Multivariate Survival Analysis of Macrovascular Events in Veterans with Type 2 Diabetes

BY
DERRICK G. KAUFMAN
BS (University of Illinois at Chicago) 2005
MS (University of Illinois at Chicago) 2008

THESIS
Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Public Health Sciences in the Graduate College of the University of Illinois at Chicago, 2014

Chicago, Illinois

Defense Committee:
Hua Yun Chen, Chair and Advisor, Biostatistics
Hui Xie, Biostatistics
Sally Freels, Biostatistics
Dulal Bhaumik, Biostatistics
Nicholas Emanuele, Loyola University of Medicine
ACKNOWLEDGMENTS

Many thanks to my advisor Hua Yun Chen for his patience, encouragement, and inspiration. To my employer Domenic Reda for supporting me in obtaining a doctoral degree and for his editorial comments on this document. To Dulal Bhaumik for listening and dispensing valuable advice. To Nicholas Emanuele for his professionalism and kindness. To my committee members Hui Xie and Sally Freels for their feedback and service on my defense committee. To my friend and coworker Kush Kapur for listening to my ideas and providing feedback. To my friend and colleague Kimberly Carlson for encouragement and support. To Liliana Aceves for her administrative support and tips on obtaining funding. To Ronald Hershow for supporting me to attend JSM in 2012. To Kevin Martin and Mark Ezzo of the VA Informatics and Computing Infrastructure, without whom I would not have been able to perform my simulations in anything approaching a reasonable amount of time. To the VADT study team for executing the study and collecting the data.

DGK
# TABLE OF CONTENTS

## CHAPTER 1 MOTIVATING EXAMPLE

1.1 Veterans Affairs Diabetes Trial .................................................. 1
1.1.1 Design .................................................................................. 1
1.1.2 Primary and Secondary Outcomes ........................................... 2
1.1.3 Results .................................................................................. 2
1.2 Veterans Affairs Diabetes Trial Follow-up Study ............................ 4
1.2.1 Design .................................................................................. 4
1.2.2 Primary and Secondary Outcomes ........................................... 4
1.3 Rosiglitazone Controversy ............................................................... 5
1.4 Analysis Challenges ...................................................................... 6

## CHAPTER 2 LITERATURE REVIEW

2.1 Survival Analysis .......................................................................... 8
2.2 Competing Risks .......................................................................... 9
2.3 Multivariate Models ....................................................................... 13
2.4 Semi-competing Risks ................................................................. 14
2.5 Recurrent Events ......................................................................... 20
2.6 Joint Modeling ............................................................................ 22

## CHAPTER 3 PARAMETRIC CONDITIONAL MODEL

3.1 The Problem ................................................................................. 25
3.1.1 Notation ................................................................................. 25
3.1.2 Joint Model A ......................................................................... 26
3.1.3 Conditional Model B ............................................................... 27
3.1.4 Goal ....................................................................................... 28
3.1.5 Likelihood Components .......................................................... 29
3.2 Expectation Maximization Algorithm .............................................. 32
3.2.1 Likelihood Notation ............................................................... 32
3.2.2 Expectation Step .................................................................... 33
3.2.3 Objective Function ................................................................. 34
3.2.4 Gauss-Laguerre Quadrature ..................................................... 35
3.2.5 Expectation Step .................................................................... 43
3.2.6 Maximization Step ................................................................. 44
3.3 Variance Estimation ..................................................................... 44
3.4 Complication-free Survival ........................................................... 47
3.4.1 Survival Curve ....................................................................... 47
3.4.2 New Proposed Test Statistic ..................................................... 48
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4.3</td>
<td>Simulation Approach</td>
</tr>
<tr>
<td>3.4.4</td>
<td>Wald Test</td>
</tr>
<tr>
<td>3.4.5</td>
<td>Marginal Survival</td>
</tr>
<tr>
<td><strong>4</strong></td>
<td>PARAMETRIC CONDITIONAL MODEL: SIMULATIONS</td>
</tr>
<tr>
<td>4.1</td>
<td>Exponential Distribution</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Log-likelihood</td>
</tr>
<tr>
<td>4.2</td>
<td>Quadrature Implementation</td>
</tr>
<tr>
<td>4.3</td>
<td>Empirical Distribution of the Minimum</td>
</tr>
<tr>
<td>4.4</td>
<td>Simulations</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Simulation Results</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Survival Curves</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Test Statistic</td>
</tr>
<tr>
<td><strong>5</strong></td>
<td>PARAMETRIC CONDITIONAL MODEL: VETERANS AFFAIRS DIABETES TRIAL ANALYSIS</td>
</tr>
<tr>
<td>5.1</td>
<td>Parametric Distribution</td>
</tr>
<tr>
<td>5.2</td>
<td>Description of the Data</td>
</tr>
<tr>
<td>5.3</td>
<td>Exponential Model</td>
</tr>
<tr>
<td>5.4</td>
<td>Weibull Model</td>
</tr>
<tr>
<td>5.5</td>
<td>Discussion of Analysis and Method</td>
</tr>
<tr>
<td><strong>6</strong></td>
<td>THREE-LEVEL HIERARCHICAL PROPORTIONAL-HAZARDS MODEL</td>
</tr>
<tr>
<td>6.1</td>
<td>Three-level Hierarchical Proportional-Hazards Model</td>
</tr>
<tr>
<td>6.2</td>
<td>New Proposed Three-level Hierarchical Proportional-Hazards Model</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Defining Time</td>
</tr>
<tr>
<td>6.2.2</td>
<td>Log-likelihood</td>
</tr>
<tr>
<td>6.2.3</td>
<td>Expectation Conditional Maximization Algorithm</td>
</tr>
<tr>
<td>6.2.4</td>
<td>Expectation Step</td>
</tr>
<tr>
<td>6.2.5</td>
<td>Conditional Maximization Step</td>
</tr>
<tr>
<td>6.3</td>
<td>Estimating the Variance-covariance Matrix</td>
</tr>
<tr>
<td>6.4</td>
<td>Wald Test</td>
</tr>
<tr>
<td><strong>7</strong></td>
<td>THREE-LEVEL HIERARCHICAL PROPORTIONAL-HAZARDS MODEL: VETERANS AFFAIRS DIABETES TRIAL ANALYSIS</td>
</tr>
<tr>
<td>7.1</td>
<td>Descriptive Statistics</td>
</tr>
<tr>
<td>7.2</td>
<td>Model Assumptions and Notation</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Full Model</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Reduced Model</td>
</tr>
<tr>
<td>7.3</td>
<td>Supplementary Four-Dimensional Analysis</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS (continued)

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.4</td>
<td>Discussion of Analysis</td>
</tr>
<tr>
<td>8</td>
<td>THREE-LEVEL HIERARCHICAL PROPORTIONAL-HAZARDS MODEL: SIMULATIONS</td>
</tr>
<tr>
<td>8.1</td>
<td>Data Generation</td>
</tr>
<tr>
<td>8.2</td>
<td>Simulation Results</td>
</tr>
<tr>
<td>8.3</td>
<td>Discussion of Simulation Results</td>
</tr>
<tr>
<td>APPENDIX</td>
<td></td>
</tr>
<tr>
<td>CITED LITERATURE</td>
<td></td>
</tr>
<tr>
<td>VITA</td>
<td></td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. SIMULATION RESULTS MODEL A</td>
<td>61</td>
</tr>
<tr>
<td>II. SIMULATION RESULTS MODEL B</td>
<td>62</td>
</tr>
<tr>
<td>III. SIMULATION RESULTS MODEL C</td>
<td>63</td>
</tr>
<tr>
<td>IV. SIMULATION RESULTS MODEL D</td>
<td>65</td>
</tr>
<tr>
<td>V. MARGINAL UNIVARIATE MODELS</td>
<td>74</td>
</tr>
<tr>
<td>VI. VADT RESULTS EXPONENTIAL</td>
<td>75</td>
</tr>
<tr>
<td>VII. VADT RESULTS WEIBULL</td>
<td>78</td>
</tr>
<tr>
<td>VIII. FREQUENCY OF VADT PATIENTS WITH MULTIPLE EVENTS</td>
<td>80</td>
</tr>
<tr>
<td>IX. NUMBER OF EVENTS BY TREATMENT GROUP</td>
<td>102</td>
</tr>
<tr>
<td>X. FULL MODEL</td>
<td>105</td>
</tr>
<tr>
<td>XI. REDUCED MODEL</td>
<td>107</td>
</tr>
<tr>
<td>XII. PRIMARY OUTCOMES GROUPED MODEL</td>
<td>109</td>
</tr>
<tr>
<td>XIII. MEAN (STD) NUMBER OF EVENTS PER DATASET</td>
<td>114</td>
</tr>
<tr>
<td>XIV. SIMULATED PARAMETER VALUES BASED ON VADT</td>
<td>115</td>
</tr>
<tr>
<td>XV. SIMULATION RESULTS SCENARIOS A1 AND A2</td>
<td>116</td>
</tr>
<tr>
<td>XVI. SIMULATION RESULTS SCENARIOS B1 AND B2</td>
<td>116</td>
</tr>
<tr>
<td>XVII. SIMULATION RESULTS MODELS C1 AND C2</td>
<td>118</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. VADT: Complication-free Survival.</td>
<td>3</td>
</tr>
<tr>
<td>2. Model A: Complication-free Survival.</td>
<td>66</td>
</tr>
<tr>
<td>3. Model B: Complication-free Survival.</td>
<td>67</td>
</tr>
<tr>
<td>4. Model A: Marginal Y Survival.</td>
<td>68</td>
</tr>
<tr>
<td>5. Model B: Marginal Y Survival.</td>
<td>69</td>
</tr>
<tr>
<td>6. Model A: Empirical CDF of T.</td>
<td>71</td>
</tr>
<tr>
<td>7. Model B: Empirical CDF of T.</td>
<td>72</td>
</tr>
<tr>
<td>8. VADT: Exponential Model, Complication-free Survival.</td>
<td>76</td>
</tr>
<tr>
<td>9. VADT: Exponential Model, Marginal CV-death Survival.</td>
<td>77</td>
</tr>
</tbody>
</table>
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amp</td>
<td>Amputation for Ischemic Gangrene</td>
</tr>
<tr>
<td>Ang</td>
<td>Angina</td>
</tr>
<tr>
<td>A-G</td>
<td>Andersen-Gill</td>
</tr>
<tr>
<td>CAD</td>
<td>Coronary Artery Disease</td>
</tr>
<tr>
<td>CAR</td>
<td>Coarsening at Random</td>
</tr>
<tr>
<td>CDF</td>
<td>Cumulative Distribution Function</td>
</tr>
<tr>
<td>CHF</td>
<td>Congestive Heart Failure</td>
</tr>
<tr>
<td>CLI</td>
<td>Critical Limb Ischemia</td>
</tr>
<tr>
<td>CV</td>
<td>Cardiovascular</td>
</tr>
<tr>
<td>ECM</td>
<td>Expectation Conditional-Maximization</td>
</tr>
<tr>
<td>EDF</td>
<td>Empirical Distribution Function</td>
</tr>
<tr>
<td>EPP</td>
<td>Events Per Patient</td>
</tr>
<tr>
<td>EM</td>
<td>Expectation-Maximization</td>
</tr>
<tr>
<td>IC</td>
<td>Intermittent Claudication</td>
</tr>
<tr>
<td>ICR</td>
<td>Invasive Cerebrovascularization</td>
</tr>
<tr>
<td>IE</td>
<td>Intermediate Event</td>
</tr>
<tr>
<td>Int</td>
<td>Intensive Glycemic Control Treatment</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>IR</td>
<td>Invasive Revascularization</td>
</tr>
<tr>
<td>IPCW</td>
<td>Inverse Probability of Censoring Weighted</td>
</tr>
<tr>
<td>IPR</td>
<td>Invasive Peripheral Revascularization</td>
</tr>
<tr>
<td>ISD</td>
<td>Integrated Squared Difference</td>
</tr>
<tr>
<td>MI</td>
<td>Myocardial Infarction</td>
</tr>
<tr>
<td>MLE</td>
<td>Maximum Likelihood Estimate</td>
</tr>
<tr>
<td>MO</td>
<td>Marshall-Olkin</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>PE</td>
<td>Prior Cardiovascular Event</td>
</tr>
<tr>
<td>PPE</td>
<td>Proportion of Patients who had an Event</td>
</tr>
<tr>
<td>PH</td>
<td>Proportional-Hazards</td>
</tr>
<tr>
<td>Str</td>
<td>Stroke</td>
</tr>
<tr>
<td>TE</td>
<td>Terminal Event</td>
</tr>
<tr>
<td>TIA</td>
<td>Transient Ischemic Attack</td>
</tr>
<tr>
<td>VADT</td>
<td>Veterans Affairs Diabetes Trial</td>
</tr>
<tr>
<td>WRT</td>
<td>With Respect To</td>
</tr>
</tbody>
</table>
SUMMARY

Many of the most challenging diseases in medicine today are multifaceted. Type 2 diabetes affects multiple components within the body such as the cardiovascular (CV), peripheral-vascular, and cerebrovascular systems. Within each system there are multiple points of failure, and any given treatment may not affect all system components the same way. Researchers are often not sure which components are most sensitive to change and tend to group these components into composite outcomes, because there is a lack of statistical methodology that can address multiple outcomes that produces inferences that are easy to interpret.

In this dissertation, two modeling approaches, one parametric conditional model and the other random effects proportional hazards model, are utilized for the analysis of multiple types of events, possibly recurrent, from the Veterans Affairs Diabetes trial (VADT). The VADT is a trial of type 2 diabetic patients randomized to intensive or standard glycemic control therapy. The primary outcome of the study was time-to-first of any one of nine macrovascular events.

Both methods strive to minimize information loss due to grouping or discarding data and to control the computational complexity involved in handling nine types of repeated macrovascular events.

In the first modeling approach, a parametric conditional model is used to model three distinct types of events: myocardial infarction (MI), congestive heart failure (CHF), and cardiovascular death (CV-death). Only the time to the first event of each of the three event types
SUMMARY (continued)

was used in the conditional model. The results did not show a significant treatment effect but
did demonstrate a positive dependence of time to CV-death on time to CHF event.

In the second modeling approach, a three-level hierarchical proportional-hazards model
was developed and applied to analyzing all eight recorded primary events (two of the nine were
combined into one as per the protocol). Analyzed in the presence of important covariates, no
treatment effect was found in any of the eight types of events. One covariate, prior CV-event
was found significant in all types of events, baseline insulin was significant in predicting CHF
and invasive revascularization (IR).

