informative Censoring

Table 14. Test Statistics and P-values across 20 $M$ Values for Karnof, Symptom, Treat and CD4 Variables in ACTG 175 for Informative Censoring
# LIST OF FIGURES

| FIGURE                                                                 | PAGE  |
|-----------------------------------------------------------------------|-------|
| Figure 1. Boxplot for Parameter Estimates of Each Parameter across 20 $M$ Values for Noninformative Censoring (a) vs Informative Censoring (b) for Scenario 2. | 40    |
| Figure 2. Coverage Probability of 95% Confidence Interval across 20 $M$ Values for Each Parameter for Noninformative Censoring (a) vs Informative Censoring (b) for Scenario 2. | 41    |
| Figure 3. Boxplot for Parameter Estimates of Each Parameter across 20 $M$ Values for Noninformative Censoring (a) vs Informative Censoring (b) for Scenario 1. | 42    |
| Figure 4. Coverage Probability of 95% Confidence Interval across 20 $M$ Values for Each Parameter for Noninformative Censoring (a) vs Informative Censoring (b) for Scenario 1. | 43    |
CHAPTER 1

INTRODUCTION

Survival Data Analysis

The statistical analysis of what are referred to as lifetime, survival time, or failure time data is an important topic in various fields and the applications of lifetime or survival time methodology range from examinations of the durability of manufactured items to studies and treatments of human diseases (Lawless, 2011).

Survival analysis is used in many areas for analyzing data involving the times of transition among several states or conditions (Leung, Elashoff, & Afifi, 1997). Survival analysis is also called lifetime data analysis, time to event analysis, reliability and event history analysis, depending on the focus and the type of application (Leung et al., 1997; Prinja, Gupta, & Verma, 2010).

In some instances, the events are actual deaths of individuals and “lifetime” referred to the length of life measured from some particular starting point. In other instances, terms such as “lifetime”, “death”, or “failure” denote the event of interest and are used in a figurative sense (Lawless, 2011).

A key characteristic that distinguishes survival analysis from other statistics fields is that survival data are usually censored or incomplete in some way (Leung et al., 1997). There are three types of censoring, left censoring, right censoring, and interval censoring.

Left Censoring, Right Censoring, & Interval Censoring
The event times are censored when they are not observed accurately (Sparling, Younes, Lachin, & Bautista, 2006). In clinical trials, censoring usually occurs when information on time to outcome event is unavailable for all participants, and observations are censored when information on time to event is unavailable due to loss to follow-up or non-occurrence of outcome event before the trial end (Prinja et al., 2010).

The event time is right-censored when follow-up is curtailed without observing the event (Sparling et al., 2006). In clinical trials, an observation is considered right-censored if the person was alive at study termination or was lost to follow-up at any time during the study (Leung et al., 1997).

The event time is left-censored when the event occurs at some unknown time prior to an individual’s inclusion in a cohort (Sparling et al., 2006). In clinical trials, an observation is considered left-censored if the person had been on risk for disease for a period before entering the study (Prinja et al., 2010).

The event time is interval-censored when the event occurs within some interval of time, but the exact event time is unknown (Sparling et al., 2006). Interval censoring can occur when observing a value requires follow-ups or inspections. Interval censoring occurs frequently in epidemiological, financial, sociological, and clinical (especially HIV and cancer) studies where the event of interest is known only to occur within an interval induced by periodic monitoring (Song & Ma, 2008; Zeng, Mao, & Lin, 2016). Practically, most observational studies dealing with non-lethal outcomes have periodic examination schedules and are interval-censored (Prinja et al., 2010; Zeng et al., 2016).
Right censoring (where the upper bound is infinity) and left censoring (where the lower bound is 0) can be considered as special cases of interval censoring. Therefore, the more general type of censoring (i.e., interval censoring) was focused in this project. This project conducted interval-censored survival analysis with time-dependent covariates.

**Time-dependent Covariates**

In many instances, factors or covariates affecting an individual’s lifetime may vary over time, and we refer to them as time-dependent covariates (Lawless, 2011). For example, in a study which investigated the age at which smokers develop chronic diseases, the type and level of smoking for each participant can vary over time and are considered as time-dependent covariates. A lifetime’s relationship to covariates cannot be discussed without considering the values the covariates take over time; a generally useful approach is to consider the hazard function at time $t$ conditional on previous covariate values (Lawless, 2011).

In this project, the interval-censored survival analysis with time-dependent covariates were conducted under two scenarios, noninformative censoring and informative censoring.

**Noninformative Censoring and Informative Censoring**

Most studies assumed that the examination time and lifetime of interest are completely independent (e.g., Groeneboom & Wellner, 1992) or conditionally independent given covariates (e.g., Rossini & Tsiatis, 1996). This is known as the assumption of noninformative censoring. Chen, Lu, Chen, and Hsu (2012) suggested that if the examination time is intrinsically related to the lifetime of interest, the
examination time is considered informative censoring. Such data may occur in many fields including epidemiological and medical studies.

Dettoni, Marra, and Radice (2020) mentioned that censoring is independent when the hazard rate of the event of interest for the censored observations is equal to the hazard rate for the uncensored observations, otherwise censoring is dependent; censoring is informative or dependent when the censoring times contain information on the parameters of the distribution of the event variable (e.g., Kalbfleisch & Prentice, 2002). Ranganathan and Pramesh (2012) stated that censoring in survival analysis should be noninformative, which means participants who drop out of the study should do so due to reasons unrelated to the study, while informative censoring occurs when participants are lost to follow up due to reasons related to the study. Although informative censoring has been well-studied in the survival analysis (e.g., Chen et al., 2012; Leung et al., 1997; Emura & Chen, 2018), the specific study which analyzes the problem of informative censoring is scarce, even though ignoring it may cause detrimental consequences on inferential conclusions (e.g., Lu & Zhang, 2012).

For this project, the analysis of interval-censored data was studied under both noninformative and informative censoring scenarios, using both simulations and the real data.

The Real Data - ACTG 175 Data

In real life, we may not be able to record the exact event time. For example, in the real data of the AIDS Clinical Trial Group (ACTG) 175 (e.g., Hammer et al., 1996), time to progression to AIDS may not be accurately observed. Some previous studies considered the data as right-censored (e.g., Hammer et al., 1996; Scharfstein &
Robins, 2002), while some studies considered the data as interval-censored (e.g., Song & Ma, 2008; Song & Wang, 2017; Wen, 2012; Wen & Chen, 2014). Several previous studies (e.g., Hammer et al., 1996; Huang & Zhang, 2008; Scharfstein & Robins, 2002; Song, Davidian, & Tsiatis, 2002; Song & Ma, 2008; Song & Wang, 2017; Wen, 2012; Wen & Chen, 2014) analyzing ACTG 175 data were discussed below.

Hammer et al. (1996) evaluated treatment with either a single nucleoside or two nucleosides in HIV-1 infected adults whose CD4 cell counts were between 200 and 500 per cubic millimeter. The ACTG 175, a randomized clinical trial that compares monotherapy with zidovudine or didanosine with combination therapy with zidovudine and didanosine or zidovudine and zalcitabine in HIV-1 infected adults with CD4 cell counts between 200 and 500 per cubic millimeter, was used in this study. The primary study end point was a decline in CD4 cell count of at least 50 percent, or an event indicating progression to AIDS or death. The variables involved in the analysis included gender, age, race, homosexuality, hemophilia, intravenous drug use, Karnofsky score, symptomatic HIV infection, previously received antiretroviral therapy, and CD4 cell count. Patients were examined at weeks 2, 4, and 8 and then every 12 weeks, with CD4 cell counts determined from week 8 onward. The results demonstrated that treatment with zidovudine alone was inferior to each of the other three treatments. Antiretroviral therapy can improve survival in patients with CD4 cell counts from 200 to 500 per cubic millimeter and such therapy can confer a survival benefit in a population with intermediate-stage disease.

Patients were examined at weeks 2, 4, and 8 and then every 12 weeks, with CD4 cell counts determined from week 8 onward. Since the examinations of CD4 cell
counts were conducted every 12 weeks and the progression to AIDS depends on the CD4 cell counts, the event time should be considered as interval-censored. As a result, the analysis using models for right-censored data on the ACTG 175 data should be viewed as an approximation, and the accuracy of such approximation needs to be evaluated. In addition, if the periodic examination schedules for patients were followed exactly, then the censoring should be noninformative; however, in practice there might be factors that affect the examination date and the event time simultaneously which will result in informative censoring.

Scharfstein and Robins (2002) also analyzed the ACTG 175 data as right-censored data and developed a method for estimating the survival curve of a continuous survival time random variable. The method allowed for adjustment of informative censoring due to measured prognostic factors for time-to-event and censoring. In the ACTG 175 data, baseline variables included gender, age, race, ethnicity, CD4 cell count, Karnofsky score, prior use of antiretroviral therapy, symptomatic HIV infection, homosexuality, hemophilia and intravenous drug use. The variables used in the analysis included CD4 count, age, intravenous drug use, Karnofsky score, and symptomatic HIV infection. One finding was that subjects who had lower time-dependent CD4, were younger, were intravenous drug users and had lower baseline Karnofsky scores were more likely to be censored.

Song et al. (2002) mentioned that time-dependent covariates are generally observed intermittently and are subject to error. They defined a joint longitudinal-survival model with multiple time-dependent covariates measured at potentially different time points whose observation is subject to error, and generalized the
conditional score approach which was for a single time-dependent covariate and required no assumption on the distribution of the random effects to multiple time-dependent covariates. Song et al. (2002) used this approach to handle mismeasured covariates. The approach was illustrated via simulations and demonstrated by application to the ACTG 175 data. The variables included in the ACTG 175 data analysis were treatment, CD4 count, and CD8 count. Application of the approach showed that it is accessible for routine use and leads to reliable inference.

Although both studies used the conditional score approach to handle mismeasured covariates, the study of Song et al. (2002) generalized the conditional score approach to multiple time-dependent covariates to handle mismeasured covariates, while the study of Song and Ma (2008) proposed a multiple data augmentation approach that can convert interval-censored data into right-censored data and then employed the conditional score approach, which was developed for analyzing right-censored data, to handle mismeasured covariates. The goal of the study of Song and Ma (2008) was to fill in the gap between methodologies for analyzing interval-censored data without mismeasured covariates and methodologies for analyzing right-censored data with mismeasured covariates. The performance of the multiple augmentation approach was assessed via simulation studies and the ACTG 175 data. The ACTG 175 data was analyzed as interval-censored data with mismeasured covariates. The variables used in the analysis included CD4 count and treatment. The findings showed that this approach had satisfactory empirical performance for interval-censored data. Song and Ma (2008) mentioned that this approach can be applied under other semiparametric model assumptions such as the
additive risk model, and that the conditional score approach can be replaced by other error-dealing approaches for right-censored data, such as correction approaches.

The study of Song and Wang (2017) also used the conditional score approach to handle mismeasured covariates. However, different from the study of Song et al. (2002) and the study of Song and Ma (2008), Song and Wang (2017) proposed a partially time-varying coefficient proportional hazards model to explore the associations between the hazard of failure and covariates (time-varying covariates and fixed covariates). The coefficients were estimated using a polynomial spline approach, and the corrected score and conditional score approaches were used to deal with mismeasured covariates. The proposed model was assessed by simulations and applied to the ACTG 175 data to explore the temporal dynamics of the effect of treatment and CD4 counts on time to AIDS or death. The variables included in the model were treatment and CD4 count. Although both corrected score and conditional score approaches were asymptotically equivalent, the conditional score approach was recommended for its better finite sample performance. The proposed approach will have applicability in a wide range of studies for dealing with survival data with longitudinal mismeasured covariates. The proportional hazards model may be generalized to include functions of the random effects by analogy to Song et al. (2002).

Similar to the study of Song and Ma (2008), which analyzed general interval-censored data with mismeasured covariates, the study of Wen (2012) also analyzed general interval-censored data with mismeasured covariates. However, while Song and Ma (2008) converted interval-censored data into right-censored data using multiple
augmentation approach and employed the conditional score approach to handle mismeasured covariates, Wen (2012) described a semiparametric maximum likelihood (SPML) method for analyzing general interval-censored data with mismeasured covariates under the proportional hazards model, and proposed a stable and efficient computation algorithm for computing the SPML estimators. The proposed method was illustrated via simulations and with the ACTG 175 data. The primary goal of the ACTG 175 data analysis was to address the effects of baseline CD4 count and treatment on the time to AIDS or death in antiretroviral-naive patients. The covariates included in the analysis were CD4 count and treatment. The proposed approach can be applied to other semiparametric models, such as the proportional odds model.

Although the studies of Song et al. (2002), Song and Ma (2008), Song and Wang (2017), and Wen (2012) also handled mismeasured covariates, the study of Wen and Chen (2014) was different. Wen and Chen (2014) proposed a functional inference method under the semiparametric proportional odds model for interval-censored survival data with mismeasured covariates. The approach used the working independence strategy to handle general interval-censored data and used the conditional score approach to handle mismeasured covariates. The proposed working independence conditional score approach was illustrated via simulations and the ACTG 175 data. The variables used in the analysis of ACTG 175 data included CD4 count and treatment. Wen and Chen (2014) suggested some future directions such as extending the proposed approach to more general regression models than the proportional odds model and developing a functional modeling approach assumption for the examination times on the same subject.
Another study analyzing the ACTG 175 data (Huang & Zhang, 2008) proposed an estimation method for the bivariate proportional hazards model for competing risks. The copula approach was used to do sensitivity analysis for the Cox proportional hazards models. The joint modeling was used to address informative censoring. By this approach, the joint distribution of the survival time and censoring time is assumed to be a function of their marginal distributions. Huang and Zhang (2008) used this approach to assess the effect of informative censoring on the parameter estimates. The proposed approach was assessed by simulation studies and applied to the ACTG 175 data. In the ACTG 175 data analysis, covariates were selected using a stepwise selection algorithm. A total of nine variables were selected for the analysis, including treatment, hemophilia, symptomatic HIV infection, gender, intravenous drug use, age, race, prior antiretroviral therapy, and CD4 count.

The ACTG 175 data was also analyzed in the present study and the robustness of findings was investigated. The details of the present study were described below.

**Present Study**

In real life, we may not be able to record the exact event time or the accurate lower bound and upper bound for interval-censored survival data. For example, in the ACTG 175 data, the examinations of CD4 cell counts may not be conducted exactly every 12 weeks, and the time bounds may be affected by unobservable event time, which may result in inaccurate time bounds. If this is the case, the estimates as well as standard errors may not be calculated consistently, which may affect the accuracy of the findings. In this project, the impact of inaccurate time bounds on the analysis of interval-censored data was studied under both noninformative and informative
censoring scenarios, using simulations and the real data. The purposes of the present study included:

- Developed an easy-to-use code for interval-censored survival data with both fixed and time-dependent covariates.
- Conducted extensive simulations to investigate the robustness of the interval-censored survival analysis with inaccurate time bounds and time-dependent covariates, particularly under noninformative censoring and informative censoring.
- Conducted a real data analysis using the ACTG 175 data to investigate the robustness of findings under inaccurate time bounds.

This study was the first to develop the code for interval-censored survival data with both fixed and time-dependent covariates, and was the first to investigate the robustness of the interval-censored survival analysis under inaccurate time bounds. This study contributed to the research literature about the topic and contributed to the methodology, procedures, and analysis of future survival studies examining the similar topic.
CHAPTER 2

METHODOLOGY

Data Setting

The data consists of independently and identically distributed (i.i.d.) measurements from \( n \) subjects. Each subject has a sequence of visit times, a vector of fixed covariates, a vector of time-dependent covariates, and an event time.

Specifically, for subject \( i = 1, \ldots, n \), let \( \tau_{i,1} < \cdots < \tau_{i,J_i} \) be a random sequence of visit times, where \( J_i \) denotes the total number of visits for the \( i \)th subject. For convenience, let \( \tau_{i,J_i+1} = \infty \). The first visit is at \( \tau_{i,1} = 0 \), the beginning of the study.

Let \( Z_i = (Z_{i,1}, \ldots, Z_{i,p})^T \) be a vector of \( p \) fixed covariates for the \( i \)th subject measured at the first visit. Let \( Y_{i,j} = (Y_{i,j,1}, \ldots, Y_{i,j,q})^T \) be a vector of \( q \) time-dependent covariates measured at the \( j \)th visit. The value of each time-dependent covariate is assumed to be constant within each interval \( (\tau_{i,j}, \tau_{i,j+1}] \), for \( j = 1, \ldots, J_i \). So, the vector of time-dependent covariates at time \( t > 0 \) can be written as

\[
Y_i(t) = \sum_{j=1}^{J_i} Y_{i,j}I(\tau_{i,j} < t \leq \tau_{i,j+1}).
\]

The event time is denoted as \( X_i \). The complete data for the \( i \)th subject consists of \( J_i, \tau_{i,j}, Z_i, Y_i(t) \), and \( X_i \). We may fail to observe the complete data due to the presence of censoring. The observed data for the \( i \)th subject consists of \( Z_i, Y_i(t) \), the event indicator \( \delta_i \), and the time interval \( (L_i, U_i] \). If the exact event time is observed,
then $\delta_i = 1$ and $L_i = U_i = X_i$; otherwise, $\delta_i = 0$, and $L_i$ and $U_i$ are the two adjacent visit times that have the event time $X_i$ in between, and can be written as Equation 1:

$$L_i = \sum_{j=1}^{I_i} \tau_{i,j} I(\tau_{i,j} < X_i \leq \tau_{i,j+1})$$

and Equation 2:

$$U_i = \sum_{j=1}^{I_i} \tau_{i,j+1} I(\tau_{i,j} < X_i \leq \tau_{i,j+1}).$$

This formula allows for left-, right-, and interval-censored observations. The observation is left-censored if $L_i = 0$; the observation is right-censored if $U_i = \infty$; the observation is interval-censored if $L_i > 0$ and $U_i < \infty$.

Let $\gamma$ and $\eta$ be the coefficient vectors for the fixed and time-dependent covariates, $\theta$ be the intercept; $\gamma$ is $p$-dimensional, $\eta$ is $q$-dimensional. With time-dependent covariates, the hazard function can be expressed as

$$\lambda_i(t; \omega) = \frac{\alpha e^{\theta + \gamma^T Z_i + \eta^T Y_i(t)}}{1 + e^{\theta + \gamma^T Z_i + \eta^T Y_i(t)}}^{\alpha - 1}$$

where $\alpha > 0$ and $\kappa$ are general hazard function parameters; the parameter vector $\omega = (\theta, \gamma, \kappa, \alpha, \eta)^T$. Specific values for $\alpha$ and $\kappa$ yield a specific distribution (Sparling et al, 2006). For instance, $\kappa = 0$ yields a Weibull hazard. With no time-dependent covariates, the hazard function can be simplified as

$$\lambda_i(t; \omega) = \frac{\alpha e^{(\theta + \gamma^T Z_i) t^{\alpha - 1}}}{1 + e^{(\theta + \gamma^T Z_i) t^{\alpha}}}.$$
\[ \Lambda_i(t; \omega) = \int_0^t \lambda_i(s; \omega) \, ds, \]

the survival function is

\[ S_i(t; \omega) = e^{-\Lambda_i(t; \omega)}, \]

the probability density function is

\[ f_i(t; \omega) = \lambda_i(t; \omega) e^{-\Lambda_i(t; \omega)}, \]

and the cumulative distribution function is

\[ F_i(t; \omega) = 1 - e^{-\Lambda_i(t; \omega)}. \]

**Model**

The likelihood approach was used to draw inferences about the unknown parameters. The log-likelihood function can be expressed as

\[ l(\omega) = \sum_{i=1}^n \left\{ \delta_i \log f_i(L_i; \omega) + (1 - \delta_i) \log [F_i(U_i; \omega) - F_i(L_i; \omega)] \right\}, \]

where the parameter vector \( \omega = (\theta, \gamma, \kappa, \alpha, \eta)^T \). The estimator, \( \hat{\omega} \), was the maximizer of the log-likelihood function and was asymptotically normal. The R function *optim* with BFGS method (Nocedal & Wright, 1999) was used to calculate the maximizer.

The R package, *numDeriv*, was used to calculate the Hessian matrix of the negative log-likelihood. The estimated variance was the diagonal of the inverse of the Hessian matrix and the standard errors of the estimates equaled the square root of the estimated variance. The 95% confidence intervals were constructed using estimates plus and minus 1.96 times standard errors.

**Development of an Easy-to-Use Implementation**

Existing software provide limited support in fitting regression models with interval-censored survival data and time-dependent covariates. For example, the
survreg function in R can be used for left, right, or interval-censored data, but it can only be applied to fixed covariates; the coxph function can be used for both fixed and time-dependent covariates, but it can only be used for right-censored data. To facilitate the study of the impact of inaccurate time bounds, I developed my own code for interval-censored survival data which included both fixed and time-dependent covariates.

**Simulation Studies**

I conducted simulations to see estimate difference when time bounds of the intervals were not exactly observed or accurately recorded. In the simulated data, the lower bound of the interval is $L_i$, and the upper bound of the interval is $U_i$. The interval-censored variable is $1 - \delta_i$. The time-independent variables are $Z_i$. The time-dependent variables are $Y_i(t)$.

In the first scenario, 500 simulated datasets were generated. Each simulated dataset consisted of 1000 i.i.d. observations and included only fixed covariates, $Z_{i,1}$ and $Z_{i,2}$. The covariate $Z_{i,1}$ was a normal random variable with mean 0 and variance 1 and $Z_{i,2}$ was a binomial random variable with probability of 0.5. The lower bound of the interval $L_i$ as well as the upper bound of the interval $U_i$ were elements of the vector of $\tau_{i,j}$, which was defined as [0, 10] evenly visit 10000 times. $L_i$ and $U_i$ were obtained using Equation 1 and 2, respectively. The event time, $X_i$, was $S_i^{-1}(Q_i)$, where $Q_i$ was i.i.d. uniform (0, 1). All the observations were interval-censored. The true values for the five parameters, $\theta, \gamma_1, \gamma_2, \kappa, \alpha$, were 0, 0.8, 0.8, 0, and 2, respectively.
In the second scenario, 500 simulated datasets were generated. Each simulated
dataset consisted of 1000 i.i.d. observations and included both fixed covariates, $Z_{i,1}$
and $Z_{i,2}$, and time-dependent covariate, $Y_i(t)$. The fixed covariate $Z_{i,1}$ was a normal
random variable with mean 0 and variance 1 and $Z_{i,2}$ was a binomial random variable
with probability of 0.5. The time-dependent covariate $Y_i(t) = I(t > W_i)$, where $W_i$
was i.i.d. uniform (0, 2). The lower bound of the interval $L_i$ as well as the upper bound
of the interval $U_i$ were elements of the vector of $\tau_{i,j}$, which was defined as [0, 13]
evenly visit 10000 times. $L_i$ and $U_i$ were obtained using Equation 1 and 2,
respectively. The event time, $X_i$, was $S_i^{-1}(Q_i)$, where $Q_i$ was i.i.d. uniform (0, 1). All
the observations were interval-censored. The true values for the six parameters,
$\theta, \gamma_1, \gamma_2, \kappa, \alpha, \eta$, were 0, 0.8, 0.8, 0, 2, and $-0.5$, respectively.