In conclusion, different modeling approaches are used to analyze the VADT data to
examine if there might be a treatment effect that went undetected in the initial analysis due
to discarding information about recurrent events. However, even when analyzing all available
data, no treatment effect was found. Some insight is gained on the relationship between different
types of events and death.
CHAPTER 1

MOTIVATING EXAMPLE

In this chapter, I describe the problem presented by the analysis of data from a clinical trial in type 2 diabetic patients, the motivating example for this dissertation. In Section 1.1 I describe in detail the first phase of the clinical trial and corresponding data. In Section 1.2 I describe the second phase of that trial, the long-term follow-up phase. In Section 1.3 I give background information about a problem with one of the study drugs. In Section 1.4 I describe the characteristics of the data and associated analytical challenges presented by the study’s design.

1.1 Veterans Affairs Diabetes Trial

1.1.1 Design

The purpose of this trial was to evaluate the efficacy of intensive glycemic control versus standard glycemic control in preventing macrovascular complications in veterans with poorly controlled type 2 diabetes (4). In this two-arm clinical trial patients were randomized to either intensive or standard glycemic control while being stratified by site, baseline history of CV disease and baseline insulin use. Treatment algorithms consisting of oral agents and insulin were specified in the protocol. The algorithms determined what drugs and in what dosages they should be prescribed to achieve a target glycated hemoglobin level (7% intensive arm, 8.5% standard arm). The goal was to achieve a between-group separation of 1.5 percentage
points in mean hemoglobin A1c between the intensive and standard treatment groups. After six months this goal was achieved with a level of 8.4% in the standard arm and 6.9% in the intensive arm. The median follow-up was 5.6 years.

### 1.1.2 Primary and Secondary Outcomes

The primary outcome was the time-to-first occurrence of any one of nine CV events: MI, new or worsening CHF, IR procedures, inoperable coronary artery disease (CAD), stroke, surgical intervention for cerebrovascular or peripheral vascular disease, amputation for ischemic gangrene and death from CV causes. Events were adjudicated by an endpoint committee (three medical doctors) blinded to treatment. Secondary CV outcomes included new or worsening angina, new transient ischemic attacks (TIA), new intermittent claudication (IC), new critical limb ischemia (CLI), and death from any cause. Non-CV secondary outcomes included microvascular complications: retinopathy, nephropathy, and neuropathy. Adverse events such as severe hypoglycemic episodes were monitored.

### 1.1.3 Results

Between December 1, 2000 and May 30, 2003, 1,791 patients were enrolled in the trial. Follow-up ended on May 30, 2008. The median follow-up time was 5.6 years. The study sample size was based on detecting a 21% difference in complication-free survival proportion during an average six-year-long follow-up period. Using an event rate of 40% in the standard group and 31.6% in the intensive group, a two-tailed hypothesis test with $\alpha=.05$ and $\beta=.14$, seven interim analyses, and a delayed treatment effect beginning in the third year, the required sample size was 850 per group. The primary results were reported in January of 2009 (4). The observed
composite event rate was 33.5% in the standard group and 29.5% in the intensive group, a relative reduction of 11.9% (hazard ratio, .88; 95% confidence interval, .74 to 1.05; p=.14). A log-rank test resulted in p=.14. There were no significant differences in individual components of the primary macrovascular outcomes.

Figure 1. VADT: Complication-free Survival.
1.2 Veterans Affairs Diabetes Trial Follow-up Study

1.2.1 Design

After follow-up ended in 2008, all responsibility for care and treatment is shifted to the patient’s primary care physician and the intervention is terminated. Toward the end of the VADT, patients were asked to consent to participation in the long-term follow-up study (ClinicalTrials.gov Identifier: NCT00756613). Those patients who consented would be followed via electronic central database searches and an annual mail-in survey questionnaire designed to assess quality-of-life and self-reported events. The purpose of this observational follow-up study was stated in the protocol as to determine the long-term effect of intensive glycemic control of type 2 diabetes on major CV complications. One major question that the investigators posed in the protocol was on the nature of metabolic memory, whether or not there is a delayed effect of intensive treatment after the transition to usual care. Secondary outcomes include: CV mortality, major microvascular complications, health-related quality of life, and total mortality. About 1,030 of the original 1,791 patients consented to the follow-up study at the time of writing.

1.2.2 Primary and Secondary Outcomes

The original nine-component composite outcome has been reduced to five components (nonfatal MI, nonfatal stroke, new CHF, amputation due to ischemic diabetic gangrene, and CV-related death). This more restrictive outcome for this long-term follow-up study is due to the fact that some of the outcomes from the original VADT are difficult to discern (worsening of CHF) or are more arbitrary (e.g., cardiac procedures) given the available data from centralized
database searches and the absence of on-study clinical visits. The proposed analysis was to combine the VADT main study data and the follow-up data and analyze the primary endpoint under the new definition for differences between treatment groups using a log-rank test. As of writing, data collection for this study is still ongoing and therefore will not be analyzed in this dissertation.

1.3 Rosiglitazone Controversy

In July of 2007 Nissen and Wolski published the results of a meta-analysis (56) that suggested Rosiglitazone may increase the risk of MI (p=.03) and CV-related death (p=.06). Ninety-five percent of VADT study patients were prescribed rosiglitazone as the second step of two primary oral agents specified in the protocol for lowering hemoglobin A1c levels in both standard and intensive groups. It was generally prescribed at higher doses in the intensive group as compared to the standard group. This raised the question of whether or not the treatment under study could decrease the risk of certain components of the composite outcome while simultaneously increasing the risk of other components. In the VADT final analysis some components of the primary composite endpoint showed a nonsignificant trend toward a benefit of standard treatment (CV-death: Intensive=64, Standard=78, log-rank, p=.26; Stroke: Intensive=36, Standard=28). At the same time, the total number of events for other components of the primary composite outcome showed nonsignificant trend towards benefit in the intensive group (MI: Intensive=64, Standard=78, log-rank p=.24; CHF: Intensive=76, Standard=82, log-rank p=.24) (25).
1.4 Analysis Challenges

In the VADT nine distinct event types, many of them possibly recurrent, are combined into one composite outcome of complication-free survival time that are compared between the two treatment arms. This method, while simple, ignores information provided by the order of event times of different types as well as recurrent events of the same type. Furthermore, by redefining the endpoint for the follow-up study to a smaller subset of event types the investigators ignore information provided by the endpoints collected in the main study but not in the follow-up study.

The unfortunate act of discarding information provided by recurrent events and by events that were collected in the first but not the second phase of a study appears to be due to a lack of suitable statistical methodology. With modern electronic medical records systems, data on clinical outcomes are plenty but information on events is coarsened without additional data provided by an in-person clinical visit. Extending a controlled clinical trial by collecting data via centralized databases is an attractive option both in terms of reduced costs and of the potential for answering questions about long-term effects of treatments. There appears to be a gap in methodology that would allow us to incorporate the full data from the first phase with the reduced data from the second phase into one analysis.
The use of composite outcomes in survival analysis is popular because it increases the event rate and consequently increases statistical power. However, difficulties of interpretation arise when components of varying importance are grouped together, particularly when treatment effect is higher for less important components than for more important ones (31) (69). In the case of the motivating example treatment effect estimates move in opposite directions in two distinct groups of component outcomes (MI and CHF in favor, stroke, and CV-death against).

The literature is full of examples of interventions that were suspected to harm as well as benefit the patient such as Vioxx (rofecoxib) (55), Natrecor (nesiritide) (66) (65), and Avandia (rosiglitazone) (56). Composite outcomes may mask such harmful effects. Mechanistic knowledge of intervention pathways is lacking in medicine and statistical modeling may provide answers (3). Multivariate modeling of the composite components is an attractive alternative to modeling of univariate composite outcomes because it can better explain the biological mechanism of the treatment effect and can detect when treatment effects move in opposite directions for different outcomes. Also, multivariate models explicitly account for correlation between components, leading to smaller standard errors and increased statistical power.
CHAPTER 2

LITERATURE REVIEW

In this chapter I review the literature relevant to the data presented by the VADT. In Section 2.1 I review some literature on univariate survival analysis and note the limitations of commonly used methods. In Section 2.2 I review the literature on analysis of competing risks data, a specific kind of multivariate survival data. In Section 2.3 I review literature on multivariate frailty models. In Section 2.4 I review literature on a topic that most closely fits the structure of the data of my motivating example, semi-competing risks. In Section 2.5 I review literature on recurrent events data, which is another feature of my motivational example, but which is not often taken into account simultaneously with semi-competing risks. In Section 2.6 I review literature on joint modeling in survival analysis which is a useful tool in the analysis of multivariate data of different types.

2.1 Survival Analysis

The distinguishing feature of survival data is censoring. Many statistical methods for analysis of survival data have been developed in the second half of the last century. Kaplan and Meier (41) proposed the now well-known Kaplan-Meier estimator for nonparametrically estimating a survival curve for right-censored data. Many nonparametric tests for comparing survival distributions, such as the log-rank test and the linear rank test, were also proposed by Gehan (35), Mantel (49), and Peto and Peto (62). Another development is the introduction of
proportional hazards regression model by Cox (18),(17), which models the effects of covariates on the survival time by a semiparametric model. The Cox’s model has been widely used in survival time data analysis. Theories for statistical inference under the nonparametric and semiparametric model had been fully developed and extended. In comparison to the nonparametric and semiparametric models, parametric models by Epstein and Sobel (29) offer more efficiency in detecting group differences when the data follow the specified parametric distribution. Theory for statistical inference under most of the parametric models is simple and straightforward. These models and methods, initially developed for the analysis of univariate survival time with right-censored data, have been extended to handle interval-censored data (60).

The key assumption that enables the analysis of censored data using the above-mentioned methods is that of independence between the survival time and the censoring time. When this assumption fails, the survival time distribution may not be identifiable from the observed censored data. This requires special attention in the analysis of censored survival time. Another complication is that when more than one type of failure can occur or the same type of events can repeatedly occur in the same individual, the failure times may not be independent. In this case, association among a group of survival times need to be taken into consideration in the survival time analysis. More complicated models and methods of analysis are required.

2.2 Competing Risks

In medical research, patients are often at risk for more than one failure type and the occurrence of one failure type may effectively censor all others. When two or more types of failures are not independent and the survival time of one particular failure type is of interest, the
censoring time, i.e., the minimum of all other failure times, is not independent of the survival time of the failure type under consideration. Those correlated failure times that are mutually censoring are known in the literature as “competing risks.” Since the common survival function is not identifiable in the presence of competing risks, other qualities have been defined and estimated in the competing risk model. Klein and Moeschenerger (42) describe crude, net, and partial-crude probabilities for competing risks. Crude probability is the probability of an event of a given type in the presence of all other risk types. Net probability is a hypothetical probability of an event of a given type in the absence of all other risk types. Partial-crude, a combination of crude and net is the probability of an event of a given type given the removal of all other types. Fine and Gray (32) have provided methods for estimating the effect of a covariate on cause-specific hazard functions.

In their classic paper Prentice et al. (63) describe three major problems in competing risks as (1) inference on the effects of covariates on types of failure, (2) interrelation among failure types, and (3) estimation of failure rates of one type given the removal of all other types. They describe the cause-specific hazard functions as “the basic estimable quantities in the competing risks framework.” They are defined as the instantaneous failure rate at time $t$ of failure type $j$ in the presence of all other failure types, and are estimable with univariate analysis procedures treating failures of type $k, k \neq j$ as censored.

The authors also address problems with the latent failure time formulation of competing risks where each failure type is associated with its own latent failure time and the observed data are the minimum of the latent failure times in each individual. Latent failure times are
sometimes described as the failure time of cause $j$ given the removal of all other causes. These latent failure times can be independent of or dependent on one another in practice. However difficulties arise in distinguishing between independence and dependence in the competing risks framework. In the VADT data, it is feasible that the different failure times of the components of the composite outcome would be dependent because they are all macrovascular events from the same individual. Tsiatis (70) has shown that the independent nonparametric model is not distinguishable from an infinite number of dependent nonparametric models when only the time until first event and event-type is observed. This implies the attempt to estimate the net survival function nonparametrically will fail in the presence of competing risks. Crowder (19) has summarized problems of identifiability in fully parametric models of competing risks and investigates systems which survive one or more failures.

Another semi-parametric approach suggested by Prentice et al. (63) is using Cox’s proportional hazards model with time-dependent covariates to describe the association between two different type-specific failure times. They suggest that if two different types of failure are associated with the same time-dependent covariate then they are correlated with each other. Cortese and Andersen (15) outline several methods for using internal time-dependent covariates in the competing risks framework to estimate quantities of interest such as cause-specific hazards and prediction of cumulative incidences and survival probabilities.

Marshall and Olkin (50) describe a bivariate exponential model for competing risks that can be used to model a special kind of dependence among two or more distinct failure types. The model is derived from a “shock model” where three types of fatal or nonfatal
shocks follow Poisson processes. For two types of events, the first, second, and third types of shocks act on the first system component only, the second system component only, and simultaneously on both system components respectively. The model allows for dependence between two system component failure times when that dependence arrives from the possibility of both components failing simultaneously. I considered using this model for the VADT data. However, the underlying mechanism of the dependence of failure times in the VADT is not explained solely by simultaneous occurrence of different events, but rather by the systemic nature of the disease.

While non-identifiability is an issue in semi-parametric models for competing risks, we can always assume a parametric model for the full data. Moeschberger and David 1971 (54) and Moeschberger 1971 (53) have extended the Marshall-Olkin (MO) model to Weibull distribution through a change of variables and to include more than two failure types in the competing risks framework. Emoto and Matthews 1990 (28) derived a different multivariate Weibull distribution through marginal transformations of a Pickands bivariate exponential distribution. Assuming the model fits, although there is no way to test the fit, their model allows for a general dependence among distinct failure types. They use the model for survival data with dependent censoring and failure as two competing risks.

Hougaard (39) points out that the MO model is a three parameter sub-model of a five-parameter homogeneous Markov model for bivariate data. A Markov model is a special multistate model, in which the transition probabilities are assumed to be constant with respect to time.
2.3 Multivariate Models

Clayton (14) introduces the bivariate gamma frailty model for disease incidence in father-son pairs. The model can be derived from two Cox models with a common covariate $X$ that is unknown and assumed to follow a gamma distribution in the population. Their focus lies in estimating an association parameter $\theta$ that is equal to the relative risk for the son if his father has had the event as opposed to if his father has not had the event. The parameter $\theta$ can take on values of 1 for no association, $< 1$ for negative association and $> 1$ for positive association. The parameter $\theta$ is estimated in absence of knowledge of the marginal distributions. The hazard ratio for the son given the father’s event time is equal to the hazard ratio for the father given the son’s event time. This restriction makes the model more appropriate for the same event type among multiple individuals rather than for different event types within the same individual. The model has the property that $F(s, 0) = F_s(s)$ and $F(0, t) = F(t)$. Oakes (57),(59) improves upon Clayton’s estimator for $\theta$ in the bivariate gamma frailty model. The author proposes an estimator based on Kendall’s coefficient of concordance and derives asymptotic properties. He discusses parametric inference where both marginals are exponentially distributed. Oakes (58) describes a semi-parametric approach, treating the marginals as nuisance functions, for estimation of $\theta$.