In this study, extensive simulations were conducted to investigate the
robustness of the interval-censored survival analysis with inaccurate time bounds and
time-dependent covariates, particularly under the following two scenarios.
1) noninformative censoring:

The new lower bound and upper bound of interval were defined as follows:

new $L_i = L_i K_{L_i}$

new $U_i = U_i K_{U_i}$

where $K_{L_i}$ was drawn from a uniform distribution on $[1/M, 1]$ and $K_{U_i}$ was
drawn from a uniform distribution on $[1, M]$ and $M \geq 1$ was a tuning parameter which
was a sequence of 20 numbers between 1 and 10.

2) informative censoring:

The new lower bound and upper bound of interval were defined as follows:
new $L_i = L_i K_{L_i}$

new $U_i = U_i K_{U_i}$

where $K_{L_i}$ was drawn from a uniform distribution on $[1/M_{L_i}, 1]$ and $K_{U_i}$ was drawn from a uniform distribution on $[1, M_{U_i}]$; $M_{L_i} \geq 1$ as well as $M_{U_i} \geq 1$ was a tuning parameter that was a function of $L_i$ and $U_i$. For example, $M_{L_i} = 1 + (M - 1) \frac{L_i}{\text{mean}(L_i)}$, $M_{U_i} = 1 + (M - 1) \frac{U_i}{\text{mean}(U_i)}$, and $M$ was a tuning parameter as in the noninformative censoring scenario.

The method described in Chapter 2 was applied to obtain the parameter estimates. Then the bias, the standard error, the mean squared error (i.e., the average squared difference between the estimated values and the true value), the coverage probability (i.e., the probability that a procedure for constructing random regions will produce an interval containing the true value) were calculated for each parameter. The programming language was R.

**Real Data Application**

The ACTG 175 (e.g., Hammer et al., 1996) was a randomized clinical trial to compare monotherapy with zidovudine or didanosine with combination therapy with zidovudine and didanosine or zidovudine and zalcitabine in HIV-1 infected adults with CD4 cell counts between 200 and 500 per cubic millimeter. The primary study endpoint was a decline in CD4 cell count of at least 50 percent, or an event indicating progression to AIDS or death. Patients were examined at weeks 2, 4, and 8 and then every 12 weeks, with CD4 cell counts determined from week 8 onward. This dataset had 2139 observations. The data included the following time-dependent covariate, $cd4$ (CD4 T cell count at baseline and at $20 \pm 5$ weeks), and some fixed covariates, such
as *treat* (treatment), *karnof* (Karnofsky score), *symptom* (symptomatic HIV infection), etc.

The variable *days* referred to the number of days until the first occurrence of a decline in CD4 T cell count of at least 50 percent, or an event indicating progression to AIDS, or death. According to the study of Hammer et al. (1996), the examinations of CD4 cell counts were conducted every 12 weeks. Thus, when interval-censored the upper bound of the interval was the variable *days* and the lower bound of the interval was *days* − 12 × 7; if *days* − 12 × 7 < 0, then the lower bound of the interval was set to be 0. When right-censored the lower bound of the interval was *days* and the upper bound of the interval was infinite.

The method described in Chapter 2 was applied to the ACTG 175 data and the parameter estimates, standard errors, and *p*-values were obtained.

For real data analysis, the robustness of findings (estimates, standard errors, and *p*-values) under different variations of time bounds were investigated. *M* was an upper bound for the errors in the time bounds which was a sequence of 20 numbers between 1 and 10, so we investigated if different choices of *M* might affect the findings.
CHAPTER 3

FINDINGS

Simulations

In the simulations, bias, standard error, mean squared error, and coverage probability of 95% confidence interval for each parameter were reported for each of the 20 $M$ values for the simulated dataset including both fixed and time-dependent covariates (i.e., scenario 2). Box plots were drawn for parameter estimates of each parameter across 20 $M$ values for noninformative censoring and informative censoring of scenario 2 (see Figure 1). Scatter plots were drawn for coverage probability of 95% confidence interval across 20 $M$ values for each parameter for noninformative censoring and informative censoring of scenario 2 (see Figure 2).

Figure 1 showed that for both noninformative censoring and informative censoring, parameter estimates for $\gamma_1, \gamma_2, \text{and } \eta$ were more robust against change across 20 $M$ values, but parameter estimates for $\theta, \kappa,$ and $\alpha$ were not that robust against change across 20 $M$ values.

The average bias across 500 data sets for each parameter across 20 $M$ values for scenario 2 were demonstrated in Table 1 for noninformative censoring and in Table 5 for informative censoring. We can see from Table 1 that for noninformative censoring, bias was more robust against change across 20 $M$ values for $\gamma_1, \gamma_2,$ and $\eta$ than for $\theta, \kappa,$ and $\alpha$. We can see from Table 5 that for informative censoring, bias was more robust against change across 20 $M$ values for $\gamma_1, \gamma_2,$ and $\eta$ than for $\theta, \kappa,$ and $\alpha$. 

19
The average standard error across 500 data sets for each parameter across 20 $M$ values for scenario 2 were demonstrated in Table 2 for noninformative censoring and in Table 6 for informative censoring. We can see from Table 2 and Table 6 that standard error slightly increased across $M$ values for each parameter, for both noninformative censoring and informative censoring.

The average mean squared error across 500 data sets for each parameter across 20 $M$ values for scenario 2 were demonstrated in Table 3 for noninformative censoring and in Table 7 for informative censoring. We can see from Table 3 that for noninformative censoring, mean squared error was more robust against change across 20 $M$ values for $\gamma_1$, $\gamma_2$, and $\eta$ than for $\theta$, $\kappa$, and $\alpha$. We can see from Table 7 that for informative censoring, mean squared error was more robust against change across 20 $M$ values for $\gamma_1$, $\gamma_2$, and $\eta$ than for $\theta$, $\kappa$, and $\alpha$.

The coverage probability of 95% confidence interval across 20 $M$ values for each parameter for scenario 2 were demonstrated in Table 4 for noninformative censoring and in Table 8 for informative censoring. Scatter plots were also drawn for coverage probability for noninformative censoring and informative censoring (see Figure 2). We can see from Table 4 and Figure 2 (a) that for noninformative censoring, coverage probability was more robust against change across 20 $M$ values for $\gamma_1$, $\gamma_2$, and $\eta$ than for $\theta$, $\kappa$, and $\alpha$. We can see from Table 8 and Figure 2 (b) that for informative censoring, coverage probability was more robust against change across 20 $M$ values for $\gamma_1$, $\gamma_2$, and $\eta$ than for $\theta$, $\kappa$, and $\alpha$.

Additionally, box plots were drawn for parameter estimates of each parameter across 20 $M$ values for noninformative censoring and informative censoring for the
simulated dataset including only fixed covariates (i.e., scenario 1). See Figure 3 for box plots. Scatter plots were drawn for coverage probability of 95% confidence interval across 20 $M$ values for each parameter for noninformative censoring and informative censoring of scenario 1 (see Figure 4).

Figure 3 and Figure 4 demonstrated that for noninformative censoring, parameter estimates and coverage probability for $\gamma_1$ and $\gamma_2$ were more robust against change across 20 $M$ values, but parameter estimates and coverage probability for $\theta$, $\kappa$, and $\alpha$ were not that robust against change across 20 $M$ values; for informative censoring, parameter estimates and coverage probability for $\gamma_1$ and $\gamma_2$ were more robust against change across 20 $M$ values, but parameter estimates and coverage probability for $\theta$, $\kappa$, and $\alpha$ were not robust against change across 20 $M$ values.

To summarize, the findings showed that for noninformative censoring of scenario 2, parameter estimates as well as coverage probability were more robust against change across $M$ values for $\gamma_1$, $\gamma_2$, and $\eta$ than for $\theta$, $\kappa$, and $\alpha$, which were not that robust against change across $M$ values; for informative censoring of scenario 2, parameter estimates as well as coverage probability were more robust against change across $M$ values for $\gamma_1$, $\gamma_2$, and $\eta$ than for $\theta$, $\kappa$, and $\alpha$, which were not robust against change across $M$ values. Moreover, for noninformative censoring of scenario 1, parameter estimates as well as coverage probability were more robust against change across $M$ values for $\gamma_1$ and $\gamma_2$ than for $\theta$, $\kappa$, and $\alpha$, which were not that robust against change across $M$ values; for informative censoring of scenario 1, parameter estimates as well as coverage probability were more robust against change across $M$ values for $\gamma_1$ and $\gamma_2$ than for $\theta$, $\kappa$, and $\alpha$, which were not robust against change across $M$ values.
In both scenario 2 and scenario 1, the parameter estimates and coverage probability for 
\( \alpha \) were more robust against change across \( M \) values for noninformative censuring than 
for informative censoring.

In the scenarios considered in this simulation study, noninformative censoring 
is almost irrelevant to survival time, so as \( M \) value changed, for noninformative 
censoring, both scale and shape parameters (\( \theta \), \( \kappa \), and \( \alpha \)) changed but were not as 
severe as that for informative censoring. Informative censoring is related to survival 
time, as \( M \) value changed, the new lower bound and upper bound of interval had 
relationship with the original lower bound and upper bound of interval, so both scale 
and shape changed for informative censoring. Since \( \theta \) affects scale, \( \kappa \) affects scale and 
shape, \( \alpha \) affects shape, thus for noninformative censoring (both scale and shape 
changed as \( M \) changed but were not as severe as that for informative censoring), the 
parameter estimates and coverage probability for \( \theta \), \( \kappa \), and \( \alpha \) were not very robust 
against change across \( M \) values, but for \( \alpha \) were kind of more robust against change 
across \( M \) values than for informative censoring. For informative censoring (both scale 
and shape changed as \( M \) changed), parameter estimates and coverage probability for \( \theta \), 
\( \kappa \), and \( \alpha \) were not robust against change across \( M \) values. In addition, noninformative 
censoring has no relationship with fixed and time-dependent covariates, thus for 
noninformative censoring, the parameter estimates as well as coverage probability for 
\( \gamma_1 \), \( \gamma_2 \), and \( \eta \) were robust against change across \( M \) values. The informative censoring 
scenario considered in this study has no strong relationship with fixed covariates and 
time-dependent covariates, thus for informative censoring, parameter estimates and
coverage probability for $\gamma_1$, $\gamma_2$, and $\eta$ were also robust against change across $M$ values.

**Real Data Analysis**

Some previous studies analyzing ACTG 175 data only used treatment and cd4 counts variables in the analysis (e.g., Song & Ma, 2008; Song & Wang, 2017; Wen, 2012; Wen & Chen, 2014), so the real data analysis in this study first used $treat$ as fixed covariate and $cd4$ as time-dependent covariate. The estimates, standard errors, and $p$-values under different variations of time bounds were obtained for the fixed covariate, $treat$, and the time-dependent covariate, $cd4$.