Hougaard (38) proposes a family of multivariate parametric models in which individuals within a group share a common unobserved covariate, assumed to follow a positive stable distribution. Some convenient properties of proportionality follow when Weibull distribution
is assumed. More recently, Bendeen-Roche and Liang (7) have incorporated parametric frailty into a multivariate competing risks framework.

2.4 Semi-competing Risks

In complex disease research such as cancer and diabetes death is not the only event of interest. Other events of interest include intermediate events (IE), such as metastasis in the case of cancer or nonfatal heart attack in the case of diabetes. These events are naturally correlated with the terminal event (TE). Patients can be followed beyond their IE event but not beyond their TE. So called "illness-to-death" models with shared frailty have been proposed by Xu et al. (72). This feature has also been referred to in the literature as "semi-competing risks" by Fine and Chappell (33). This framework closely fits the VADT data in which there are potentially multiple IEs and one TE per individual. The observation of both TE and IE times within the same individual provides information on the dependence structure between events. Crawley and Hu (20) have used time-to-IE as a time-dependent covariate in a Cox proportional hazards (PH) model for survival time in the analysis of heart transplant data. However, my interest in the VADT study is not solely in the TE and the influence of the IEs on the TE, but also in the IEs themselves, their degree of dependence on one another, and the effect of treatment on all events.

Fine and Chappell (33) approaches the semi-competing risks problem using a semi-parametric gamma frailty model proposed in Clayton 1978 (14), also described as a copula model. However, as Clayton notes, the gamma frailty may be suitable for clustered individuals experiencing the same type of event but not for a single patient experiencing different types
of events. Their goal is estimation of the marginal survivor function of time-to-IE, designated $X$, in the presence of censoring by the TE, whose time is designated $Y$. They demonstrate that the distribution of $(X, Y)$ is nonparametrically identifiable in the “upper wedge” where $X < Y$ while marginal survivor function of $X$ is not. The method is flexible in that it accounts for dependence between two events via the shared frailty and allows for both parametric and nonparametric specification of the marginal distributions. The authors propose a consistent estimator for the gamma frailty parameter $\theta$, with the restriction $\theta > 1$. Using the proposed estimator for $\theta$ and the identifiable product limit estimators of the cumulative distribution functions $F_y$ and $F_z$ ($Z = \min(X, Y)$) they propose a “plug-in estimator” for $F_x$.

While the joint distribution of survival times is not nonparametrically identifiable in the “lower wedge” $X > Y$, as these observations are not observable, I may assume a parametric distribution in the “lower wedge.” This method in general is problematic for the interpretation of the case $X > Y$ (i.e., heart attack occurs after death), which is assumed possible for purposes of model fitting. However, the assumption that the same relationship holds for $X > Y$ as for $X < Y$ is biologically plausible in a systemic disease such as diabetes. I propose that the nature of the dependence stems from the fact that diabetes affects the entire CV system, and the events analyzed are all failures of the macrovascular system, a subsystem of the CV system. Therefore, the dependence stems from the fact that all events under analysis are different types of failures of the same macrovascular system, that each individual has only one system, and that individuals differ in terms of their stage of CV disease and co-morbidities that define the state of that system and thus the probability of experiencing an event. Therefore, I argue that
the order of the events is not important in this application. The occurrence of the IE itself is not the mechanism that increases the probability of a TE, but the underlying state of the system is. Furthermore, the argument is strengthened by the fact that in the VADT, the IE (heart attack) and TE (CV-death) may be physically the same events with different outcomes (i.e., the event being classified as an IE if nonfatal heart attack and a TE if fatal). If a patient suffers a severe heart attack, the chance of fatality is dependent on external as well as internal factors, such as access to medical treatment at the time of onset of symptoms. If a patient receives the appropriate life-saving treatment and the patient survives, what might have been a TE is instead classified as an IE.

Dignam et al. (24) propose a parametric conditional model for modeling semi-competing risk data consisting of a nonterminal event $X$ and a terminal event $Y$ in the presence of covariates $Z$. They leverage the information about the relationship between $X$ and $Y$ observable in the "upper wedge" to extend the parametric model into the unobserved "lower wedge," by treating the “lower wedge” as missing data. Specifically, joint parametric models for $X|Z$ and for $Y|X,Z$ are specified and fitted using the expectation-maximization (EM) algorithm with unobserved $X$ when $X > Y$ treated as missing. Their interest was on modeling $X|Z$ marginally over $Y$ using the information provided by $Y$.

Zeng et al. (73) have taken a pattern-mixture model approach, treating those who experienced the nonterminal event and those who did not as two distinct groups at the outset of the trial, modeled by a Bernoulli random variable. For those who did not experience the nonterminal event $X$, the terminal event $Y|Z$ is modeled nonparametrically. For those who did
experience the nonterminal event $X|Z$ and $Y' = Y - X|X, Z$ is also modeled nonparametrically. This method has the advantage that it automatically implies the order $X < Y$ and $X > Y$ need not be considered. Both approaches of Dignam et al. and Zeng et al. explicitly model the dependence of the different failure types, which may be useful when the dependence is of interest by investigators. However, in the VADT the pattern mixture approach seems inappropriate since the scientific community does not believe there is any mechanism that would induce two distinct populations, one that is destined to experience the IE and one that is not. But rather, there are fatal and nonfatal events and which one happens first is a matter of chance.

Epstein and Munoz (30) propose a bivariate Weibull joint distribution for the IE and TE with the restriction that IE occurs before the TE, if it occurs at all. They argue that models that provide for the possibility that the IE occurs after the TE can distort the relationship and often overestimate the time-to-IE. They use the concept of sub-cumulative distribution for the probability that the IE occurs at time $r$ before the TE occurs at time $d$ may be less than one. They use this model to estimate four quantities of interest: (1) the probability that a patient presents with the IE, (2) the IE time, given that the patient presents the IE, (3) the conditional probability that the patient survives $v$ years after presenting the IE, and (4) the residual survival in the overall population ignoring the IE. The authors fit their model to AIDS data and also describe a nonparametric estimation approach and compares the the results with the parametric approach, and find that the resulting survivor and hazard functions closely agree. While this method appears to side-step the complication of modeling the “lower wedge,”
it is essentially the same as the pattern mixture model above and is inadequate for the same reasons.

Multistate or multiple decrement models have been used to model data where time-to-IE and TE are recorded and the IE may or may not be observed (22). The three "states" can be described as healthy, having experienced the IE, or having experienced the TE with the last as an absorbing state. Aalen (1),(2) proposed a nonparametric estimator for the cumulative hazard of the transition probabilities. Andersen (5) used separate semi-parametric PH models for each of the transition probabilities in a study on nephropathy and mortality in diabetic patients.

Bebchuk and Betensky (8) use a multistate model while their interest is on modeling the minimum of time of IE and time of TE in the absence of IE. These are modeled as three transition probabilities, with the transition from IE to TE modeled with the time-to-IE as a covariate. The hazards are modeled semi-parametrically with local likelihood estimation and an iterative imputation method for right-censored data. This model can handle cases when the IE censor time is different from the TE censor time. In the example, they model the treatment groups separately and compare the estimates as the procedure cannot handle covariates. They suggest that the Cox PH model using local likelihood estimation can be used when covariates are present. In a later paper Bebchuk and Betensky (9) propose a similar model while the interest is on modeling the transition probability from the IE state to the TE state. Without the ability to handle covariates, they propose two distance-based tests for testing equivalence of two survival curves. The first is a Kolmogorov-Smirnov test statistic based on the maximum
difference between survival curves of two groups. The second test statistic is an integrated weighted difference.

To summarize, multistate models are very flexible as different models can be specified for different state transition probabilities. However, if there are more than two events, all pairwise transitions must be modeled. In the VADT there are nine different types of events, but even a multistate model of just three event types would mean six models, and some transitions might not have enough observations to fit their own model. For this reason, multi-state models are too cumbersome for this application.

In a closely related problem, there is much recent research surrounding the problem of selecting biological markers that are predictive of death or disease progression. When the marker is a binary indicator that occurs at time $X$ before the disease progression or terminal event at time $Y$ the problem is similar to semi-competing risks. Gail (34) has proposed using $X$ as a time-dependent covariate in Cox’s proportional-hazards model. Day and Bryant have proposed (21) using bivariate frailty models using biological markers to predict survival.

Lagakos (43) proposes a parametric model to analyze survival time in the presence of a single IE. His interest is in quantifying the relationship between the IE and TE times and thus getting more precise estimates of the marginal distribution of the TE time. The model is built on a bivariate exponential model and related to a semi-Markov model. It assumes that the IE time is exponentially distributed and that the conditional distributions of the time from IE to TE (given the IE occurs) and time to TE (given no IE) are exponential and equal. Perhaps more problematic is the assumption of independence of time from IE to TE and time-to-IE, as
in some medical applications in long-term diseases you would expect those to be correlated as lifespan is limited.

Shen and Thall (67) have built a model for the situation when a patient can experience at most one of possibly many IEs before succumbing to the TE. This means that many IEs can be considered as competing risks. The times to each IE, from each IE to TE and to TE without IE are specified marginally by a three-parameter generalized odds rate model. The time of each IE and subsequent residual survival are combined under a bivariate generalized von Morgenstern distribution. The model accounts for positive or negative association between the time of each nonfatal event and subsequent survival while accommodating covariates and the usual administrative censoring. The model is not suitable if more than one type of IE is observable in one individual because it would require that transition probabilities be modeled of every permutation, and is therefore not suitable for our application.

2.5 Recurrent Events

Most of the research on recurrent events involve modeling the mean of the counting process $E[d\hat{N}(t)|\hat{N}(u), 0 \leq u < t, X(t)]$ called the intensity process. The approaches generally differ by what they condition on: the full $\hat{N}(u), 0 \leq u < t$, just the number of events at time $t$: $\hat{N}(t-)$, or not depending on any aspect of the history of the event process. Many models can be used for the intensity process with some form of Cox PH model the most common.

In addition to semi-competing risks, recurrent events is a distinguishing feature of the VADT data. Recurrent events can be modeled as following a counting process. Andersen and Gill (6) have extended the Cox model to the intensity process of a multivariate counting process.
In this paper, the authors explore the large sample properties of such a model using martingale theory for the proofs of asymptotic normality of the estimators. Miloslavsky et al. 2004 (52) describe a method of modeling recurrent events in the presence of dependent censoring that is based on the Andersen-Gill multiplicative intensity model (6). Briefly, the A-G model is given by

$$E\{dN(t)|\bar{X}(t-)) = Y\lambda(t)\lambda_0(t)\exp(\beta\zeta(t))$$

Here $Y\lambda(t)$ is the at-risk process, $\lambda_0(t)$ is the baseline intensity process, $\zeta(t)$ is a known function of the history data $\bar{X}(t-) = \{\bar{N}(t-), \bar{Z}(t-))$, including history of the event $\bar{N}(t-)$. A distinguishing feature of the A-G model is it does not avoid modeling the dependency of $dN(t)$ on $\bar{N}(t-)$, which other models described in Wei et al. (71), Pepe and Cai (61), Lawless and Nadeau (45) and Lin et al. (46) do avoid. The author proposes a marginal A-G multiplicative intensity model

$$E\{dN(t)|\bar{W}(t-)) = Y\lambda(t)\lambda_0(t)\exp(\beta\zeta(t))$$

This model differs from the previous in that $\bar{W}(t) = \{\bar{N}(t), \bar{Z}^*(t))$ and $\bar{Z}^*(t)$ is part of the full data covariate process $\bar{Z}(t)$. The full data are not observed because of censoring. Assuming coarsening at random (CAR) means that the censoring event given the full data depends only on the observed part of the data ($\lambda_C\{t|\bar{X}(\tau)\} = \lambda_C\{t|\bar{X}(t))\}$ for $t < \tau$ and $\tau$ is time of the end of the study. They specify another A-G multiplicative intensity model for the censoring denoted as $\lambda_C\{t|\bar{X}(t-))$. Most existing methods focus on modeling the observed data counting process intensity $\lambda^*(t)$. Assuming CAR holds, the full data intensity is equal to
the observed data intensity. However, in the case of the marginal A-G multiplicative intensity model, not only must CAR hold but censoring cannot depend on covariates outside of \( \tilde{W}(t) \). Because of this limitation, the authors model the full data intensity instead. The authors establish that the unbiased regression parameters \( \beta \) of the full data counting process model can be obtained from the observed data through the use of inverse probability of censoring weighted (IPCW) mappings of full data estimating functions.

Prentice et al. 1981 (64) utilize a stratified approach, with a different Cox-type model for all patients at risk for each of the ordered events. The stratified models can come in two forms, one where time begins at the beginning of the study for all strata and another where time begins at the occurrence of the previous event for each strata. In the VADT data, where more than two events of a given type are rare, it may be feasible to have just two or three strata for each event type and throw out the rest of information provided by the fourth events and beyond. It also means specifying covariate effects in each stratum despite the fact that I am not interested in covariate effects on specific ordered occurrences of recurrent events. Therefore this method is problematic in that it discards information provided by patients who had more than a few events of a given type, and wastes degrees of freedom estimating parameters that I am not interested in.

2.6 Joint Modeling

Zeng and Lin 2007 (74) have proposed a broad framework of semi-parametric linear transformation models that include the Cox PH model as a special case. Their general model consists of an unspecified link function of \( T \) with a vector of covariates \( Z \) with an error term
from a parametric distribution. They also propose an unspecified transformation function $G(.)$ of the cumulative intensity function of which they consider Box-Cox and logarithmic transformations as examples. This gives the model considerable flexibility to accommodate non-PH functions. The parameters for the $G(.)$ transformation are chosen by some goodness-of-fit criteria such as Akaike information criterion. They establish consistency, asymptotic normality, and asymptotic efficiency through the use of empirical process theory and semi-parametric efficiency theory. They also present methods for computation using the EM algorithm. The framework can accommodate univariate failure times, multivariate failure times, recurrent events, joint longitudinal-survival models, and frailty models. Besides computational complexity, a serious drawback of the methods described is lack of easy interpretation of covariate parameters given the very general nature of the transformation function $G(.)$.