The estimates, standard errors, test statistics (estimates/ standard errors), and $p$-values across 20 $M$ values for the $treat$ and $cd4$ variables in ACTG 175 were demonstrated in Table 9 for noninformative censoring and in Table 10 for informative censoring. We can see from Table 9 that for noninformative censoring, the estimates for $treat$ and $cd4$ were relatively robust against change across $M$ values; the $p$-values for $treat$ and $cd4$ were also relatively robust against change across $M$ values. We can see from Table 10 that for informative censoring, the estimates for $treat$ and $cd4$ were very robust against change across $M$ values; the $p$-values for $treat$ and $cd4$ were also very robust against change across $M$ values. The noninformative censoring has no relationship with fixed and time-dependent covariates, thus for noninformative censoring, the estimates for $treat$ and $cd4$ were robust against change across $M$ values. The informative censoring scenario considered in this study has no strong relationship with fixed covariates and time-dependent covariates, thus for informative censoring, the estimates for $treat$ and $cd4$ were also robust against change across $M$ values. The
estimates for treat and cd4 were all negative for noninformative and informative censoring and moved towards zero. Moreover, the p-values for treat and cd4 were all significant across M values for noninformative and informative censoring. The negative and significant estimates indicated longer survival time and event occurred later. Besides, the estimates under informative censoring were more robust against change across M values, compared to the estimates under noninformative censoring, which was worth further investigation.

In addition to the treatment and cd4 variables, some previous studies analyzing ACTG 175 data also included some other variables in the analysis, such as Karnofsky score and symptomatic HIV infection (e.g., Hammer et al., 1996; Scharfstein & Robins, 2002), so the real data analysis in this study also included some more covariates, karnof and symptom, to see if there would be any differences in the robustness of findings. The estimates, standard errors, and p-values under different variations of time bounds were obtained for the fixed covariates, treat, karnof, and symptom, and the time-dependent covariate, cd4. The findings for this analysis when more covariates were included were compared with the findings for the analysis when only treat and cd4 were used.

The estimates, standard errors, test statistics (estimates/standard errors), and p-values across 20 M values for the treat, karnof, symptom and cd4 variables in ACTG 175 were demonstrated in Table 11 and Table 12 for noninformative censoring and in Table 13 and Table 14 for informative censoring. We can see from Table 11 and 12 that for noninformative censoring, the estimates as well as p-values for treat and cd4 were relatively robust against change across M values, and the values and range were
similar to those in the analysis when only treat and cd4 were used. For example, when only treat and cd4 were used, the range of estimates for treat were around (-0.334, -0.042); when more covariates were included, the range of estimates for treat were around (-0.349, -0.057); the p-values for treat and cd4 in both analyses were all less than 0.05 and significant. We can see from Table 13 and 14 that for informative censoring, the estimates as well as p-values for treat and cd4 were very robust against change across M values, and were similar to those in the analysis when only treat and cd4 were used. For example, when only treat and cd4 were used, the range of estimates for treat were around (-0.334, -0.282); when more covariates were included, the range of estimates for treat were around (-0.349, -0.298); the p-values for treat and cd4 in both analyses were all less than 0.0001 and significant. The estimates for treat and cd4 were all negative for noninformative and informative censoring. Moreover, the p-values for treat and cd4 were all significant across M values for noninformative and informative censoring. The negative estimates indicated longer survival time and event occurred later. The estimates for symptom were all positive for noninformative and informative censoring, which indicated shorter survival time with symptomatic HIV infection. Still, the estimates under informative censoring were more robust against change across M values, compared to the estimates under noninformative censoring, which was worth further investigation.
Table 1

*Bias across 20 M Values for Each Parameter for Noninformative Censoring of Scenario 2*

| M   | $\theta$ | $\kappa$ | $\alpha$ | $\gamma_1$ | $\gamma_2$ | $\eta$ |
|-----|----------|----------|----------|------------|------------|--------|
| 1.00| -0.0072  | -0.0213  | 0.0043   | 0.0026     | 0.0021     | 0.0006 |
| 1.47| -0.0498  | -0.0847  | 0.0042   | 0.0016     | 0.0005     | -0.0023|
| 1.95| -0.1099  | -0.1420  | 0.0202   | 0.0076     | 0.0047     | -0.0036|
| 2.42| -0.1755  | -0.1942  | 0.0402   | 0.0153     | 0.0158     | -0.0057|
| 2.89| -0.2400  | -0.2433  | 0.0602   | 0.0227     | 0.0210     | -0.0026|
| 3.37| -0.3083  | -0.2945  | 0.0699   | 0.0258     | 0.0274     | 0.0012 |
| 3.84| -0.3772  | -0.3717  | 0.0767   | 0.0277     | 0.0296     | -0.0002|
| 4.32| -0.4361  | -0.4273  | 0.0916   | 0.0329     | 0.0321     | 0.0020 |
| 4.79| -0.5072  | -0.5267  | 0.0884   | 0.0317     | 0.0344     | 0.0059 |
| 5.26| -0.5558  | -0.5538  | 0.0994   | 0.0374     | 0.0365     | 0.0100 |
| 5.74| -0.6248  | -0.6685  | 0.0951   | 0.0356     | 0.0336     | 0.0182 |
| 6.21| -0.6658  | -0.6988  | 0.1125   | 0.0441     | 0.0433     | 0.0088 |
| 6.68| -0.7418  | -0.8837  | 0.0953   | 0.0339     | 0.0400     | 0.0163 |
| 7.16| -0.7838  | -0.9050  | 0.1086   | 0.0421     | 0.0423     | 0.0198 |
| 7.63| -0.8210  | -0.9772  | 0.1215   | 0.0467     | 0.0455     | 0.0155 |
| 8.11| -0.8748  | -1.0894  | 0.1171   | 0.0441     | 0.0436     | 0.0164 |
| 8.58| -0.9234  | -1.1863  | 0.1214   | 0.0461     | 0.0502     | 0.0269 |
| 9.05| -0.9628  | -1.2892  | 0.1242   | 0.0472     | 0.0424     | 0.0260 |
| 9.53| -1.0103  | -1.3817  | 0.1204   | 0.0445     | 0.0481     | 0.0326 |
| 10.00| -1.0486 | -1.5003  | 0.1092   | 0.0390     | 0.0439     | 0.0249 |
Table 2

*Standard Error across 20 M Values for Each Parameter for Noninformative Censoring of Scenario 2*

| M   | θ   | κ   | α   | γ₁  | γ₂  | η   |
|-----|-----|-----|-----|-----|-----|-----|
| 1.00| 0.0955 | 0.1782 | 0.1101 | 0.0549 | 0.0776 | 0.0799 |
| 1.47| 0.1031 | 0.2083 | 0.1173 | 0.0578 | 0.0805 | 0.0858 |
| 1.95| 0.1127 | 0.2453 | 0.1277 | 0.0622 | 0.0857 | 0.0933 |
| 2.42| 0.1213 | 0.2811 | 0.1379 | 0.0667 | 0.0908 | 0.1001 |
| 2.89| 0.1286 | 0.3161 | 0.1475 | 0.0707 | 0.0958 | 0.1063 |
| 3.37| 0.1346 | 0.3494 | 0.1556 | 0.0742 | 0.1000 | 0.1112 |
| 3.84| 0.1412 | 0.3933 | 0.1640 | 0.0775 | 0.1039 | 0.1155 |
| 4.32| 0.1464 | 0.4299 | 0.1719 | 0.0809 | 0.1076 | 0.1195 |
| 4.79| 0.1519 | 0.4813 | 0.1791 | 0.0838 | 0.1107 | 0.1223 |
| 5.26| 0.1546 | 0.5045 | 0.1847 | 0.0864 | 0.1140 | 0.1260 |
| 5.74| 0.1600 | 0.5673 | 0.1919 | 0.0892 | 0.1166 | 0.1280 |
| 6.21| 0.1631 | 0.5951 | 0.1980 | 0.0920 | 0.1199 | 0.1314 |
| 6.68| 0.1702 | 0.6941 | 0.2052 | 0.0942 | 0.1219 | 0.1326 |
| 7.16| 0.1726 | 0.7152 | 0.2112 | 0.0970 | 0.1247 | 0.1356 |
| 7.63| 0.1760 | 0.7744 | 0.2175 | 0.0994 | 0.1280 | 0.1384 |
| 8.11| 0.1804 | 0.8452 | 0.2238 | 0.1019 | 0.1303 | 0.1401 |
| 8.58| 0.1843 | 0.9123 | 0.2295 | 0.1043 | 0.1330 | 0.1417 |
| 9.05| 0.1887 | 0.9863 | 0.2355 | 0.1066 | 0.1348 | 0.1434 |
| 9.53| 0.1919 | 1.0552 | 0.2403 | 0.1085 | 0.1375 | 0.1455 |
| 10.00| 0.1973 | 1.1346 | 0.2459 | 0.1106 | 0.1390 | 0.1465 |
Table 3

*Mean Squared Error across 20 M Values for Each Parameter for Noninformative Censoring of Scenario 2*

| M   | Parameter | \(\theta\) | \(\kappa\) | \(\alpha\) | \(\gamma_1\) | \(\gamma_2\) | \(\eta\) |
|-----|-----------|-------------|-------------|-------------|---------------|---------------|----------|
| 1.00|           | 0.0081      | 0.0288      | 0.0106      | 0.0029        | 0.0059        | 0.0067   |
| 1.47|           | 0.0122      | 0.0482      | 0.0123      | 0.0033        | 0.0062        | 0.0077   |
| 1.95|           | 0.0234      | 0.0772      | 0.0149      | 0.0040        | 0.0072        | 0.0096   |
| 2.42|           | 0.0445      | 0.1163      | 0.0193      | 0.0047        | 0.0086        | 0.0110   |
| 2.89|           | 0.0739      | 0.1730      | 0.0242      | 0.0058        | 0.0093        | 0.0117   |
| 3.37|           | 0.1122      | 0.2144      | 0.0290      | 0.0065        | 0.0110        | 0.0125   |
| 3.84|           | 0.1617      | 0.3181      | 0.0333      | 0.0072        | 0.0121        | 0.0143   |
| 4.32|           | 0.2107      | 0.3967      | 0.0386      | 0.0087        | 0.0134        | 0.0134   |
| 4.79|           | 0.2807      | 0.5606      | 0.0417      | 0.0088        | 0.0143        | 0.0170   |
| 5.26|           | 0.3329      | 0.5881      | 0.0459      | 0.0099        | 0.0155        | 0.0170   |
| 5.74|           | 0.4164      | 0.8417      | 0.0462      | 0.0097        | 0.0145        | 0.0180   |
| 6.21|           | 0.4679      | 0.8738      | 0.0530      | 0.0113        | 0.0174        | 0.0186   |
| 6.68|           | 0.5827      | 1.4497      | 0.0554      | 0.0115        | 0.0187        | 0.0185   |
| 7.16|           | 0.6437      | 1.3686      | 0.0589      | 0.0121        | 0.0188        | 0.0198   |
| 7.63|           | 0.7096      | 1.8358      | 0.0680      | 0.0140        | 0.0198        | 0.0220   |
| 8.11|           | 0.7989      | 2.1821      | 0.0631      | 0.0124        | 0.0194        | 0.0216   |
| 8.58|           | 0.8902      | 2.6299      | 0.0749      | 0.0143        | 0.0219        | 0.0243   |
| 9.05|           | 0.9666      | 3.1198      | 0.0788      | 0.0154        | 0.0216        | 0.0230   |
| 9.53|           | 1.0641      | 3.5497      | 0.0745      | 0.0154        | 0.0213        | 0.0213   |
| 10.00|          | 1.1415      | 3.8314      | 0.0730      | 0.0145        | 0.0216        | 0.0252   |
Table 4