Zeng and Lin 2007 (75) have used this framework to model recurrent events with a univariate normal random effect. Zeng and Lin 2009 (76) have applied the framework to a joint model of death and recurrent hospitalizations in AIDS patients with univariate normal random effect. Zhu et al. 2011 (77) have taken the work of Zeng and Lin 2009 (76) a step further to model multivariate recurrent and terminal events with bivariate normal random effects using two-dimensional Gauss quadrature. Chen 2012 (13) has also implemented the semi-parametric linear transformation models of Zeng and Lin 2007 (74) for the marginal survival distributions while modeling the joint association between semi-competing risks with a copula function.

Liu and Huang 2007 (47) have implemented a joint model of longitudinal repeated measures, recurrent events, and a terminal event. First the mixed model with normal random
effect for the longitudinal measure is specified. Second, the recurrent event model is specified
dependent on the random effect of the longitudinal measure and a second normal random effect.
Third, the model for the terminal event is specified, dependent on both random effects of the
longitudinal and recurrent events models. The hazards for the recurrent and terminal event
processes are modeled piecewise exponential and the joint model is fit by maximum likelihood
estimation using Gaussian quadrature in SAS PROC NLMIXED. They have applied the method
to an example of an AIDS study with repeated longitudinal cluster of differentiation four counts,
recurrent events of opportunistic diseases, and death as outcomes. Many researchers have
utilized joint survival-longitudinal models with both univariate (44) and multiple types of events
(26), (27).
CHAPTER 3

PARAMETRIC CONDITIONAL MODEL

In this chapter I define the model and estimation procedure for modeling survival time for three distinct types of events. In Section 3.1 I define the notation and specify the model and its likelihood function. In Section 3.2 I describe in detail how the EM algorithm is used to estimate the model. In Section 3.3 I describe the estimation of the variance of parameter estimates using Louis’s Formula. In Section 3.4 I specify the survival functions derived from the specified model and define several test statistics useful for comparison between two treatment groups.

3.1 The Problem

3.1.1 Notation

Assuming there are two nonterminal types of failure, let \((X_{1i}, X_{2i}, Y_i)\) be trivariate failure times for patient \(i\) where \(Y_i\) is a terminal failure time for the \(i^{th}\) individual for \(i = 1...n\). Let \(C_i\) be an independent censoring time that follows an unknown distribution. In addition, \(C_{1i}\) is an administrative censoring time at which follow-up of the event associated with \(X_{1i}\), \(C_{2i}\) with \(X_{2i}\), and \(C_{3i}\) with \(Y_i\). In general, define \(Y_i^* = \min(Y_i, C_i, C_{3i})\), \(X_{1i}^* = \min(X_{1i}, C_{1i}, Y_i^*)\), and \(X_{2i}^* = \min(X_{2i}, C_{2i}, Y_i^*)\).
In the motivating example \( \forall i : C_{2i} = C_{3i} \), therefore a special case of the general case above is

\[
C^*_{1i} = \begin{cases} 
C_i & \text{if } C_i < C_{1i} \\
C_{1i} & \text{otherwise}
\end{cases}
\quad \text{and} \quad
C^*_{2i} = \begin{cases} 
C_i & \text{if } C_i < C_{2i} \\
C_{2i} & \text{otherwise}
\end{cases}
\]

where \( C_{1i} < C_{2i} \) and \((C_{1i}, C_{2i})\) are administrative censoring times denoting termination of follow-up at two distinct planned calendar dates for \( X_{1i} \) and \((X_{2i}, Y_i)\) respectively. The quantity \( C_i \) follows an arbitrary density and \( C_i \) is assumed independent of \((X_{1i}, X_{2i}, Y_i)\). Define \( Y^*_i = \min(Y_i, C^*_1) \), \( X^*_2 = \min(X_{2i}, Y^*_i) \) and \( X^*_1 = \min(X_i, C^*_1, Y^*_i) \).

In the general case, define a censoring indicator vector \( \Delta_i = (\delta_{x1i}, \delta_{x2i}, \delta_{yi})^T \) where

\[
\delta_{x1i} = \begin{cases} 
1 & \text{if } X^*_1 = X_{1i} \\
0 & \text{otherwise}
\end{cases}, \quad \delta_{x2i} = \begin{cases} 
1 & \text{if } X^*_2 = X_{2i} \\
0 & \text{otherwise}
\end{cases}, \quad \text{and} \quad \delta_{yi} = \begin{cases} 
1 & \text{if } Y^*_i = Y_i \\
0 & \text{otherwise}
\end{cases}
\]

The observed data are of the form \((W_i^T, Z_i^T)^T\), where \( W_i^T = (X^*_1, X^*_2, Y^*_i, \Delta_i^T) \) Assume that the \( W_i \)'s are independent of each other given the \( Z_i \)'s.

**3.1.2 Joint Model A**

To model

\[
f(x_1, x_2, y|Z) = f_{X_1, X_2}(x_1, x_2|Z) f_{Y|X_1, X_2}(y|x_1, x_2, Z) \quad (3.1)
\]
I specify a parametric model $f_{X_1,X_2}(x_1,x_2|\alpha,Z)$ for the marginal distribution where $\alpha$ is vector of regression parameters corresponding to $Z$. Similarly, I specify a parametric distribution $f_{Y|X_1,X_2}(y|x_1,x_2,\beta,Z)$ for the conditional distribution with regression parameter vector $\beta$. Let $\theta = (\alpha^T,\beta^T)^T$. For the censoring time assume an arbitrary density $g(c|Z)$. Let $S_{X_1,X_2}(x_1,x_2|\theta,Z)$, $S_{Y|X_1,X_2}(y|x_1,x_2,\theta,Z)$, and $S_C(c|Z)$ denote the survivor functions corresponding to the parametric density functions.

### 3.1.3 Conditional Model B

I have the option to model $f(x_1,x_2,y|Z) = f(x_1|Z)f(x_2|x_1Z)$ and substitute into Equation 3.1 to obtain a conditional model

$$f(x_1,x_2,y|Z) = f_{X_1}(x_1|Z,\alpha^T)f_{X_2|X_1}(x_2|x_1,Z,\beta^T)f_{Y|X_1,X_2}(y|x_1,x_2,Z,\gamma^T) \quad (3.2)$$

Equation 3.2 is a special case of Equation 3.1. Let $\theta = (\alpha^T,\beta^T,\gamma^T)^T$. It is worth a note of caution that due to the censoring mechanism of $Y$ on $(X_1,X_2)$ the observed data only allow me to identify the distribution $f(x_1^*,x_2^*,y^*)$. The distribution for $f(x_1,x_2,y)$ is nonparametrically unidentifiable given the observed data but is parametrically identifiable by extrapolation. For the censoring time assume an arbitrary density $g(c|Z)$. Let $S_{X_1}(x_1|\theta,Z)$, $S_{X_2|X_1}(x_2|x_1,\theta,Z)$, $S_{Y|X_1,X_2}(y|x_1,x_2,\theta,Z)$, and $S_C(c|Z)$ denote the survivor functions corresponding to the parametric density functions.
3.1.4 Goal

The model Equation 3.2 allows us to estimate the effects of \( Z \) on \( X_1 \), on \( X_2 \) independent of their effects on \( X_2 \) and on \( Y \) independent of their effects on \( X_1 \) and \( X_2 \). This allows us to see how \( Z \) affects the specified joint distribution and to better understand the biological mechanism of the treatment. Interactions between \( Z \) and the dependence of \( X_2 \) on \( X_1 \) or the dependence of \( Y \) on \( X_1, X_2 \) are also possible. When armed with a fully specified joint distribution, event-specific censoring mechanisms of the form described in Sections 1.2.2 and 3.1.1 can be handled with missing data techniques.

Once a joint distribution is specified and model parameters are estimated questions about the effects of \( Z \) on the marginal distributions of \( X_1, X_2 \) or \( Y \) or on complication-free survival time (\( \min(X_1, X_2, Y) \)) may be addressed. For questions about complication-free survival time, define \( M_i = \min(X_{1i}, X_{2i}, Y_i) \) where \( (X_{1i}, X_{2i}, Y_i) \) follows a parametric distribution as described in Equation 3.2. If \( Z \) is a treatment indicator, let \( g(m_i|Z_i) \) be the distribution of \( M_i|Z \). Our final tasks are to plot the survival curves for the two treatment groups and to test the hypothesis: \( H_0 : g(m|Z = 1) = g(m|Z = 0) \) versus the alternative \( H_a : g(m|Z = 1) \neq g(m|Z_i = 0) \).
3.1.5 Likelihood Components

Assume the more general model given in Equation 3.1

1. If all three events are observed ($\Delta_i=(1,1,1)$):

$$\tilde{L}_1(\theta) = S_C(Y^*_i | Z_i) f_{X_1, X_2}(X^*_1, X^*_2 | \theta, Z_i) f_{Y|X_1, X_2}(Y^*_i | X^*_1, X^*_2, \theta, Z_i)$$

$$= S_C(Y^*_i | Z_i) L_1(\theta)$$

2. If $X_{2i}$ and $Y_i$ are observed but $X_{1i}$ is not ($\Delta_i=(0,1,1)$):

$$\tilde{L}_2(\theta) = S_C(Y^*_i | Z_i) \int_{X^*_1}^{\infty} f_{X_1, X_2}(r, X^*_2 | \theta, Z_i) f_{Y|X_1, X_2}(Y^*_i | r, X^*_2, \theta, Z_i) dr$$

$$= S_C(Y^*_i | Z_i) L_2(\theta)$$

3. If $X_{1i}$ and $Y_i$ are observed but $X_{2i}$ is not ($\Delta_i=(1,0,1)$ and $X^*_2 = Y^*_i$):

$$\tilde{L}_3(\theta) = S_C(Y^*_i | Z_i) \int_{X^*_2}^{\infty} f_{X_1, X_2}(X^*_1, s | \theta, Z_i) f_{Y|X_1, X_2}(Y^*_i | X^*_1, s, \theta, Z_i) ds$$

$$= S_C(Y^*_i | Z_i) L_3(\theta)$$
4. If $Y_i$ is observed but $X_{1i}$ and $X_{2i}$ are not ($\Delta_i=(0,0,1)$ and $X_{2i}^{*}=Y_{i}^{*}$)):

$$
\tilde{L}_{4i}(\theta) = S_C(Y_{i}^{*}|Z_{i}) \int_{X_{2i}}^{\infty} \int_{X_{1i}}^{\infty} f_{X_{1},X_{2}}(r,s|\theta,Z_{i}) f_{Y|X_{1},X_{2}}(Y_{i}^{*}|r,s,\theta,Z_{i}) \, dr \, ds \\
= S_C(Y_{i}^{*}|Z_{i}) \tilde{L}_{4i}(\theta)
$$

5. If $X_{1i}$ is observed but $X_{2i}$ and $Y_i$ are not ($\Delta_i=(1,0,0)$ and $X_{2i}^{*}=Y_{i}^{*}$)):

$$
\tilde{L}_{5i}(\theta) = g(Y_{i}^{*}|Z_{i}) \int_{X_{2i}}^{\infty} \int_{X_{1i}}^{\infty} f_{X_{1},X_{2}}(X_{1i}^{*},s|\theta,Z_{i}) f_{Y|X_{1},X_{2}}(t|X_{1i}^{*},s,\theta,Z_{i}) \, dt \, ds \\
= g(Y_{i}^{*}|Z_{i}) \int_{X_{1i}}^{\infty} f_{X_{1},X_{2}}(X_{1i}^{*},s|\theta,Z_{i}) S_y|X_{1},X_{2}(Y_{i}^{*}|X_{1i}^{*},s,\theta,Z_{i}) \, ds \\
= g(Y_{i}^{*}|Z_{i}) \tilde{L}_{5i}
$$

6. If $X_{2i}$ is observed but $X_{1i}$ and $Y_i$ are not ($\Delta_i=(0,1,0)$):

$$
\tilde{L}_{6i}(\theta) = g(Y_{i}^{*}|Z_{i}) \int_{X_{1i}}^{\infty} \int_{X_{2i}}^{\infty} f_{X_{1},X_{2}}(r,X_{2i}^{*}|\theta,Z_{i}) f_{Y|X_{1},X_{2}}(t|r,X_{2i}^{*},\theta,Z_{i}) \, dt \, dr \\
= g(Y_{i}^{*}|Z_{i}) \int_{X_{1i}}^{\infty} f_{X_{1},X_{2}}(r,X_{2i}^{*}|\theta,Z_{i}) S_y|X_{1},X_{2}(Y_{i}^{*}|r,X_{2i}^{*},\theta,Z_{i}) \, dr \\
= g(Y_{i}^{*}|Z_{i}) \tilde{L}_{6i}
$$
7. If $X_{1i}$ and $X_{2i}$ are observed but $Y_i$ is not (\(\Delta_i=(1,1,0)\)):

$$
\hat{L}_{7i}(\theta) = \int_{Y_i}^\infty \cdots f_{X_1, X_2}(X_{1i}, X_{2i} | \theta, Z_i) f_{Y | X_1, X_2}(t | X_{1i}, X_{2i}, \theta, Z_i) \, dt
$$

$$
= g(Y_i | Z_i) f_{X_1, X_2}(X_{1i}, X_{2i} | \theta, Z_i) S_{Y | X_1, X_2}(Y_i | X_{1i}, X_{2i}, \theta, Z_i)
$$

$$
= g(Y_i | Z_i) \hat{L}_{7i}
$$

8. Neither $X_{1i}$ nor $X_{2i}$ nor $Y_i$ is observed (\(\Delta_i=(0,0,0)\) and $X_{2i}^* = Y_i^*$)):

$$
\hat{L}_{8i}(\theta) = \int_{X_1}^\infty \cdots \int_{X_2}^\infty \cdots \int_{Y_i}^\infty \cdots f_{X_1, X_2}(r, s | \theta, Z_i) f_{Y | X_1, X_2}(t | r, s, \theta, Z_i) \, dt \, ds \, dr
$$

$$
= g(Y_i^* | Z_i) \int_{X_1}^\infty \cdots \int_{X_2}^\infty \cdots \int_{Y_i}^\infty \cdots f_{X_1, X_2}(r, s | \theta, Z_i) S_{Y | X_1, X_2}(Y_i^* | r, s, \theta, Z_i) \, ds \, dr
$$

$$
= g(Y_i^* | Z_i) \hat{L}_{8i}
$$

Combining these yields

$$
\hat{L}_i(\theta) \equiv \hat{L}_{7_1}^{1-\delta_{x1i}} \hat{L}_{7_2}^{1-\delta_{x2i}} \hat{L}_{7_3}^{1-\delta_{yi}} \hat{L}_{7_4}^{1-\delta_{x1i}} \hat{L}_{7_5}^{1-\delta_{x2i}} \hat{L}_{7_6}^{1-\delta_{yi}} \hat{L}_{8_1}^{1-\delta_{x1i}} \hat{L}_{8_2}^{1-\delta_{x2i}} \hat{L}_{8_3}^{1-\delta_{yi}} \hat{L}_{8_4}^{1-\delta_{x1i}} \hat{L}_{8_5}^{1-\delta_{x2i}} \hat{L}_{8_6}^{1-\delta_{yi}}
$$
Factoring out \( g(c|Z) \) and \( S_C(y|x_1, x_2, \theta, Z) \) yields a likelihood proportional to

\[
\tilde{\mathcal{L}}_i(\theta) = \mathcal{L}_1^{x_i \delta_x \delta_y \delta_y} \mathcal{L}_2^{(1-\delta_x) \delta_x \delta_y \delta_y} \mathcal{L}_3^{(1-\delta_y) \delta_x \delta_y \delta_y} \mathcal{L}_4^{(1-\delta_x)(1-\delta_y) \delta_y} \mathcal{L}_5^{(1-\delta_x)(1-\delta_y) \delta_y} \ldots
\]

Use the EM algorithm (23) to compute the maximum likelihood estimate (MLE) \( \hat{\theta} \) of \( \theta \) by maximizing \( \mathcal{L}(\theta) = \prod_i \mathcal{L}_i(\theta) \), over \( i = 1 \ldots n \).

### 3.2 Expectation-Maximization Algorithm

Assume Equation 3.2

#### 3.2.1 Likelihood Notation

The complete-data log-likelihood given \((X_1, X_2, Y^*_i)\) is

\[
\ell_i^c = \ln f_{X_1}(X_1|Z_i) + \ln f_{X_2|X_1}(X_2|X_1, Z_i) + \delta_y \ln f_{Y|X_1, X_2}(Y^*_i|X_1, X_2, Z_i) + \ldots
\]

\[
(1-\delta_y) \ln S_{Y|X_1, X_2}(Y^*_i|X_1, X_2, Z_i)
\]

Defining

\[
\ell_1(X_1, X_2, Y^*_i, \theta) = \ln f_{X_1}(X_1|Z_i) + \ln f_{X_2|X_1}(X_2|X_1, Z_i) + \ln f_{Y|X_1, X_2}(Y^*_i|X_1, X_2, Z_i)
\]

\[
\ell_2(X_1, X_2, Y^*_i, \theta) = \ln f_{X_1}(X_1|Z_i) + \ln f_{X_2|X_1}(X_2|X_1, Z_i) + \ln S_{Y|X_1, X_2}(Y^*_i|X_1, X_2, Z_i)
\]

leads to

\[
\ell_i^c = \delta_y \ell_1(X_1, X_2, Y^*_i, \theta) + (1-\delta_y) \ell_2(X_1, X_2, Y^*_i, \theta)
\]
3.2.2 Expectation Step

The goal is to estimate the function $Q(\theta|\theta^{(t)}) = \sum_i Q_i(\theta|\theta^{(t)})$ where

$$Q_i(\theta|\theta^{(t)}) = E_{(X_{1i}, X_{2i}, Y_i)}[\ell_i^1(\theta; (X_{1i}, X_{2i}, Y_i))|W_i, Z_i; \theta^{(t)}]$$

To simplify the notation define the conditional density functions

$$h_1(r|x_2, y; \theta, Z) = \frac{f_{X_1}(r; \theta, Z)f_{X_2|X_1}(x_2|r; \theta, Z)f_{y|x_1,x_2}(y|r, x_2; \theta, Z)}{N_1(x_2, y; \theta, Z)}$$

$$h_2(s|x_1, y; \theta, Z) = \frac{f_{X_1}(s|x_1; \theta, Z)f_{y|x_1,x_2}(y|x_1, s; \theta, Z)}{N_2(x_1, y, \theta, Z)}$$

$$h_3(r, s|y; \theta, Z) = \frac{f_{X_1}(r; \theta, Z)f_{X_2|X_1}(s|r; \theta, Z)f_{y|x_1,x_2}(y|r, s; \theta, Z)}{N_3(y, \theta, Z)}$$

$$h_4(r|x_2, y; \theta, Z) = \frac{f_{X_1}(r; \theta, Z)f_{X_2|X_1}(x_2|r; \theta, Z)S_{y|x_1,x_2}(y|r, x_2; \theta, Z)}{N_4(x_2, y, \theta, Z)}$$

$$h_5(s|x_1, y; \theta, Z) = \frac{f_{X_1}(s|x_1; \theta, Z)f_{y|x_1,x_2}(y|x_1, s; \theta, Z)}{N_5(x_1, y, \theta, Z)}$$

$$h_6(r, s|y; \theta, Z) = \frac{f_{X_1}(r; \theta, Z)f_{X_2|X_1}(s|r; \theta, Z)S_{y|x_1,x_2}(y|r, s; \theta, Z)}{N_6(y, \theta, Z)}$$

where

$$N_1(x_2, y; \theta, Z) = \int_{x_1}^{\infty} f_{X_1}(r; \theta, Z)f_{X_2|X_1}(x_2|r; \theta, Z)f_{y|x_1,x_2}(y|r, x_2; \theta, Z) \, dr$$

$$N_2(x_1, y; \theta, Z) = \int_{x_2}^{\infty} f_{X_1}(s|x_1; \theta, Z)f_{y|x_1,x_2}(y|x_1, s; \theta, Z) \, ds$$

$$N_3(y, \theta, Z) = \int_{x_1}^{\infty} \int_{x_2}^{\infty} f_{X_1}(r; \theta, Z)f_{X_2|X_1}(s|r; \theta, Z)f_{y|x_1,x_2}(y|r, s; \theta, Z) \, ds \, dr$$

$$N_4(x_2, y; \theta, Z) = \int_{x_1}^{\infty} f_{X_1}(r; \theta, Z)f_{X_2|X_1}(x_2|r; \theta, Z)S_{y|x_1,x_2}(y|r, x_2; \theta, Z) \, dr$$

$$N_5(x_1, y; \theta, Z) = \int_{x_2}^{\infty} f_{X_1}(s|x_1; \theta, Z) \, ds$$

$$N_6(y, \theta, Z) = \int_{x_1}^{\infty} \int_{x_2}^{\infty} f_{X_1}(r; \theta, Z)f_{X_2|X_1}(s|r; \theta, Z)S_{y|x_1,x_2}(y|r, s; \theta, Z) \, ds \, dr$$
\[ N_5(x_1, y, \theta, Z) = \int_{x_2}^{\infty} f_{X_1}(x_1; \theta, Z) f_{X_2|X_1}(s|x_1; \theta, Z) S_{y|x_1,x_2}(y|x_1, s; \theta, Z) \, ds \]

\[ N_6(y, \theta, Z) = \int_{x_1}^{\infty} \int_{x_2}^{\infty} f_{X_1}(r; \theta, Z) f_{X_2|X_1}(s|r; \theta, Z) S_{y|y_1,x_2}(y|r, s; \theta, Z) \, ds \, dr \]

are normalizing constants after \( r \) and \( s \) are integrated out.

### 3.2.3 Objective Function

The target function can be written in terms of the log-likelihood components as

\[
Q(\theta|\theta(t)) = \sum_{i: \Delta = (1,1,1)} \ell_{1i}(X_{1i}^*, X_{2i}^*, Y_i^*, \theta) \\
+ \sum_{i: \Delta = (0,1,1)} \int_{X_{1i}^*}^{\infty} \ell_{1i}(r, X_{2i}^*, Y_i^*, \theta) \ h_1(r|X_{2i}^*, Y_i^*; \theta(t), Z) \, dr \\
+ \sum_{i: \Delta = (1,0,1)} \int_{Y_i^*}^{\infty} \ell_{1i}(X_{1i}^*, s, Y_i^*, \theta) \ h_2(s|X_{1i}^*, Y_i^*; \theta(t), Z) \, ds \\
+ \sum_{i: \Delta = (0,0,1)} \int_{X_{1i}^*}^{\infty} \int_{Y_i^*}^{\infty} \ell_{1i}(r, s, Y_i^*, \theta) \ h_3(r, s|Y_i^*; \theta(t), Z) \, dr \, ds \\
+ \sum_{i: \Delta = (1,1,0)} \ell_{2i}(X_{1i}^*, X_{2i}^*, Y_i^*, \theta) \\
+ \sum_{i: \Delta = (0,1,0)} \int_{X_{1i}^*}^{\infty} \ell_{2i}(r, X_{2i}^*, Y_i^*, \theta) \ h_4(r|X_{2i}^*, Y_i^*; \theta(t), Z) \, dr \\
+ \sum_{i: \Delta = (1,0,0)} \int_{Y_i^*}^{\infty} \ell_{2i}(X_{1i}^*, s, Y_i^*, \theta) \ h_5(s|X_{1i}^*, Y_i^*; \theta(t), Z) \, ds \\
+ \sum_{i: \Delta = (0,0,0)} \int_{X_{1i}^*}^{\infty} \int_{Y_i^*}^{\infty} \ell_{2i}(r, s, Y_i^*, \theta) \ h_6(r, s|Y_i^*; \theta(t), Z) \, dr \, ds
\]
3.2.4 Gauss-Laguerre Quadrature

Gauss-Laguerre quadrature is used to approximate integrals of the form

\[ \int_0^\infty x^\alpha e^{-x} f(x) dx \]

with \( M \) nodes \( \{u_{j\alpha}\} \) and weights \( \{v_{j\alpha}\} \) for \( j=1...M \). In my situation I can choose \( \alpha=0 \) or \( \alpha=1 \) and find out which one gives us more accurate results. Let \( \{u_{j0}\} \) and \( \{v_{j0}\} \) denote nodes and weights when \( \alpha=0 \) and \( \{u_{j1}\} \) and \( \{v_{j1}\} \) when \( \alpha=1 \). When time scale of the application is such that the tabulated quadrature nodes do not adequately cover the support of the time distribution, it is necessary to use a scaling factor \( \lambda \).

\( \Delta=(0,1,1) \)

When \( \alpha=0 \)

\[ \int_{X_{1i}}^{\infty} \ell_{1\ell}(r, X_{2i}^*, Y_i^*, \theta) h_{1}(r|X_{2i}^*, Y_i^*; \theta^{(t)}, Z) \, dr \]

Let \( \lambda_{ri}^{(t)} \) be the chosen scaling factor. With the change of variables \( u = \lambda_{ri}^{(t)}(r - X_{1i}^*) \) and \( dr = 1/\lambda_{ri}^{(t)} du \), the integral is equivalent to

\[ = \int_0^{\infty} \ell_{1\ell}(u/\lambda_{ri} + X_{1i}^*, X_{2i}^*, Y_i^*, \theta) h_{1}(u/\lambda_{ri} + X_{1i}^*|X_{2i}^*, Y_i^*; \theta^{(t)}, Z)/\lambda_{ri}^{(t)} \, du \]
\[ = \int_0^{\infty} e^{-u} \ell_{1\ell}(u/\lambda_{ri} + X_{1i}^*, X_{2i}^*, Y_i^*, \theta) h_{1}(u/\lambda_{ri} + X_{1i}^*|X_{2i}^*, Y_i^*; \theta^{(t)}, Z)/\lambda_{ri}^{(t)} \, du \]
\[ = \int_0^{\infty} e^{-u} g_1(u) \, du \]
Using standard weights and nodes for Laguerre quadrature when α=0:

\[
\sum_{j=1}^{M} \ell_{1i}(u_{j0}/\lambda_{ri}) + X_{1i}^* X_{2i}^* Y_i^* \theta \left( \frac{h_1(u_{j0}/\lambda_{ri}) + X_{1i}^* X_{2i}^* Y_i^* \theta^{(t)}; Z}{ue^{-u_{j0}/\lambda_{ri}}} \right) v_{j0} \\
= \sum_{j=1}^{M} \ell_{1i}(u_{j0}/\lambda_{ri}) + X_{1i}^* X_{2i}^* Y_i^* \theta w_{j0}^{(x)}
\]

Where \( w_{j0}^{(x)} = \frac{h_1(u_{j0}/\lambda_{ri}) + X_{1i}^* X_{2i}^* Y_i^* \theta^{(t)}; Z}{ue^{-u_{j0}/\lambda_{ri}}} v_{j0} \) for \( j=1...M \).