*Coverage Probability of 95% Confidence Interval across 20 M Values for Each Parameter for Noninformative Censoring of Scenario 2*

| M   | Parameter | \( \theta \) | \( \kappa \) | \( \alpha \) | \( \gamma_1 \) | \( \gamma_2 \) | \( \eta \) |
|-----|-----------|--------------|--------------|--------------|--------------|--------------|--------------|
| 1.00|           | 0.964        | 0.970        | 0.978        | 0.944        | 0.940        | 0.944        |
| 1.47|           | 0.934        | 0.976        | 0.980        | 0.950        | 0.950        | 0.950        |
| 1.95|           | 0.860        | 0.974        | 0.958        | 0.942        | 0.950        | 0.940        |
| 2.42|           | 0.708        | 0.974        | 0.946        | 0.948        | 0.946        | 0.940        |
| 2.89|           | 0.554        | 0.972        | 0.928        | 0.936        | 0.950        | 0.944        |
| 3.37|           | 0.360        | 0.984        | 0.932        | 0.934        | 0.948        | 0.952        |
| 3.84|           | 0.204        | 0.986        | 0.938        | 0.940        | 0.944        | 0.946        |
| 4.32|           | 0.110        | 0.986        | 0.918        | 0.916        | 0.932        | 0.958        |
| 4.79|           | 0.068        | 0.992        | 0.920        | 0.908        | 0.936        | 0.942        |
| 5.26|           | 0.044        | 0.994        | 0.902        | 0.916        | 0.932        | 0.938        |
| 5.74|           | 0.022        | 0.998        | 0.922        | 0.950        | 0.962        | 0.926        |
| 6.21|           | 0.006        | 0.996        | 0.920        | 0.926        | 0.954        | 0.948        |
| 6.68|           | 0.008        | 0.998        | 0.924        | 0.926        | 0.948        | 0.934        |
| 7.16|           | 0.002        | 1.000        | 0.910        | 0.924        | 0.928        | 0.934        |
| 7.63|           | 0.000        | 1.000        | 0.904        | 0.928        | 0.944        | 0.928        |
| 8.11|           | 0.000        | 0.996        | 0.922        | 0.946        | 0.942        | 0.932        |
| 8.58|           | 0.000        | 0.998        | 0.910        | 0.928        | 0.934        | 0.916        |
| 9.05|           | 0.000        | 0.998        | 0.910        | 0.918        | 0.946        | 0.918        |
| 9.53|           | 0.002        | 0.996        | 0.920        | 0.924        | 0.944        | 0.944        |
| 10.00|          | 0.000        | 1.000        | 0.918        | 0.934        | 0.934        | 0.920        |
### Table 5

**Bias across 20 M Values for Each Parameter for Informative Censoring of Scenario 2**

| M    | Parameter | \( \theta \) | \( \kappa \) | \( \alpha \) | \( \gamma_1 \) | \( \gamma_2 \) | \( \eta \) |
|------|-----------|---------------|---------------|---------------|---------------|---------------|---------------|
| 1.00 |           | -0.0072       | -0.0213       | 0.0043        | 0.0026        | 0.0021        | 0.0006        |
| 1.47 |           | -0.0090       | -0.0627       | 0.0073        | 0.0033        | 0.0027        | 0.0003        |
| 1.95 |           | -0.0020       | -0.0342       | 0.0165        | 0.0204        | 0.0197        | -0.0273       |
| 2.42 |           | -0.0213       | -0.0283       | 0.0051        | 0.0290        | 0.0290        | -0.0398       |
| 2.89 |           | -0.0520       | -0.0399       | -0.0118       | 0.0344        | 0.0316        | -0.0442       |
| 3.37 |           | -0.0973       | -0.0753       | -0.0368       | 0.0332        | 0.0359        | -0.0451       |
| 3.84 |           | -0.1433       | -0.1204       | -0.0628       | 0.0323        | 0.0321        | -0.0404       |
| 4.32 |           | -0.1933       | -0.1745       | -0.0910       | 0.0266        | 0.0266        | -0.0312       |
| 4.79 |           | -0.2325       | -0.2144       | -0.1096       | 0.0254        | 0.0258        | -0.0304       |
| 5.26 |           | -0.2818       | -0.2792       | -0.1342       | 0.0204        | 0.0207        | -0.0204       |
| 5.74 |           | -0.3272       | -0.3511       | -0.1559       | 0.0149        | 0.0147        | -0.0177       |
| 6.21 |           | -0.3752       | -0.4087       | -0.1780       | 0.0107        | 0.0138        | -0.0082       |
| 6.68 |           | -0.4114       | -0.4627       | -0.1917       | 0.0089        | 0.0083        | 0.0006        |
| 7.16 |           | -0.4548       | -0.5508       | -0.2157       | 0.0011        | -0.0021       | 0.0066        |
| 7.63 |           | -0.4903       | -0.6129       | -0.2269       | -0.0015       | 0.0006        | 0.0071        |
| 8.11 |           | -0.5264       | -0.6710       | -0.2414       | -0.0054       | -0.0049       | 0.0103        |
| 8.58 |           | -0.5630       | -0.7384       | -0.2551       | -0.0086       | -0.0048       | 0.0238        |
| 9.05 |           | -0.6028       | -0.8353       | -0.2715       | -0.0145       | -0.0145       | 0.0322        |
| 9.53 |           | -0.6282       | -0.8970       | -0.2818       | -0.0175       | -0.0210       | 0.0294        |
| 10.00|           | -0.6637       | -0.9847       | -0.2975       | -0.0235       | -0.0238       | 0.0335        |
Table 6

*Standard Error across 20 M Values for Each Parameter for Informative Censoring of Scenario 2*

| M   | Parameter | \( \theta \) | \( \kappa \) | \( \alpha \) | \( \gamma_1 \) | \( \gamma_2 \) | \( \eta \) |
|-----|-----------|--------------|--------------|-------------|--------------|-------------|---------|
| 1.00|           | 0.0955       | 0.1782       | 0.1101      | 0.0549       | 0.0776      | 0.0799  |
| 1.47|           | 0.1075       | 0.2138       | 0.1181      | 0.0587       | 0.0825      | 0.0883  |
| 1.95|           | 0.1161       | 0.2281       | 0.1227      | 0.0618       | 0.0877      | 0.0984  |
| 2.42|           | 0.1223       | 0.2420       | 0.1258      | 0.0642       | 0.0917      | 0.1055  |
| 2.89|           | 0.1268       | 0.2563       | 0.1280      | 0.0664       | 0.0948      | 0.1110  |
| 3.37|           | 0.1310       | 0.2757       | 0.1302      | 0.0682       | 0.0972      | 0.1147  |
| 3.84|           | 0.1345       | 0.2967       | 0.1318      | 0.0698       | 0.0990      | 0.1175  |
| 4.32|           | 0.1378       | 0.3197       | 0.1335      | 0.0711       | 0.1004      | 0.1195  |
| 4.79|           | 0.1406       | 0.3390       | 0.1352      | 0.0725       | 0.1020      | 0.1216  |
| 5.26|           | 0.1427       | 0.3649       | 0.1365      | 0.0736       | 0.1032      | 0.1226  |
| 5.74|           | 0.1460       | 0.3947       | 0.1384      | 0.0749       | 0.1041      | 0.1233  |
| 6.21|           | 0.1477       | 0.4180       | 0.1395      | 0.0759       | 0.1051      | 0.1240  |
| 6.68|           | 0.1499       | 0.4413       | 0.1413      | 0.0771       | 0.1061      | 0.1248  |
| 7.16|           | 0.1529       | 0.4781       | 0.1425      | 0.0777       | 0.1065      | 0.1252  |
| 7.63|           | 0.1554       | 0.5070       | 0.1444      | 0.0789       | 0.1077      | 0.1260  |
| 8.11|           | 0.1568       | 0.5313       | 0.1455      | 0.0797       | 0.1083      | 0.1265  |
| 8.58|           | 0.1590       | 0.5616       | 0.1471      | 0.0806       | 0.1093      | 0.1266  |
| 9.05|           | 0.1616       | 0.6036       | 0.1488      | 0.0816       | 0.1099      | 0.1263  |
| 9.53|           | 0.1640       | 0.6336       | 0.1502      | 0.0823       | 0.1103      | 0.1269  |
|10.00|           | 0.1660       | 0.6713       | 0.1514      | 0.0828       | 0.1109      | 0.1270  |
Table 7

*Mean Squared Error across 20 M Values for Each Parameter for Informative Censoring of Scenario 2*

| M  | Parameter | \( \theta \) | \( \kappa \) | \( \alpha \) | \( \gamma_1 \) | \( \gamma_2 \) | \( \eta \) |
|----|-----------|---------------|---------------|---------------|---------------|---------------|---------------|
| 1.00 | 0.0081   | 0.0288        | 0.0106        | 0.0029        | 0.0059        | 0.0067        |
| 1.47 | 0.0103   | 0.0456        | 0.0121        | 0.0035        | 0.0068        | 0.0076        |
| 1.95 | 0.0124   | 0.0540        | 0.0142        | 0.0044        | 0.0084        | 0.0107        |
| 2.42 | 0.0129   | 0.0530        | 0.0139        | 0.0047        | 0.0088        | 0.0120        |
| 2.89 | 0.0170   | 0.0664        | 0.0145        | 0.0055        | 0.0096        | 0.0140        |
| 3.37 | 0.0243   | 0.0731        | 0.0156        | 0.0057        | 0.0108        | 0.0161        |
| 3.84 | 0.0365   | 0.1024        | 0.0194        | 0.0058        | 0.0108        | 0.0156        |
| 4.32 | 0.0542   | 0.1253        | 0.0233        | 0.0054        | 0.0107        | 0.0154        |
| 4.79 | 0.0702   | 0.1463        | 0.0266        | 0.0054        | 0.0102        | 0.0160        |
| 5.26 | 0.0977   | 0.2100        | 0.0340        | 0.0057        | 0.0108        | 0.0165        |
| 5.74 | 0.1262   | 0.2746        | 0.0411        | 0.0059        | 0.0114        | 0.0148        |
| 6.21 | 0.1589   | 0.3115        | 0.0472        | 0.0051        | 0.0108        | 0.0138        |
| 6.68 | 0.1874   | 0.3717        | 0.0523        | 0.0056        | 0.0104        | 0.0156        |
| 7.16 | 0.2274   | 0.5236        | 0.0633        | 0.0058        | 0.0110        | 0.0152        |
| 7.63 | 0.2625   | 0.6209        | 0.0701        | 0.0062        | 0.0116        | 0.0167        |
| 8.11 | 0.2992   | 0.7125        | 0.0759        | 0.0059        | 0.0114        | 0.0163        |
| 8.58 | 0.3391   | 0.8280        | 0.0820        | 0.0061        | 0.0109        | 0.0168        |
| 9.05 | 0.3866   | 1.0426        | 0.0924        | 0.0068        | 0.0123        | 0.0166        |
| 9.53 | 0.4168   | 1.1711        | 0.0980        | 0.0068        | 0.0124        | 0.0173        |
| 10.00| 0.4647   | 1.3717        | 0.1075        | 0.0071        | 0.0123        | 0.0179        |
Table 8

Coverage Probability of 95% Confidence Interval across 20 M Values for Each Parameter for Informative Censoring of Scenario 2

| M   | Parameter | \(\theta\) | \(\kappa\) | \(\alpha\) | \(\gamma_1\) | \(\gamma_2\) | \(\eta\) |
|-----|-----------|-------------|-------------|-------------|---------------|---------------|---------|
| 1.00| \(\theta\) | 0.964       | 0.970       | 0.978       | 0.944         | 0.940         | 0.944   |
| 1.47| \(\kappa\) | 0.966       | 0.966       | 0.970       | 0.950         | 0.954         | 0.956   |
| 1.95| \(\alpha\) | 0.962       | 0.958       | 0.962       | 0.938         | 0.936         | 0.944   |
| 2.42| \(\gamma_1\) | 0.968       | 0.962       | 0.962       | 0.940         | 0.952         | 0.952   |
| 2.89| \(\gamma_2\) | 0.960       | 0.966       | 0.974       | 0.918         | 0.942         | 0.944   |
| 3.37| \(\eta\)   | 0.902       | 0.972       | 0.970       | 0.936         | 0.948         | 0.930   |
| 3.84| \(\theta\) | 0.838       | 0.968       | 0.940       | 0.940         | 0.942         | 0.954   |
| 4.32| \(\kappa\) | 0.740       | 0.984       | 0.918       | 0.952         | 0.962         | 0.944   |
| 4.79| \(\alpha\) | 0.634       | 0.994       | 0.906       | 0.954         | 0.956         | 0.940   |
| 5.26| \(\gamma_1\) | 0.482       | 0.990       | 0.842       | 0.958         | 0.948         | 0.950   |
| 5.74| \(\gamma_2\) | 0.370       | 0.978       | 0.816       | 0.944         | 0.940         | 0.962   |
| 6.21| \(\eta\)   | 0.240       | 0.990       | 0.762       | 0.958         | 0.954         | 0.964   |
| 6.68| \(\theta\) | 0.172       | 0.994       | 0.740       | 0.954         | 0.968         | 0.950   |
| 7.16| \(\kappa\) | 0.120       | 0.990       | 0.670       | 0.954         | 0.950         | 0.956   |
| 7.63| \(\alpha\) | 0.074       | 0.986       | 0.622       | 0.944         | 0.940         | 0.956   |
| 8.11| \(\gamma_1\) | 0.040       | 0.982       | 0.610       | 0.952         | 0.942         | 0.940   |
| 8.58| \(\gamma_2\) | 0.036       | 0.980       | 0.588       | 0.946         | 0.948         | 0.940   |
| 9.05| \(\eta\)   | 0.018       | 0.970       | 0.548       | 0.954         | 0.946         | 0.938   |
| 9.53| \(\theta\) | 0.014       | 0.984       | 0.518       | 0.940         | 0.932         | 0.932   |
| 10.00| \(\kappa\) | 0.014       | 0.974       | 0.470       | 0.946         | 0.940         | 0.940   |
Table 9

Estimates, Standard Errors, Test Statistics, and P-values across 20 M Values for Treat and CD4 Variables in ACTG 175 for Noninformative Censoring

| M  | Treat Estimate | Treat SE  | Treat Test | Treat P  | CD4 Estimate | CD4 SE  | CD4 Test | CD4 P  |
|----|----------------|----------|------------|----------|--------------|---------|----------|--------|
| 1.00 | -0.3338       | 0.0596   | -5.5967    | <0.0001  | -1.1157      | 0.0749  | -14.9004 | <0.0001|
| 1.47 | -0.2569       | 0.0524   | -4.9052    | <0.0001  | -0.9993      | 0.0853  | -11.7196 | <0.0001|
| 1.95 | -0.2000       | 0.0508   | -3.9365    | 0.0001   | -0.8625      | 0.1004  | -8.5918  | <0.0001|
| 2.42 | -0.1159       | 0.0308   | -3.7585    | 0.0002   | -0.4179      | 0.0840  | -4.9778  | <0.0001|
| 2.89 | -0.1161       | 0.0396   | -2.9276    | 0.0034   | -0.4609      | 0.1325  | -3.4792  | 0.0005  |
| 3.37 | -0.1258       | 0.0381   | -3.3057    | 0.0009   | -0.4993      | 0.1034  | -4.8272  | <0.0001|
| 3.84 | -0.0570       | 0.0219   | -2.6061    | 0.0092   | -0.2804      | 0.0753  | -3.7250  | 0.0002  |
| 4.32 | -0.1026       | 0.0328   | -3.1290    | 0.0018   | -0.4184      | 0.0957  | -4.3714  | <0.0001|
| 4.79 | -0.1253       | 0.0354   | -3.5446    | 0.0004   | -0.4682      | 0.0942  | -4.9717  | <0.0001|
| 5.26 | -0.0612       | 0.0143   | -4.2821    | <0.0001  | -0.2659      | 0.0214  | -12.4245 | <0.0001|
| 5.74 | -0.0950       | 0.0356   | -2.6716    | 0.0075   | -0.3776      | 0.0950  | -3.9753  | 0.0001  |
| 6.21 | -0.0421       | 0.0164   | -2.5586    | 0.0105   | -0.2312      | 0.0201  | -11.4909 | <0.0001|
| 6.68 | -0.0675       | 0.0156   | -4.3216    | <0.0001  | -0.2755      | 0.0226  | -12.2082 | <0.0001|
| 7.16 | -0.0528       | 0.0225   | -2.3466    | 0.0189   | -0.3168      | 0.0829  | -3.8194  | 0.0001  |
| 7.63 | -0.0677       | 0.0281   | -2.4100    | 0.0160   | -0.3130      | 0.0978  | -3.1994  | 0.0014  |
| 8.11 | -0.0619       | 0.0250   | -2.4727    | 0.0134   | -0.2841      | 0.0718  | -3.9563  | 0.0001  |
| 8.58 | -0.0642       | 0.0281   | -2.2829    | 0.0224   | -0.3248      | 0.0864  | -3.7602  | 0.0002  |
| 9.05 | -0.1013       | 0.0383   | -2.6471    | 0.0081   | -0.4562      | 0.1265  | -3.6064  | 0.0003  |
| 9.53 | -0.1037       | 0.0510   | -2.0339    | 0.0420   | -0.4057      | 0.1654  | -2.4531  | 0.0142  |
| 10.00| -0.0715       | 0.0355   | -2.0167    | 0.0437   | -0.3961      | 0.1441  | -2.7494  | 0.0060  |
Table 10

Estimates, Standard Errors, Test Statistics, and P-values across 20 M Values for Treat and CD4 Variables in ACTG 175 for Informative Censoring

| M   | Treat Estimate | SE  | Test   | P    | CD4 Estimate | SE  | Test   | P    |
|-----|----------------|-----|--------|------|--------------|-----|--------|------|
| 1.00| -0.3338        | 0.0596 | -5.5967| <0.0001 | -1.1157      | 0.0749 | -14.9004| <0.0001 |
| 1.47| -0.2977        | 0.0563 | -5.2895| <0.0001 | -1.0544      | 0.0773 | -13.6413| <0.0001 |
| 1.95| -0.3027        | 0.0550 | -5.5048| <0.0001 | -1.0383      | 0.0768 | -13.5231| <0.0001 |
| 2.42| -0.3067        | 0.0590 | -5.1971| <0.0001 | -1.0608      | 0.0781 | -13.5901| <0.0001 |
| 2.89| -0.3249        | 0.0591 | -5.5020| <0.0001 | -1.0817      | 0.0796 | -13.5952| <0.0001 |
| 3.37| -0.3234        | 0.0578 | -5.5941| <0.0001 | -1.0650      | 0.0802 | -13.2871| <0.0001 |
| 3.84| -0.3230        | 0.0568 | -5.6869| <0.0001 | -1.0709      | 0.0811 | -13.2108| <0.0001 |
| 4.32| -0.3383        | 0.0592 | -5.7111| <0.0001 | -1.0983      | 0.0800 | -13.7285| <0.0001 |
| 4.79| -0.2854        | 0.0601 | -4.7467| <0.0001 | -1.0958      | 0.0825 | -13.2794| <0.0001 |
| 5.26| -0.3178        | 0.0602 | -5.2766| <0.0001 | -1.0779      | 0.0799 | -13.4905| <0.0001 |
| 5.74| -0.3237        | 0.0564 | -5.7421| <0.0001 | -1.0428      | 0.0811 | -12.8631| <0.0001 |
| 6.21| -0.2817        | 0.0587 | -4.7947| <0.0001 | -1.0700      | 0.0816 | -13.1177| <0.0001 |
| 6.68| -0.3247        | 0.0580 | -5.5948| <0.0001 | -1.0396      | 0.0822 | -12.6410| <0.0001 |
| 7.16| -0.2809        | 0.0585 | -4.8045| <0.0001 | -1.0891      | 0.0838 | -12.9958| <0.0001 |
| 7.63| -0.3139        | 0.0604 | -5.1940| <0.0001 | -1.1010      | 0.0843 | -13.0537| <0.0001 |
| 8.11| -0.2912        | 0.0558 | -5.2190| <0.0001 | -1.0058      | 0.0835 | -12.0474| <0.0001 |
| 8.58| -0.3040        | 0.0565 | -5.3760| <0.0001 | -1.0581      | 0.0857 | -12.3508| <0.0001 |
| 9.05| -0.2913        | 0.0585 | -4.9821| <0.0001 | -1.0917      | 0.0842 | -12.9649| <0.0001 |
| 9.53| -0.3603        | 0.0627 | -5.7482| <0.0001 | -1.0652      | 0.0834 | -12.7686| <0.0001 |
| 10.00| -0.2846       | 0.0571 | -4.9843| <0.0001 | -1.0445      | 0.0857 | -12.1912| <0.0001 |
Table 11

*Estimates and Standard Errors across 20 M Values for Karnof, Symptom, Treat and CD4 Variables in ACTG 175 for Noninformative Censoring*

| M   | Karnof | Symptom  | Treat  | CD4  | Karnof | Symptom  | Treat  | CD4  |
|-----|--------|----------|--------|------|--------|----------|--------|------|
| 1.00 | -0.1967 | 0.2342  | -0.3490 | -1.0675 | 0.0567 | 0.0607  | 0.0607 | 0.0738 |
| 1.47 | -0.1383 | 0.1895  | -0.2699 | -0.9844 | 0.0526 | 0.0494  | 0.0549 | 0.0815 |
| 1.95 | -0.0944 | 0.1451  | -0.2288 | -0.9046 | 0.0508 | 0.0479  | 0.0557 | 0.0969 |
| 2.42 | -0.0813 | 0.0853  | -0.1260 | -0.4431 | 0.0300 | 0.0288  | 0.0333 | 0.0874 |
| 2.89 | -0.0563 | 0.0531  | -0.1332 | -0.5269 | 0.0330 | 0.0331  | 0.0423 | 0.1314 |
| 3.37 | -0.0315 | 0.0664  | -0.1283 | -0.5098 | 0.0313 | 0.0324  | 0.0405 | 0.1075 |
| 3.84 | -0.0214 | 0.0450  | -0.0709 | -0.3310 | 0.0212 | 0.0216  | 0.0247 | 0.0674 |
| 4.32 | -0.0368 | 0.0940  | -0.1076 | -0.4079 | 0.0259 | 0.0306  | 0.0329 | 0.0921 |
| 4.79 | -0.0132 | 0.0468  | -0.1269 | -0.4651 | 0.0284 | 0.0282  | 0.0368 | 0.0987 |
| 5.26 | -0.0258 | 0.0546  | -0.0795 | -0.3590 | 0.0213 | 0.0217  | 0.0242 | 0.0699 |
| 5.74 | -0.0615 | 0.0951  | -0.1115 | -0.3753 | 0.0319 | 0.0337  | 0.0381 | 0.0972 |
| 6.21 | -0.0471 | 0.0766  | -0.0709 | -0.3450 | 0.0295 | 0.0333  | 0.0308 | 0.0978 |
| 6.68 | -0.0652 | 0.0689  | -0.1370 | -0.4916 | 0.0332 | 0.0347  | 0.0524 | 0.1649 |
| 7.16 | -0.0351 | 0.0652  | -0.0572 | -0.3501 | 0.0226 | 0.0249  | 0.0240 | 0.0772 |
| 7.63 | -0.0286 | 0.0996  | -0.0921 | -0.3656 | 0.0295 | 0.0316  | 0.0314 | 0.0843 |
| 8.11 | -0.0045 | 0.0928  | -0.0695 | -0.3459 | 0.0294 | 0.0314  | 0.0281 | 0.0764 |
| 8.58 | -0.0274 | 0.0499  | -0.0606 | -0.3242 | 0.0265 | 0.0271  | 0.0292 | 0.0927 |
| 9.05 | -0.0330 | 0.0378  | -0.1041 | -0.4988 | 0.0314 | 0.0326  | 0.0386 | 0.1184 |
| 9.53 | -0.0188 | 0.0697  | -0.1096 | -0.4303 | 0.0340 | 0.0378  | 0.0487 | 0.1428 |
| 10.00 | -0.1166 | 0.0886  | -0.0972 | -0.4768 | 0.0444 | 0.0357  | 0.0397 | 0.1388 |
Table 12