When α=1:

\[
\int_{X_{1i}}^{\infty} \ell_{1i}(r, X_{2i}^*, Y_i^*, \theta) h_1(r, X_{2i}^*, Y_i^*; \theta^{(t)}; Z) \, dr
\]

Let \( \lambda_{ri}^{(t)} = \ell_{1i}^{(t)}(r, X_{2i}^*, Y_i^*, \theta^{(t)}) \) evaluated at \( r=0 \) where \( \ell_{1i}^{(t)} \) is the derivative w.r.t. \( r \). With the change of variables \( u = \lambda_{ri}^{(t)} (r - X_{1i}^*) \) and \( dr = 1/\lambda_{ri}^{(t)} du \), the integral is equivalent to

\[
= \int_{0}^{\infty} \ell_{1i}(u/\lambda_{1i} + X_{1i}^* X_{2i}^* Y_i^* \theta \ h_1(u/\lambda_{1i} + X_{1i}^* X_{2i}^* Y_i^* \theta^{(t)}; Z)/\lambda_{ri}^{(t)} \, du
\]

\[
= \int_{0}^{\infty} ue^{-u} \ell_{1i}(u/\lambda_{1i} + X_{1i}^* X_{2i}^* Y_i^* \theta \ h_1(u/\lambda_{1i} + X_{1i}^* X_{2i}^* Y_i^* \theta^{(t)}; Z)/\lambda_{ri}^{(t)} \, du
\]

\[
= \int_{0}^{\infty} ue^{-u} g_1(u) \, du
\]

Using standard weights and nodes for Laguerre quadrature when \( \alpha=1 \):

\[
\sum_{j=1}^{M} \ell_{1i}(u_{j1}/\lambda_{ri}) + X_{1i}^* X_{2i}^* Y_i^* \theta \left( \frac{h_1(u_{j1}/\lambda_{ri}) + X_{1i}^* X_{2i}^* Y_i^* \theta^{(t)}; Z}{u_{j1}e^{-u_{j1}/\lambda_{ri}}} \right) v_{j1} \\
= \sum_{j=1}^{M} \ell_{1i}(u_{j1}/\lambda_{ri}) + X_{1i}^* X_{2i}^* Y_i^* \theta w_{j1}^{(x)}
\]
Where \( w_{j1}^{(x)} = \frac{h_3(u_{j1}/\lambda_{s1}^{(t)} + X_{1i}^*|X_{1i}^*, Y_{1i}^*, \theta^{(t)}, Z)}{u_{j1} e^{-u_{j1}/\lambda_{s1}^{(t)}}} \) for \( j = 1 \ldots M \).

\[ \Delta = (0, 1, 0) : \text{Repeat Section 3.2.4, substituting } \ell_{1i} \text{ with } \ell_{2i} \text{ and } h_1 \text{ with } h_4. \]

\[ \Delta = (1, 0, 1) \]

When \( \alpha = 0 \):
\[
\int_0^{\infty} \ell_{1i}(X_{1i}^*, s, Y_{i}^*, \theta) h_2(s|X_{1i}^*, Y_{i}^*; \theta^{(t)}, Z) \, ds
\]

Let \( \lambda_{si}^{(t)} = \ell_{1i}'(X_{1i}^*, s, Y_{i}^*, \theta^{(t)}) \) evaluated at \( s = 0 \) where \( \ell_{1i}' \) is the derivative w.r.t. \( s \). With the change of variables \( u = \lambda_{si}^{(t)}(s - X_{2i}^*) \) and \( ds = 1/\lambda_{si}^{(t)} du \), the integral is equivalent to
\[
= \int_0^{\infty} \ell_{1i}(X_{1i}^*, u/\lambda_{si}^{(t)} + X_{2i}^*, Y_{i}^*, \theta) h_2(u/\lambda_{si}^{(t)} + X_{2i}^*|X_{1i}^*, Y_{i}^*; \theta^{(t)}, Z)/\lambda_{si}^{(t)} \, du
\]
\[
= \int_0^{\infty} e^{-u} \ell_{1i}(X_{1i}^*, u/\lambda_{si}^{(t)} + X_{2i}^*, Y_{i}^*, \theta) h_2(u/\lambda_{si}^{(t)} + X_{2i}^*|X_{1i}^*, Y_{i}^*; \theta^{(t)}, Z) \, du
\]
\[
= \int_0^{\infty} e^{-u} g_2(u) \, du
\]

Using standard weights and nodes for Laguerre quadrature when \( \alpha = 0 \)
\[
\simeq \sum_{k=1}^{M} \ell_{1i}(X_{1i}^*, u_{k0}/\lambda_{si}^{(t)} + X_{2i}^*, Y_{i}^*, \theta) \left( \frac{h_2(u_{k0}/\lambda_{si}^{(t)} + X_{2i}^*|X_{1i}^*, Y_{i}^*; \theta^{(t)}, Z)}{e^{-u_{k0}/\lambda_{si}^{(t)}}} v_{k0} \right)
\]
\[
= \sum_{k=1}^{M} \ell_{1i}(u_{k0}/\lambda_{si}^{(t)} + X_{2i}^*|X_{1i}^*, Y_{i}^*, \theta) w_{ki}^{(y)}
\]

Where \( w_{k0i}^{(y)} = \frac{h_2(u_{k0}/\lambda_{si}^{(t)} + X_{2i}^*|X_{1i}^*, Y_{i}^*, \theta^{(t)}, Z)}{e^{-u_{k0}/\lambda_{si}^{(t)}}} v_{k0} \).
When \( \alpha = 1 \) I will evaluate

\[
\int_{X^*_{2i}}^{\infty} \ell_{1ij}(X^*_{1i}, s, Y^*_i, \theta) h_2(s|X^*_{1i}, Y^*_i; \theta^{(t)}, Z) \, ds
\]

Let \( \lambda^{(t)}_{si} = \ell'_{1i}(X^*_{1i}, s, Y^*_i, \theta^{(t)}) \) evaluated at \( s=0 \) where \( \ell'_{1i} \) is the derivative w.r.t. \( s \). With the change of variables \( u = \lambda^{(t)}_{si}(s - X^*_{2i}) \) and \( ds = 1/\lambda^{(t)}_{si} \, du \), the integral is equivalent to

\[
= \int_0^\infty \ell_{1i}(X^*_{1i}, u/\lambda^{(t)}_{si} + X^*_{2i}, Y^*_i, \theta) h_2(u/\lambda^{(t)}_{si} + X^*_{2i}|X^*_{1i}, Y^*_i; \theta^{(t)}, Z) / \lambda^{(t)}_{si} \, du
\]

\[
= \int_0^\infty u e^{-u} \ell_{1i}(X^*_{1i}, u/\lambda^{(t)}_{si} + X^*_{2i}, Y^*_i, \theta) h_2(u/\lambda^{(t)}_{si} + X^*_{2i}|X^*_{1i}, Y^*_i; \theta^{(t)}, Z) / u e^{-u} \lambda^{(t)}_{si} \, du
\]

\[
= \int_0^\infty u e^{-u} g_2(u) \, du
\]

Using standard weights and nodes for Laguerre quadrature when \( \alpha = 1 \)

\[
\simeq \sum_{k=1}^{M} \ell_{1i}(X^*_{1i}, u_{k1}/\lambda^{(t)}_{si} + X^*_{2i}, Y^*_i, \theta) \left( \frac{h_2(u_{k1}/\lambda^{(t)}_{si} + X^*_{2i}|X^*_{1i}, Y^*_i; \theta^{(t)}, Z)}{u_{k1} e^{-u_{k1} \lambda^{(t)}_{si}}} v_{k1} \right)
\]

\[
= \sum_{k=1}^{M} \ell_{1i}(X^*_{1i}, u_{k1}/\lambda^{(t)}_{si} + X^*_{2i}, Y^*_i, \theta) w^{(y)}_{k1i}
\]

Where \( w^{(y)}_{k1i} = \frac{h_2(u_{k1}/\lambda^{(t)}_{si} + X^*_{2i}|X^*_{1i}, Y^*_i; \theta^{(t)}, Z)}{u_{k1} e^{-u_{k1} \lambda^{(t)}_{si}}} v_{k1} \).

\( \Delta = (1, 0, 0) \): Repeat Section 3.2.4, substituting \( \ell_{1i} \) with \( \ell_{2i} \) and \( h_2 \) with \( h_5 \).

\( \Delta = (0, 0, 1) \)

I denote \( u^{(x_1)} \) for the change of variable \( r \) and \( u^{(x_2)} \) for the change of variable \( s \) for integration over \( X_1 \) and \( X_2 \) respectively. When \( \alpha = 0 \):

\[
\int_{X_1}^{\infty} \int_{X_2}^{\infty} \ell_{1i}(r, s, Y_i^*, \theta) h_3(r, s|Y_i^*; \theta(t), Z) \, ds \, dr
\]

With the change of variables \( u^{(x_2)} = \lambda_{si}^{(t)} (s - X_{2i}^*) \) and \( ds = 1/\lambda_{si}^{(t)} du^{(x_2)} \), the integral is equivalent to

\[
= \int_{X_1}^{\infty} \int_{0}^{\infty} e^{-u^{(x_2)}} \ell_{1i}(r, u^{(x_2)}/\lambda_{si}^{(t)} + X_{2i}^*, Y_i^*, \theta) h_3(r, u^{(x_2)}/\lambda_{si}^{(t)} + X_{2i}^*, Y_i^*; \theta(t), Z) \, du^{(x_2)} \, dr
\]

Using standard weights and nodes for Laguerre quadrature when \( \alpha = 0 \):

\[
\approx \sum_{k=1}^{M} \left\{ \int_{X_{1i}}^{\infty} \ell_{1i}(r, u_{k0}/\lambda_{si}^{(t)} + X_{2i}^*, Y_i^*, \theta) \left( \frac{h_3(r, u_{k0}/\lambda_{si}^{(t)} + X_{2i}^*, Y_i^*; \theta(t), Z)}{e^{-u_{k0}} \lambda_{si}^{(t)}} v_{k0} \right) dr \right\}
\]

\[
= \sum_{k=1}^{M} \left\{ \int_{X_{1i}}^{\infty} \ell_{1i}(r, u_{k0}/\lambda_{si}^{(t)} + X_{2i}^*, Y_i^*, \theta) \left( h_3(r, u_{k0}/\lambda_{si}^{(t)} + X_{2i}^*, Y_i^*; \theta(t), Z) \right) v_{k0} \right\} dr
\]
With the change of variables \( u^{(x_1)} = \lambda_{r_i}^{(t)}(r - X_{1i}^*) \) and \( dr = 1/\lambda_{r_i}^{(t)} du \), the expression is equal to

\[
= \sum_{k=1}^{M} \left\{ \frac{v_{k0}}{\lambda_{s_i}^{(t)} e^{-u_{k0}}} \int_0^\infty \ell_{1i}(u_{j0}/\lambda_{r_i}^{(t)} + X_{1i}^*, u_{k0}/\lambda_{s_i}^{(t)} + X_{2i}^*, Y_i^*, \theta) \\
\times h_3(u_{j0}/\lambda_{r_i}^{(t)} + X_{1i}^*, u_{k0}/\lambda_{s_i}^{(t)} + X_{2i}^*, Y_i^*, \theta) h_3(u_{j0}/\lambda_{r_i}^{(t)} + X_{1i}^*, u_{k0}/\lambda_{s_i}^{(t)} + X_{2i}^*, Y_i^*, \theta) \right\}
\]

Using standard weights and nodes for Laguerre quadrature when \( \alpha = 0 \)

\[
\approx \sum_{k=1}^{M} \left\{ \frac{v_{k0}}{\lambda_{s_i}^{(t)} e^{-u_{k0}}} \sum_{j=1}^{M} \left[ \ell_{1i}(u_{j0}/\lambda_{r_i}^{(t)} + X_{1i}^*, u_{k0}/\lambda_{s_i}^{(t)} + X_{2i}^*, Y_i^*, \theta) \right] \\
\times \left( \frac{h_3(u_{j0}/\lambda_{r_i}^{(t)} + X_{1i}^*, u_{k0}/\lambda_{s_i}^{(t)} + X_{2i}^*, Y_i^*, \theta) h_3(u_{j0}/\lambda_{r_i}^{(t)} + X_{1i}^*, u_{k0}/\lambda_{s_i}^{(t)} + X_{2i}^*, Y_i^*, \theta) }{u_{j0}} \right) \right\}
\]

\[
= \sum_{k=1}^{M} \left\{ \sum_{j=1}^{M} \ell_{1i}(u_{j0}/\lambda_{r_i}^{(t)} + X_{1i}^*, u_{k0}/\lambda_{s_i}^{(t)} + X_{2i}^*, Y_i^*, \theta) h_3(u_{j0}/\lambda_{r_i}^{(t)} + X_{1i}^*, u_{k0}/\lambda_{s_i}^{(t)} + X_{2i}^*, Y_i^*, \theta) \right\}
\]

\[
= \sum_{j=1}^{M} \sum_{k=1}^{M} \ell_{1i}(u_{j0}/\lambda_{r_i}^{(t)} + X_{1i}^*, u_{k0}/\lambda_{s_i}^{(t)} + X_{2i}^*, Y_i^*, \theta) w^{(xy)}_{j_{k0}i}
\]

Where \( w^{(xy)}_{j_{k0}i} = h_3(u_{j0}/\lambda_{r_i}^{(t)} + X_{1i}^*, u_{k0}/\lambda_{s_i}^{(t)} + X_{2i}^*, Y_i^*, \theta) \frac{v_{j0}}{e^{-u_{j0}/\lambda_{r_i}^{(t)}}} \frac{v_{k0}}{e^{-u_{k0}/\lambda_{s_i}^{(t)}}} \)
I denote $u^{(x_1)}$ for the change of variable $r$ and $u^{(x_2)}$ for the change of variable $s$ for integration over $X_1$ and $X_2$ respectively. When $\alpha=1$:

$$\int_{X_1^*}^{\infty} \int_{X_2^*}^{\infty} \ell_1(r, s, Y_i^*, \theta) h_3(r, s|Y_i^*; \theta(t), Z) \, ds \, dr$$

With the change of variables $u^{(x_2)} = \lambda_{si}^{(t)} (s - X_{2i}^*)$ and $ds = 1/\lambda_{si}^{(t)} du^{(x_2)}$, the integral is equivalent to:

$$= \int_{X_1^*}^{\infty} \int_{0}^{\infty} \ell_1(r, u^{(x_2)}/\lambda_{si}^{(t)} + X_{2i}^*, Y_i^*, \theta) h_3(r, u^{(x_2)}/\lambda_{si}^{(t)} + X_{2i}^*|Y_i^*; \theta(t), Z)/\lambda_{si}^{(t)} \, du^{(x_2)} \, dr$$

$$= \int_{X_1^*}^{\infty} \int_{0}^{\infty} u^{(x_2)} e^{-s} \ell_1(r, u^{(x_2)}/\lambda_{si}^{(t)} + X_{2i}^*, Y_i^*, \theta) h_3(r, u^{(x_2)}/\lambda_{si}^{(t)} + X_{2i}^*|Y_i^*; \theta(t), Z) u^{(x_2)} e^{-u^{(x_2)}} \lambda_{si}^{(t)} du^{(x_2)} \, dr$$

$$= \int_{X_1^*}^{\infty} \int_{0}^{\infty} u^{(x_2)} e^{-u^{(x_2)}} g_3(r, u^{(x_2)}) \, du^{(x_2)} \, dr$$