Test Statistics and P-values across 20 M Values for Karnof, Symptom, Treat and CD4 Variables in ACTG 175 for Noninformative Censoring

| M     | Test Statistics |  | P-values |
|-------|-----------------|---|----------|
|       | Karnof          |  |          |          |      |
| 1.00  | -3.2415         |  | 4.1281   | -5.7495  | -14.4636 |
| 1.47  | 3.8389          | -4.9124 | -12.0758 | 0.0086   | <0.0001  |
| 1.95  | 3.0311          | -4.1086 | -9.3374  | 0.0632   | <0.0001  |
| 2.42  | 2.9591          | -3.7801 | -5.0695  | 0.0067   | 0.0002   |
| 2.89  | 1.6035          | -3.1484 | -4.0097  | 0.0875   | 0.1088   |
| 3.37  | 2.0473          | -3.1652 | -4.7429  | 0.3136   | 0.0406   |
| 3.84  | 2.0818          | -2.8696 | -4.9113  | 0.3131   | 0.0374   |
| 4.32  | 3.0721          | -3.2718 | -4.4277  | 0.1560   | 0.0021   |
| 4.79  | 1.6565          | -3.4505 | -4.7122  | 0.6418   | 0.0976   |
| 5.26  | 2.5146          | -3.2871 | -5.1358  | 0.2262   | 0.0119   |
| 5.74  | 2.8206          | -2.9293 | -3.8615  | 0.0539   | 0.0048   |
| 6.21  | 2.3013          | -2.2984 | -3.5274  | 0.1110   | 0.0214   |
| 6.68  | 1.9871          | -2.6132 | -2.9809  | 0.0500   | 0.0469   |
| 7.16  | 2.6241          | -2.3810 | -4.5327  | 0.1192   | 0.0087   |
| 7.63  | 3.1520          | -2.9362 | -4.3388  | 0.3326   | 0.0016   |
| 8.11  | 2.9539          | -2.4730 | -4.5250  | 0.8779   | 0.0031   |
| 8.58  | 1.8442          | -2.0759 | -3.4978  | 0.3011   | 0.0651   |
| 9.05  | 1.1605          | -2.6932 | -4.2114  | 0.2923   | 0.2459   |
| 9.53  | 1.8414          | -2.2520 | -3.0130  | 0.5797   | 0.0656   |
| 10.00 | 2.4788          | -2.4499 | -3.4352  | 0.0086   | 0.0132   |
Table 13

*Estimates and Standard Errors across 20 M Values for Karnof, Symptom, Treat and CD4 Variables in ACTG 175 for Informative Censoring*

| M   | Karnof | Symptom | Treat | CD4  | Karnof | Symptom | Treat | CD4  |
|-----|--------|---------|-------|------|--------|---------|-------|------|
| 1.00| -0.1967| 0.2342  | -0.3490| -1.0675| 0.0607 | 0.0567  | 0.0607| 0.0738|
| 1.47| -0.1857| 0.2156  | -0.3127| -1.0200| 0.0569 | 0.0536  | 0.0580| 0.0746|
| 1.95| -0.1559| 0.2002  | -0.3125| -1.0053| 0.0548 | 0.0520  | 0.0564| 0.0750|
| 2.42| -0.1616| 0.2283  | -0.3231| -1.0235| 0.0583 | 0.0551  | 0.0606| 0.0761|
| 2.89| -0.1760| 0.2130  | -0.3362| -1.0422| 0.0583 | 0.0539  | 0.0604| 0.0783|
| 3.37| -0.1674| 0.2171  | -0.3363| -1.0299| 0.0576 | 0.0541  | 0.0591| 0.0777|
| 3.84| -0.1512| 0.2257  | -0.3371| -1.0366| 0.0583 | 0.0548  | 0.0586| 0.0789|
| 4.32| -0.1961| 0.2176  | -0.3584| -1.0656| 0.0604 | 0.0559  | 0.0616| 0.0783|
| 4.79| -0.2122| 0.1940  | -0.2984| -1.0515| 0.0613 | 0.0565  | 0.0611| 0.0811|
| 5.26| -0.1764| 0.1871  | -0.3312| -1.0447| 0.0582 | 0.0551  | 0.0615| 0.0788|
| 5.74| -0.1592| 0.2193  | -0.3451| -1.0178| 0.0584 | 0.0544  | 0.0583| 0.0792|
| 6.21| -0.1608| 0.1984  | -0.2999| -1.0497| 0.0597 | 0.0558  | 0.0612| 0.0803|
| 6.68| -0.1694| 0.2001  | -0.3411| -1.0045| 0.0595 | 0.0546  | 0.0592| 0.0802|
| 7.16| -0.1945| 0.2003  | -0.3023| -1.0661| 0.0604 | 0.0566  | 0.0607| 0.0806|
| 7.63| -0.1713| 0.2249  | -0.3225| -1.0590| 0.0567 | 0.0536  | 0.0609| 0.0825|
| 8.11| -0.1762| 0.2009  | -0.3074| -0.9850| 0.0585 | 0.0538  | 0.0581| 0.0802|
| 8.58| -0.1583| 0.1823  | -0.3162| -1.0235| 0.0564 | 0.0530  | 0.0584| 0.0838|
| 9.05| -0.2024| 0.2294  | -0.3110| -1.0637| 0.0608 | 0.0575  | 0.0612| 0.0825|
| 9.53| -0.2189| 0.2226  | -0.3878| -1.0319| 0.0625 | 0.0588  | 0.0650| 0.0811|
|10.00| -0.1993| 0.1876  | -0.3107| -1.0439| 0.0606 | 0.0564  | 0.0601| 0.0825|
Table 14

*Test Statistics and P-values across 20 M Values for Karnof, Symptom, Treat and CD4 Variables in ACTG 175 for Informative Censoring*

| M   | Test Statistics | P-values |   |   |   |   |   |   |   |
|-----|-----------------|----------|---|---|---|---|---|---|---|
|     | Karnof          | Symptom  | Treat | CD4 | Karnof | Symptom | Treat | CD4 |
| 1.00| -3.2415         | 4.1281   | -5.7495 | -14.4636 | 0.0012 | <0.0001 | <0.0001 | <0.0001 |
| 1.47| -3.2621         | 4.0248   | -5.3929 | -13.6792 | 0.0011 | 0.0001 | <0.0001 | <0.0001 |
| 1.95| -2.8432         | 3.8500   | -5.5376 | -13.4064 | 0.0045 | 0.0001 | <0.0001 | <0.0001 |
| 2.42| -2.7696         | 4.1463   | -5.3341 | -13.4444 | 0.0056 | <0.0001 | <0.0001 | <0.0001 |
| 2.89| -3.0189         | 3.9535   | -5.5625 | -13.3162 | 0.0025 | 0.0001 | <0.0001 | <0.0001 |
| 3.37| -2.9054         | 4.0113   | -5.6897 | -13.2598 | 0.0037 | 0.0001 | <0.0001 | <0.0001 |
| 3.84| -2.5934         | 4.1231   | -5.7556 | -13.1307 | 0.0095 | <0.0001 | <0.0001 | <0.0001 |
| 4.32| -3.2478         | 3.8926   | -5.8205 | -13.6000 | 0.0012 | 0.0001 | <0.0001 | <0.0001 |
| 4.79| -3.4648         | 3.4327   | -4.8865 | -12.9653 | 0.0005 | 0.0006 | <0.0001 | <0.0001 |
| 5.26| -3.0316         | 3.3955   | -5.3874 | -13.2565 | 0.0024 | 0.0007 | <0.0001 | <0.0001 |
| 5.74| -2.7284         | 4.0280   | -5.9197 | -12.8519 | 0.0064 | 0.0001 | <0.0001 | <0.0001 |
| 6.21| -2.6944         | 3.5535   | -4.9006 | -13.0641 | 0.0071 | 0.0004 | <0.0001 | <0.0001 |
| 6.68| -2.8489         | 3.6647   | -5.7581 | -12.5192 | 0.0044 | 0.0002 | <0.0001 | <0.0001 |
| 7.16| -3.2203         | 3.5375   | -4.9827 | -13.2207 | 0.0013 | 0.0004 | <0.0001 | <0.0001 |
| 7.63| -3.0219         | 4.1981   | -5.2982 | -12.8412 | 0.0025 | <0.0001 | <0.0001 | <0.0001 |
| 8.11| -3.0116         | 3.7334   | -5.2926 | -12.2856 | 0.0026 | 0.0002 | <0.0001 | <0.0001 |
| 8.58| -2.8085         | 3.4378   | -5.4153 | -12.2076 | 0.0050 | 0.0006 | <0.0001 | <0.0001 |
| 9.05| -3.3304         | 3.9869   | -5.0835 | -12.8886 | 0.0009 | 0.0001 | <0.0001 | <0.0001 |
| 9.53| -3.5042         | 3.7837   | -5.9702 | -12.7288 | 0.0005 | 0.0002 | <0.0001 | <0.0001 |
| 10.00| -3.2869        | 3.3265   | -5.1680 | -12.6592 | 0.0010 | 0.0009 | <0.0001 | <0.0001 |
Figure 1

*Boxplot for Parameter Estimates of Each Parameter across 20 M Values for Noninformative Censoring (a) vs Informative Censoring (b) for Scenario 2*
Figure 2

Coverage Probability of 95% Confidence Interval across 20 M Values for Each Parameter for Noninformative Censoring (a) vs Informative Censoring (b) for Scenario 2
Figure 3

Boxplot for Parameter Estimates of Each Parameter across 20 M Values for Noninformative Censoring (a) vs Informative Censoring (b) for Scenario 1

(a)

(b)
Figure 4

Coverage Probability of 95% Confidence Interval across 20 M Values for Each Parameter for Noninformative Censoring (a) vs Informative Censoring (b) for Scenario 1
CHAPTER 4

CONCLUSION

The present study developed an easy-to-use code for interval-censored survival data with both fixed and time-dependent covariates. Extensive simulations were conducted to investigate the robustness of the interval-censored survival analysis with inaccurate time bounds and time-dependent covariates, particularly under noninformative censoring and informative censoring. Moreover, a real data analysis on the ACTG 175 data was conducted to investigate the robustness of findings under different variations of time bounds.

The likelihood approach was used to draw inferences about the unknown parameters. The parameter estimator was the maximizer of the log-likelihood function and was asymptotically normal. The R package, numDeriv, was used to calculate the Hessian matrix of the negative log-likelihood. The estimated variance was the diagonal of the inverse of the Hessian matrix and the standard errors of the estimates equaled the square root of the estimated variance. The confidence intervals were constructed using estimates plus and minus 1.96 times standard errors.

In the simulations, the parameter estimates were obtained, and then the bias, the standard error, the mean squared error, and the coverage probability of 95% confidence interval for each parameter were calculated and reported for each of the 20 $M$ values for the simulated dataset including both fixed and time-dependent covariates. The findings demonstrated that for both noninformative censoring and informative
censoring, parameter estimates as well as coverage probability were more robust against change across $M$ values for $\gamma_1$, $\gamma_2$, and $\eta$ than for $\theta$, $\kappa$, and $\alpha$, which were not that robust against change across $M$ values. The parameter estimates and coverage probability were also demonstrated for each of the 20 $M$ values for the simulated dataset including only fixed covariates. The findings demonstrated that for both noninformative censoring and informative censoring, parameter estimates as well as coverage probability were more robust against change across $M$ values for $\gamma_1$ and $\gamma_2$ than for $\theta$, $\kappa$, and $\alpha$, which were not that robust against change across $M$ values.

In the scenarios considered in this simulation study, for both noninformative censoring and informative censoring, the scale and shape parameters ($\theta$, $\kappa$, and $\alpha$) changed as $M$ value changed, so parameter estimates and coverage probability for $\theta$, $\kappa$, and $\alpha$ were not robust against change across $M$ values. However, since noninformative censoring is not related to survival time but informative censoring is related to survival time, the change of scale and shape parameters ($\theta$, $\kappa$, and $\alpha$) was not as severe as that for informative censoring. In addition, the informative censoring scenario considered in this study has no strong relationship with fixed covariates and time-dependent covariates, thus for informative censoring, parameter estimates and coverage probability for $\gamma_1$, $\gamma_2$, and $\eta$ were also robust against change across $M$ values.

In the real data analysis, the estimates, standard errors, and $p$-values across 20 $M$ values for the fixed covariate, $treat$, and the time-dependent covariate, $cd4$, in ACTG 175 were obtained. The findings demonstrated that for noninformative censoring, the estimates as well as the $p$-values for $treat$ and $cd4$ were relatively robust
against change across $M$ values; for informative censoring, the estimates as well as the $p$-values for treat and cd4 were more robust against change across $M$ values. The estimates for treat and cd4 were all negative and significant for both noninformative censoring and informative censoring, which indicated longer survival time and event occurred later. The estimates under informative censoring were most robust against change across $M$ values, which was consistent with the findings in the simulation study and worth further investigation.

The real data analysis in this study also included some more covariates. The estimates, standard errors, and $p$-values across 20 $M$ values were obtained for the fixed covariates, treat, karnof, and symptom, and the time-dependent covariate, cd4. The findings for this analysis when more covariates were included were compared with the findings for the analysis when only treat and cd4 were used. The results for treat and cd4 in this analysis were very similar to those in the analysis when only treat and cd4 were used.

This study was the first to develop the code for interval-censored survival data with both fixed and time-dependent covariates, and was the first to investigate the robustness of the interval-censored survival analysis under inaccurate time bounds. This study contributed to the research literature about the topic and contributed to the methodology, procedures, and analysis of future survival studies examining the similar topic.

This project used ACTG 175 as the real data and was an illustration of how to evaluate the robustness of findings under inaccurate time bounds. Similar procedures can be conducted on other datasets with inaccurate time bounds. Moreover, this study
analyzed clinical trial data and included one time-dependent covariate in the analysis. Future studies can include more time-dependent covariates and can be extended to other areas, such as the financial, accounting, and marketing areas; for example, be used for bankruptcy prediction and credit default prediction. What’s more, this study selected covariates that were included in the real data analysis based on previous studies analyzing ACTG 175 data (e.g., Hammer et al., 1996; Huang & Zhang, 2008; Scharfstein & Robins, 2002; Song et al., 2002; Song & Ma, 2008; Song & Wang, 2017; Wen, 2012; Wen & Chen, 2014). Future studies can use automatic variable selection methods such as forward regression, backward regression, stepwise regression, and LASSO (e.g., Tibshirani, 1996), in order to select the best subset of predictors for the analysis. In addition, for interval-censored data of a large number of observations, future studies can use a possible approach, the divide and conquer strategy, to handle the data. For example, for the simulated data with 500000 observations, the data can be divided into 500 blocks and the parameter estimates and their variance within each block can be calculated. Finally, the 500 results will be merged into one.

Besides, future studies may include joint models for longitudinal and time-to-event data which bring these two data types together into a single model so that one can infer the association between the longitudinal biomarker and time to event to better examine the treatment effects and reduce bias in estimates of the treatment effects. One example of the joint model can be described as follows.

We can utilize subject-specific random effects to account for the possible dependence between the event time $X_{i}$ and the visit time $\tau_{i,j}$. This approach involves
two models; one for the gap time between adjacent visits, and the other for the event time. As an illustration, we can assume that the hazard function for the gap time \( \tau_{i,j+1} - \tau_{i,j} \) is

\[
\lambda_{\tau,i}(t; \omega_{\tau}) = \frac{\alpha_{\tau} e^{(\theta_{\tau} + \gamma_{\tau} T_{z,i} + \eta_{\tau} T_{Y_i}(\tau_{i,j}) + \zeta_{i}) t^{\alpha_{\tau}-1}}}{1 + e^{(\theta_{\tau} + \gamma_{\tau} T_{z,i} + \eta_{\tau} T_{Y_i}(\tau_{i,j}) + \zeta_{i}) t^{\alpha_{\tau}}}}
\]

where the parameter vector \( \omega_{\tau} = (\theta_{\tau}, \gamma_{\tau}, \kappa_{\tau}, \alpha_{\tau}, \eta_{\tau}, \sigma_{\tau}^2)^T \); \( \zeta_{i} \) is random effects for the \( i \)th subject, which is assumed to be distributed as \( N(0, \sigma_{\tau}^2) \). The corresponding probability density function is

\[
f_{\tau,i}(t; \omega_{\tau}) = \lambda_{\tau,i}(t; \omega_{\tau}) e^{-\int_0^t \lambda_{\tau,i}(s; \omega_{\tau}) \, ds}.
\]

The hazard function for event time \( X_i \) is

\[
\lambda_{X,i}(t; \omega_X) = \frac{\alpha_{X} e^{(\theta_{X} + \gamma_{X} T_{z,i} + \eta_{X} T_{Y_i}(t) + \rho_{X} \zeta_{i}) t^{\alpha_{X}-1}}}{1 + e^{(\theta_{X} + \gamma_{X} T_{z,i} + \eta_{X} T_{Y_i}(t) + \rho_{X} \zeta_{i}) t^{\alpha_{X}}}}
\]

where the parameter vector \( \omega_X = (\theta_{X}, \gamma_{X}, \kappa_{X}, \alpha_{X}, \eta_{X}, \rho_{X})^T \). The corresponding probability density function is

\[
f_{X,i}(t; \omega_X) = \lambda_{X,i}(t; \omega_X) e^{-\int_0^t \lambda_{X,i}(s; \omega_X) \, ds},
\]

and the cumulative distribution function is

\[
F_{X,i}(t; \omega_X) = 1 - e^{-\int_0^t \lambda_{X,i}(s; \omega_X) \, ds}.
\]

We can still use the likelihood approach to estimate \( \omega_{\tau} \) and \( \omega_X \). The likelihood function is
\[ l(\omega_T, \omega_X) = \prod_{i=1}^{n} \left\{ \int_{-\infty}^{\infty} \left[ f_{X,i}(L_i; \omega_X) \right]^{\delta_i} \left[ F_{X,i}(U_i; \omega_X) - F_{X,i}(L_i; \omega_X) \right]^{1-\delta_i} \prod_{j=1}^{J_i-1} f_{\tau,i}(\tau_{i,j+1} \right. \\
\left. \left( - \tau_{i,j}; \omega_T \right) \phi(\zeta_i; \sigma_{\tau}^2) \, d\zeta_i \right\}, \]

where \( \phi(\zeta; \sigma^2) \) is the density function of \( N(0, \sigma^2) \). We can also test for noninformative censoring by testing \( H_0: \rho_X = 0 \) against \( H_a: \rho_X \neq 0 \).
Chen, C. M., Lu, T. F. C., Chen, M. H., & Hsu, C. M. (2012). Semiparametric transformation models for current status data with informative censoring. *Biometrical Journal, 54*, 641-656.

Dettoni, R., Marra, G., & Radice, R. (2020). Generalized Link-Based Additive Survival Models with Informative Censoring. *Journal of Computational and Graphical Statistics, 29*, 503-512.

Emura, T., & Chen, Y. H. (2018). *Analysis of Survival Data with Dependent Censoring: Copula-Based Approaches*, Singapore: Springer.

Groeneboom, P., & Wellner, J. A. (1992). *Information Bounds and Non-parametric Maximum Likelihood Estimation*, Birkhauser, Boston.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., ... & Merigan, T. C. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine, 335*, 1081-1090.

Huang, X., & Zhang, N. (2008). Regression survival analysis with an assumed copula for dependent censoring: a sensitivity analysis approach. *Biometrics, 64*, 1090-1099.

Juraska, M., & Juraska, M. M. (2010). Package ‘speff2trial’. *R software*.

Kalbfleisch, J. D., & Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data* (2nd ed.), Hoboken, NJ: Wiley.
Lawless, J. F. (2011). *Statistical models and methods for lifetime data* (Vol. 362). John Wiley & Sons.

Leung, K. M., Elashoff, R. M., & Afifi, A. A. (1997). Censoring issues in survival analysis. *Annual Review of Public Health, 18*, 83-104.

Lu, Z., & Zhang, W. (2012). “Semiparametric Likelihood Estimation in Survival Models With Informative Censoring,” *Journal of Multivariate Analysis, 106*, 187-211.

National Heart, Lung, and Blood Institute. (2017). Atherosclerosis Risk in Communities Study (ARIC). *Operations Manual, 4*.

Nocedal, J., & Wright, S. (1999). *Numerical optimization*. Springer.

Prinja, S., Gupta, N., & Verma, R. (2010). Censoring in clinical trials: review of survival analysis techniques. *Indian Journal of Community Medicine: Official Publication of Indian Association of Preventive & Social Medicine, 35*, 217.

Ranganathan, P., & Pramesh, C. S. (2012). Censoring in survival analysis: Potential for bias. *Perspectives in Clinical Research, 3*, 40.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rossini, A. J., & Tsiatis, A. A. (1996). A semiparametric proportional odds regression model for the analysis of current status data. *Journal of the American Statistical Association, 91*, 713-721.

RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA.
Scharfstein, D. O., & Robins, J. M. (2002). Estimation of the failure time distribution in the presence of informative censoring. *Biometrika, 89*, 617-634.

Song, X., Davidian, M., & Tsiatis, A. A. (2002). An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics, 3*, 511-528.

Song, X., & Ma, S. (2008). Multiple augmentation for interval-censored data with measurement error. *Statistics in Medicine, 27*, 3178-3190.

Song, X., & Wang, L. (2017). Partially time-varying coefficient proportional hazards models with error-prone time-dependent covariates—an application to the AIDS Clinical Trial Group 175 data. *The Annals of Applied Statistics, 11*, 274-296.

Sparling, Y. H., Younes, N., Lachin, J. M., & Bautista, O. M. (2006). Parametric survival models for interval-censored data with time-dependent covariates. *Biostatistics, 7*, 599-614.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*, 267-288.

Wen, C. C. (2012). Cox regression for mixed case interval-censored data with covariate errors. *Lifetime Data Analysis, 18*, 321-338.

Wen, C. C., & Chen, Y. H. (2014). Functional inference for interval-censored data in proportional odds model with covariate measurement error. *Statistica Sinica, 24*, 1301-1317.
Zeng, D., Mao, L., & Lin, D. Y. (2016). Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika, 103*, 253-271.