Using standard weights and nodes for Laguerre quadrature when $\alpha=1$

$$\simeq \int_{X_1^*}^{\infty} \sum_{k=1}^{M} \ell_1(r, u_{k_1}/\lambda_{si}^{(t)} + X_{2i}^*, Y_i^*, \theta) \frac{h_3(r, u_{k_1}/\lambda_{si}^{(t)} + X_{2i}^*|Y_i^*; \theta(t), Z)}{u_{k_1} e^{-u_{k_1} \lambda_{si}^{(t)}}} v_{k_1} \, dr$$

$$= \sum_{k=1}^{M} \left\{ \int_{X_1^*}^{\infty} \ell_1(r, u_{k_1}/\lambda_{si}^{(t)} + X_{2i}^*, Y_i^*, \theta) \frac{h_3(r, u_{k_1}/\lambda_{si}^{(t)} + X_{2i}^*|Y_i^*; \theta(t), Z)}{u_{k_1} e^{-u_{k_1} \lambda_{si}^{(t)}}} v_{k_1} \, dr \right\}$$

$$= \sum_{k=1}^{M} \frac{v_{k_1}}{\lambda_{si}^{(t)} u_{k_1} e^{-u_{k_1}}} \int_{X_1^*}^{\infty} \ell_1(r, u_{k_1}/\lambda_{si}^{(t)} + X_{2i}^*, Y_i^*, \theta) h_3(r, u_{k_1}/\lambda_{si}^{(t)} + X_{2i}^*|Y_i^*; \theta(t), Z) \, dr$$
With the change of variables $u^{(x_1)} = \lambda^{(t)}_{ri} (r - X^{*}_{1i})$ and $dr = 1/\lambda^{(t)}_{ri} du$, the expression is equivalent to

\[
\begin{align*}
&= \sum_{k=1}^{M} \left\{ \frac{v_{k1}}{\lambda^{(t)}_{si} u_{k1} e^{-u_{k1}}} \int_{0}^{\infty} \ell_{1i}(u_{j1}/\lambda^{(t)}_{ri} + X^{*}_{1i}, u_{k1}/\lambda^{(t)}_{si} + X^{*}_{2i}, Y^{*}_{i}, \theta) \times h_{3}(u_{j1}/\lambda^{(t)}_{ri} + X^{*}_{1i}, u_{k1}/\lambda^{(t)}_{si} + X^{*}_{2i} Y^{*}_{i}, \theta) h_{3}(u_{j1}/\lambda^{(t)}_{ri} + X^{*}_{1i}, u_{k1}/\lambda^{(t)}_{si}) \times X^{*}_{2i}, Y^{*}_{i}; \theta^{(t)}, Z) / \lambda^{(t)}_{ri} du^{(x_1)} \right\} \\
&= \sum_{k=1}^{M} \left\{ \frac{v_{k1}}{\lambda^{(t)}_{si} u_{k1} e^{-u_{k1}}} \int_{0}^{\infty} u^{(x_1)} e^{-u^{(x_1)}} g_{4}(u^{(x_1)}) du^{(x_1)} \right\}
\end{align*}
\]

Using standard weights and nodes for Laguerre quadrature when $\alpha=1$
Where

\[
w_{jk1i}^{(xy)} = h_3(u_{j1}/\lambda_{ri}^{(t)} + X_{1i}^s, u_{k1}/\lambda_{si}^{(t)} + X_{2i}^s, Y_i^s, \theta(t)^{,}, Z) \frac{v_{j1}}{u_{j1}e^{-u_{j1}/\lambda_{ri}^{(t)}}} \frac{v_{k1}}{u_{k1}e^{-u_{k1}/\lambda_{si}^{(t)}}}
\]

\[\Delta=(0,0,0):\] Repeat Section 3.2.4, substituting \(\ell_{1i}\) with \(\ell_{2i}\) and \(h_3\) with \(h_6\).

### 3.2.5 Expectation Step

\[
Q(\theta|\theta^{(t)}) \simeq \sum_{i: \Delta=(1,1)} \ell_{1i}(X_{1i}^s, X_{2i}^s, Y_i^s, \theta)
\]

\[+ \sum_{i: \Delta=(0,1,1)} \sum_{j=1}^M \ell_{1i}(u_{j1}/\lambda_{ri}^{(t)} + X_{1i}^s, X_{2i}^s, Y_i^s, \theta) w_{ji}^{(x)} + \sum_{i: \Delta=(1,0,1)} \sum_{k=1}^M \ell_{1i}(X_{1i}^s, u_{k1}/\lambda_{si}^{(t)} + X_{2i}^s, Y_i^s, \theta) w_{ki}^{(y)}
\]

\[+ \sum_{i: \Delta=(0,0,1)} \sum_{j=1}^M \sum_{k=1}^M \ell_{1i}(u_{j1}/\lambda_{ri}^{(t)} + X_{1i}^s, u_{k1}/\lambda_{si}^{(t)} + X_{2i}^s, Y_i^s, \theta) w_{jki}^{(xy)}
\]

\[+ \sum_{i: \Delta=(1,1,0)} \ell_{2i}(X_{1i}^s, X_{2i}^s, Y_i^s, \theta)
\]

\[+ \sum_{i: \Delta=(0,1,0)} \sum_{j=1}^M \ell_{2i}(u_{j1}/\lambda_{ri}^{(t)} + X_{1i}^s, X_{2i}^s, Y_i^s, \theta) w_{ji}^{(x)}
\]

\[+ \sum_{i: \Delta=(1,0,0)} \sum_{k=1}^M \ell_{2i}(X_{1i}^s, u_{k1}/\lambda_{si}^{(t)} + X_{2i}^s, Y_i^s, \theta) w_{ki}^{(y)}
\]

\[+ \sum_{i: \Delta=(0,0,0)} \sum_{j=1}^M \sum_{k=1}^M \ell_{2i}(u_{j1}/\lambda_{ri}^{(t)} + X_{1i}^s, u_{k1}/\lambda_{si}^{(t)} + X_{2i}^s, Y_i^s, \theta) w_{jki}^{(xy)}
\]
3.2.6 Maximization Step

The maximization method will depend on the parametric models specified for the components of Equation 3.2: $\theta^{(t+1)} = \arg \max_{\theta} Q(\theta|\theta^{(t)})$.

3.3 Variance Estimation

Using Louis’s method the observed information matrix $I$ is

$$I = \sum_{i=1}^{n} E_{(X_1i, X_2i, Y_i)}[B_i|W_i, Z_i; \hat{\theta}] \quad \text{and} \quad \sum_{i=1}^{n} E_{(X_1i, X_2i, Y_i)}[U_i U_i^T|W_i, Z_i; \hat{\theta}]$$

$$= \sum_{i=1}^{n} E_{(X_1i, X_2i, Y_i)}[U_i^T W_i, Z_i; \hat{\theta}] E_{(X_1i, X_2i, Y_i)}[U_i^T|W_i, Z_i; \hat{\theta}]$$

$$U_i = \begin{cases} \frac{\partial \ell_i}{\partial \theta} & \text{if } \delta_{yi} = 1 \\ \frac{\partial \ell_i}{\partial \theta} & \text{if } \delta_{yi} = 0 \end{cases}$$

$$B_i = \begin{cases} \frac{-\partial^2 \ell_i}{\partial \theta^2} & \text{if } \delta_{yi} = 1 \\ \frac{-\partial^2 \ell_i}{\partial \theta^2} & \text{if } \delta_{yi} = 0 \end{cases}$$
Expressing $I$ in terms of summation of $I_i$'s for which the form depends on $\Delta_i,$

\[
I = \sum_{i : \Delta = (1,1)} B_i(X_{1i}^*, X_{2i}^*, Y_i^*, \hat{\theta}, Z) \\
+ \sum_{i : \Delta = (0,1)} \int_{X_{1i}}^{\infty} B_i(r, X_{2i}^*, Y_i^*, \hat{\theta}, Z) h_1(r, X_{2i}^*, Y_i^*; \hat{\theta}, Z) \, dr \\
- \int_{X_{1i}}^{\infty} U_i(r, X_{2i}^*, Y_i^*, \hat{\theta}, Z) U_i(r, X_{2i}^*, Y_i^*; \hat{\theta}, Z)^\top h_1(r, X_{2i}^*, Y_i^*; \hat{\theta}, Z) \, dr \\
+ \left\{ \int_{X_{1i}}^{\infty} U_i(r, X_{2i}^*, Y_i^*, \hat{\theta}, Z) h_1(r, X_{2i}^*, Y_i^*; \hat{\theta}, Z) \, dr \right\} \\
\times \left\{ \int_{X_{1i}}^{\infty} U_i(r, X_{2i}^*, Y_i^*; \hat{\theta}, Z)^\top h_1(r, X_{2i}^*, Y_i^*; \hat{\theta}, Z) \, dr \right\} \\
+ \sum_{i : \Delta = (1,0)} \int_{X_{2i}}^{\infty} B_i(X_{1i}^*, s, Y_i^*, \hat{\theta}, Z) h_2(X_{1i}^*, s, Y_i^*; \hat{\theta}, Z) \, ds \\
- \int_{X_{2i}}^{\infty} U_i(X_{1i}^*, s, Y_i^*, \hat{\theta}, Z) U_i(X_{1i}^*, s, Y_i^*; \hat{\theta}, Z)^\top h_2(X_{1i}^*, s, Y_i^*; \hat{\theta}, Z) \, ds \\
+ \left\{ \int_{X_{2i}}^{\infty} U_i(X_{1i}^*, s, Y_i^*, \hat{\theta}, Z) h_2(X_{1i}^*, s, Y_i^*; \hat{\theta}, Z) \, ds \right\} \\
\times \left\{ \int_{X_{2i}}^{\infty} U_i(X_{1i}^*, s, Y_i^*; \hat{\theta}, Z)^\top h_2(X_{1i}^*, s, Y_i^*; \hat{\theta}, Z) \, ds \right\} 
\]
\[\begin{align*}
&+ \sum_{\Delta=(0,0,1)} \int_{X_{1i}}^{\infty} \int_{X_{2i}}^{\infty} B_i(r, s, Y_{1i}^*, \hat{\theta}, Z) h_3(r, s, Y_{1i}^*; \hat{\theta}, Z) \, ds \, dr \\
&- \int_{X_{1i}}^{\infty} \int_{X_{2i}}^{\infty} U_i(r, s, Y_{1i}^*, \hat{\theta}, Z) U_i(r, s, Y_{1i}^*; \hat{\theta}, Z)^T h_3(r, s, Y_{1i}^*; \hat{\theta}, Z) \, ds \, dr \\
&+ \left\{ \int_{X_{1i}}^{\infty} \int_{X_{2i}}^{\infty} U_i(r, s, Y_{1i}^*, \hat{\theta}, Z) h_3(r, s, Y_{1i}^*; \hat{\theta}, Z) \, ds \, dr \right\} \\
&\times \left\{ \int_{X_{1i}}^{\infty} \int_{X_{2i}}^{\infty} U_i(r, s, Y_{1i}^*, \hat{\theta}, Z)^T h_3(r, s, Y_{1i}^*; \hat{\theta}, Z) \, ds \, dr \right\} \\
&+ \sum_{\Delta=(1,1,0)} B_i(X_{1i}^*, X_{2i}^*, Y_{1i}^*, \hat{\theta}, Z) \\
&+ \sum_{\Delta=(1,0,1)} \int_{X_{1i}}^{\infty} B_i(r, X_{2i}^*, Y_{1i}^*, \hat{\theta}, Z) h_4(r, X_{2i}^*, Y_{1i}^*; \hat{\theta}, Z) \, dr \\
&- \int_{X_{1i}}^{\infty} U_i(r, X_{2i}^*, Y_{1i}^*, \hat{\theta}, Z) U_i(r, X_{2i}^*, Y_{1i}^*; \hat{\theta}, Z)^T h_4(r, X_{2i}^*, Y_{1i}^*; \hat{\theta}, Z) \, dr \\
&+ \left\{ \int_{X_{1i}}^{\infty} U_i(r, X_{2i}^*, Y_{1i}^*, \hat{\theta}, Z) h_4(r, X_{2i}^*, Y_{1i}^*; \hat{\theta}, Z) \, dr \right\} \\
&\times \left\{ \int_{X_{1i}}^{\infty} U_i(r, X_{2i}^*, Y_{1i}^*; \hat{\theta}, Z)^T h_4(r, X_{2i}^*, Y_{1i}^*; \hat{\theta}, Z) \, dr \right\}
\end{align*}\]
3.4 Complication-free Survival

3.4.1 Survival Curve

Since our interest is on comparing the distribution of complication-free survival times between two groups \((Z=1\text{ and } Z=0)\) I will define a new event \(M\) as the minimum of \((X_1, X_2, Y)\), which follows the survival curves \(g(t|\hat{\theta}; Z=1)\) and \(g(t|\hat{\theta}; Z=0)\).

\[
g(t|\theta; Z) = \int_t^\infty \int_t^\infty \int_t^\infty f_{X_1}(x_1|\theta; Z)f_{X_2}(x_2|x_1, \theta; Z)f_Y(y|x_1, x_2, \theta; Z) \, dx_1 \, dx_2 \, dy \quad (3.4)
\]
3.4.2 New Proposed Test Statistic

Since I want to compare two heterogeneous curves without making assumptions as to the nature of the difference (e.g., proportional-hazards) I propose a test statistic based on the integrated squared difference (ISD).

\[
ISD = n \int_0^\infty \left[ g(t|\hat{\theta}; Z = 1) - g(t|\hat{\theta}; Z = 0) \right]^2 dt
\]

Let \( \hat{\theta}_1 \) correspond to \( Z=1 \) and \( \hat{\theta}_2 \) correspond to \( Z=0 \). Using a first-order Taylor approximation:

\[
ISD = n \int_0^\infty \left[ g(t|\hat{\theta}_2) + \frac{\partial g(t|\hat{\theta}_2)}{\partial \theta} (\hat{\theta}_1 - \hat{\theta}_2) + o(1) - g(t|\hat{\theta}_2) \right]^2 dt
\]

\[
= n \int_0^\infty \left[ \frac{\partial g(t|\hat{\theta}_2)}{\partial \theta} (\hat{\theta}_1 - \hat{\theta}_2) + o(1/n) - g(t|\hat{\theta}_2) \right] dt
\]

\[
\approx n(\hat{\theta}_1 - \hat{\theta}_2)^T \int_0^\infty \left[ \frac{\partial g(t|\hat{\theta}_2)}{\partial \theta} \right]^T \left[ \frac{\partial g(t|\hat{\theta}_2)}{\partial \theta} \right] dt (\hat{\theta}_1 - \hat{\theta}_2)
\]

as \( n \to \infty \) because \( \sqrt{n}|\hat{\theta}_1 - \hat{\theta}_2| = O(1) \). Where,

\[
\frac{\partial g(t|\theta_2)}{\partial \theta} = \int_t^\infty \int_t^\infty \int_t^\infty \frac{\partial}{\partial \theta} f_{X_1}(x_1|\theta_2) f_{X_2}(x_2|x_1, \theta_2) f_Y(y|x_1, x_2, \theta_2) \, dx_1 \, dx_2 \, dy
\]
I see that $(\hat{\theta}_1 - \hat{\theta}_2)^T A (\hat{\theta}_1 - \hat{\theta}_2)$ is a quadratic form and I wish to know its distribution. Under $H_0$, $\sqrt{n}(\hat{\theta}_1 - \hat{\theta}_2) \sim N(0, \Sigma)$. Given the Cholesky decomposition $\Sigma = U^T U$ where $U$ is upper-triangular and choose the matrix $C = (U^T)^{-1}$: $C \Sigma C^T = I$ I use the transformation $Y = C \sqrt{n}(\hat{\theta}_1 - \hat{\theta}_2) \sim N(0, I)$.

$$n(\hat{\theta}_1 - \hat{\theta}_2)^T A (\hat{\theta}_1 - \hat{\theta}_2) = Y^T U A U^T Y = Y^T B Y$$

Using the transformation $Z = \Gamma Y$ where $\Gamma$ is an orthogonal matrix such that $\Gamma^T B \Gamma = \Lambda$ a $3 \times 3$ diagonal matrix whose diagonal entries are eigenvalues ($\lambda_1, \lambda_2, \lambda_3$) for of $B$. Therefore, ISD is distributed weighted Chi-squared where the weights are equal to the eigenvalues of $B$. Such a distribution can be analytically calculated when $\lambda$s are known (40).

3.4.3 Simulation Approach

Let $\theta$ be the full set of parameters used to fit the whole dataset $Z=1,0$. $\theta$ can be divided into $[\theta_P, \theta_Z]$, $\theta_Z$ for the effects of $Z$ and $\theta_P$ for all others. The testing procedure is as follows:

1. Find $\hat{\theta} = [\hat{\theta}_P, \hat{\theta}_Z]$ given the observed data.

2. Calculate the test statistic $T_{data} = \int_0^\infty \left[ g(t|\hat{\theta}_P, \hat{\theta}_Z) - g(t|\hat{\theta}_P, \hat{\theta}_Z = 0) \right]^2 dt$ given the observed data.

3. Simulate a large number $N = 1000$ datasets given $\theta = [\hat{\theta}_P, \hat{\theta}_Z = 0]$

4. In $k^{th}$ simulated dataset $k = 1 \ldots N$
(a) Find \( \hat{\theta}_S^{(k)} = [\hat{\theta}_P^{(k)} , \hat{\theta}_Z^{(k)}] \) given the simulated data.

(b) Calculate the test statistic \( T^{(k)} = \int_0^\infty \left[ g(t|\hat{\theta}_P^{(k)}, \hat{\theta}_Z^{(k)}) - g(t|\hat{\theta}_P^{(k)}, \hat{\theta}_Z^{(k)} = 0) \right]^2 dt \)

(c) Store \( T^{(k)} \)

5. Estimate p-value \( p = P(T_{\text{data}} \geq T^{(k)}) \) empirically.

3.4.4 Wald Test

Testing \( H_0 : g(t|Z = 1) = g(t|Z = 0) \Rightarrow H_0 : \theta_Z = 0 \). Find the sub-matrix \( B_Z(\theta_Z) \) of the full information matrix \( B(\theta) \). Calculate the test statistic \( W = n\hat{\theta}_Z^T B_Z(\hat{\theta}_Z)^{-1} \hat{\theta}_Z \xrightarrow{d} \chi^2_r \) is Chi-squared with \( r \) degrees of freedom corresponding to the number of parameters being tested.

3.4.5 Marginal Survival

Marginal survival curves for time until death \( Y \) can also be calculated with slight modification of the limits of integration to Equation 3.4.

\[
g_Y(t|\theta; Z) = \int_t^\infty \int_0^\infty \int_0^\infty f_{X_1}(x_1|\theta; Z) f_{X_2}(x_2|x_1, \theta; Z) f_Y(y|x_1, x_2, \theta; Z) \, dx_1 \, dx_2 \, dy \quad (3.7)
\]

Marginal survival curves for \( X_1 \) and \( X_2 \) can be obtained similarly. The testing procedure for Equation 3.5 can be applied to the marginal as well as the complication-free survival curves.
CHAPTER 4

PARAMETRIC CONDITIONAL MODEL: SIMULATIONS

In this chapter I implement the model in Section 3.1.3. In Section 4.1 I give the likelihood function, score vector, and information matrix of Equation 3.2 under the assume exponential distribution. In Section 4.2 I discuss the practical implementation of the quadrature routine required for the EM algorithm. In Section 4.3 I derive the first derivatives of the likelihood necessary for calculating the test statistic in Equation 3.6. In Section 4.4 I report and discuss the results of simulations testing my algorithm on data generated from both exponential and Weibull distributions.

4.1 Exponential Distribution

4.1.1 Log-likelihood

\[ \ell_1 = \ln(f_{X_1}(x_1|Z, \alpha^T)) + \ln(f_{X_2|X_1}(x_2|x_1, Z, \beta^T)) + \ln(f_{Y|X_1,X_2}(y|x_1, x_2, Z, \gamma^T)) \]

\[ \ell_2 = \ln(f_{X_1}(x_1|Z, \alpha^T)) + \ln(f_{X_2|X_1}(x_2|x_1, Z, \beta^T)) + \ln(S_{Y|X_1,X_2}(y|x_1, x_2, Z, \gamma^T)) \]
Where

\[
\tilde{Z}_\alpha^T \alpha = \alpha_0 + Z_p^T \alpha_p \\
\tilde{Z}_\beta^T \beta = \beta_0 + h_1(x_1|\beta_t) + Z_p^T \beta_p \\
\tilde{Z}_\gamma^T \gamma = \gamma_0 + h_2(x_1, x_2|\gamma_t) + Z_p^T \gamma_p
\]

(4.1)

The vectors \(Z_p, \alpha_p, \beta_p, \gamma_p\) are column vectors of length \(p\). \(h_1(x_1|\beta_t)\) is a linear function of \(x_1\) and possibly its transformations taking on the functional form dependence of \(x_2\) on \(x_1\). It could be a polynomial or log transformation. Similarly \(h_2(x_1, x_2|\gamma_t)\) is a linear function of \(x_1\) and \(x_2\).

Assume that they follow exponential distributions:

\[
f_{X_1}(x_1|Z_p, \alpha) = \exp(-Z_\alpha^T \alpha) \exp(-\exp(-Z_\alpha^T \alpha)x_1)
\]

\[
f_{X_2|x_1}(x_2|x_1, Z_p, \beta) = \exp(-Z_\beta^T \beta) \exp(-\exp(-Z_\beta^T \beta)x_2)
\]

\[
f_{Y|x_1,x_2}(y|x_1, x_2, Z_p, \gamma) = \exp(-Z_\gamma^T \gamma) \exp(-\exp(-Z_\gamma^T \gamma)y)
\]

\[
S_{Y|x_1,x_2}(y|x_1, x_2, Z_p, \gamma) = \exp(-\exp(-Z_\gamma^T \gamma)y)
\]

(4.2)
\[ \ell_1i(X_{1i}, X_{2i}, Y_i^*, \theta) = \ln \left( \exp(-Z_{\alpha i}^T \alpha) \exp(-\exp(-Z_{\alpha i}^T \alpha))x_{1i} \right) \]
\[ + \ln \left( \exp(-Z_{\beta i}^T \beta) \exp(-\exp(-Z_{\beta i}^T \beta))x_{2i} \right) + \ln \left( \exp(-Z_{\gamma i}^T \gamma) \exp(-\exp(-Z_{\gamma i}^T \gamma))y_i \right) \]
\[ = -Z_{\alpha i}^T \alpha - \exp(-Z_{\alpha i}^T \alpha)x_{1i} - Z_{\beta i}^T \beta - \exp(-Z_{\beta i}^T \beta)x_{2i} - Z_{\gamma i}^T \gamma - \exp(-Z_{\gamma i}^T \gamma)y_i \]

\[ \ell_2i(X_{1i}, X_{2i}, Y_i^*, \theta) = \ln \left( \exp(-Z_{\alpha i}^T \alpha) \exp(-\exp(-Z_{\alpha i}^T \alpha))x_{1i} \right) \]
\[ + \ln \left( \exp(-Z_{\beta i}^T \beta) \exp(-\exp(-Z_{\beta i}^T \beta))x_{2i} \right) + \ln \left( \exp(-\exp(-Z_{\gamma i}^T \gamma))y_i \right) \]
\[ = -Z_{\alpha i}^T \alpha - \exp(-Z_{\alpha i}^T \alpha)x_{1i} - Z_{\beta i}^T \beta - \exp(-Z_{\beta i}^T \beta)x_{2i} - Z_{\gamma i}^T \gamma \]

In order to calculate the information from Equation 3.3 I need the score and information statistics. For brevity I will remove the \( i \) from the notation.

\[ U(\alpha_j) = \frac{\partial \ell_1}{\partial \alpha_j} = \frac{\partial \ell_2}{\partial \alpha_j} = \tilde{Z}_{\alpha j} \exp(-\tilde{Z}_\alpha^T \alpha)x_{1i} - 1 \]

\[ U(\beta_j) = \frac{\partial \ell_1}{\partial \beta_j} = \frac{\partial \ell_2}{\partial \beta_j} = \tilde{Z}_{\beta j} \exp(-\tilde{Z}_\beta^T \beta)x_{2i} - 1 \]

when \( \delta_y = 1 \) \[ U(\gamma_j) = \frac{\partial \ell_1}{\partial \gamma_j} = \tilde{Z}_{\gamma j} \exp(-\tilde{Z}_\gamma^T \gamma)y_i - 1 \]

when \( \delta_y = 0 \) \[ U(\gamma_j) = \frac{\partial \ell_2}{\partial \gamma_j} = \tilde{Z}_{\gamma j} \exp(-\tilde{Z}_\gamma^T \gamma)y_i \]
The vectors $\tilde{Z}_{\alpha j}$, $\tilde{Z}_{\beta j}$, and $\tilde{Z}_{\gamma j}$ are the $j^{th}$ elements of $\tilde{Z}_{\alpha}$, $\tilde{Z}_{\beta}$, and $\tilde{Z}_{\gamma}$ respectively.

$$U(\theta) = \begin{bmatrix} U(\alpha) \\ U(\beta) \\ U(\gamma) \end{bmatrix}$$

$$B(\alpha_{jk}) = -\frac{\partial^2 \ell_1}{\partial \alpha_j \partial \alpha_k} = -\frac{\partial^2 \ell_2}{\partial \alpha_j \partial \alpha_k} = \tilde{Z}_{\alpha j} \tilde{Z}_{\alpha k} \exp(-\tilde{Z}_{\alpha}^T \alpha)x_1$$

$$B(\beta_{jk}) = -\frac{\partial^2 \ell_1}{\partial \beta_j \partial \beta_k} = -\frac{\partial^2 \ell_2}{\partial \beta_j \partial \beta_k} = \tilde{Z}_{\beta j} \tilde{Z}_{\beta k} \exp(-\tilde{Z}_{\beta}^T \beta)x_2$$

$$B(\gamma_{jk}) = -\frac{\partial^2 \ell_1}{\partial \gamma_j \partial \gamma_k} = -\frac{\partial^2 \ell_2}{\partial \gamma_j \partial \gamma_k} = \tilde{Z}_{\gamma j} \tilde{Z}_{\gamma k} \exp(-\tilde{Z}_{\gamma}^T \gamma)y$$

$$B(\alpha_j \beta_k) = B(\alpha_j \gamma_k) = B(\beta_j \gamma_k) = 0$$

$$B(\theta) = \begin{bmatrix} B(\alpha)_{(p+1)\times(p+1)} & 0_{(p+1)\times(p+2)} & 0_{(p+1)\times(p+3)} \\ 0_{(p+2)\times(p+1)} & B(\beta)_{(p+2)\times(p+2)} & 0_{(p+2)\times(p+3)} \\ 0_{(p+3)\times(p+1)} & 0_{(p+3)\times(p+2)} & B(\gamma)_{(p+3)\times(p+3)} \end{bmatrix}$$
Assume that survival times follow log-normal distributions:

\[ f_{X_1}(x_1|Z_p, \alpha) = \frac{1}{\sqrt{2\pi}\sigma_1 x_1} \exp \left( -\frac{1}{2} \left( \frac{\ln(x_1) - Z_\alpha^T \alpha}{\sigma_1} \right)^2 \right) \]

\[ f_{X_2|x_1}(x_2|x_1, Z_p, \beta) = \frac{1}{\sqrt{2\pi}\sigma_2 x_2} \exp \left( -\frac{1}{2} \left( \frac{\ln(x_2) - Z_\beta^T \beta}{\sigma_2} \right)^2 \right) \]

\[ f_{Y|x_1,x_2}(y|x_1, x_2, Z_p, \gamma) = \frac{1}{\sqrt{2\pi}\sigma_3 y} \exp \left( -\frac{1}{2} \left( \frac{\ln(y) - Z_\gamma^T \gamma}{\sigma_3} \right)^2 \right) \]

\[ S_{Y|x_1,x_2}(y|x_1, x_2, Z_p, \gamma) = 1 - \Phi \left( \frac{\ln(y) - Z_\beta^T \gamma}{\sigma_3} \right) \]

Where \( \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \)

The log-likelihoods \( \ell_1 \) and \( \ell_2 \) can be written as

\[ \ell_1(X_1, X_2, Y^*, \theta) = -\ln(\sqrt{2\pi}\sigma_1 x_1) - \frac{1}{2} \left( \frac{\ln(x_1) - Z_\alpha^T \alpha}{\sigma_1} \right)^2 - \ln(\sqrt{2\pi}\sigma_2 x_2) - \frac{1}{2} \left( \frac{\ln(x_2) - Z_\beta^T \beta}{\sigma_2} \right)^2 - \ln(\sqrt{2\pi}\sigma_3 y) - \frac{1}{2} \left( \frac{\ln(y) - Z_\gamma^T \gamma}{\sigma_3} \right)^2 \]

\[ \ell_2(X_1, X_2, Y^*, \theta) = -\ln(\sqrt{2\pi}\sigma_1 x_1) - \frac{1}{2} \left( \frac{\ln(x_1) - Z_\alpha^T \alpha}{\sigma_1} \right)^2 - \ln(\sqrt{2\pi}\sigma_2 x_2) - \frac{1}{2} \left( \frac{\ln(x_2) - Z_\beta^T \beta}{\sigma_2} \right)^2 + \ln \left( 1 - \Phi \left( \frac{\ln(y) - Z_\gamma^T \gamma}{\sigma_3} \right) \right) \]
The first derivatives are:

\[
U(\alpha_j) = \frac{\partial \ell_1}{\partial \alpha_j} = \frac{\partial \ell_2}{\partial \alpha_j} = \frac{\dot{Z}_{\alpha j}}{\sigma_1^2} (\ln x_1 - Z_{1}^T \alpha)
\]

\[
U(\beta_j) = \frac{\partial \ell_1}{\partial \beta_j} = \frac{\partial \ell_2}{\partial \beta_j} = \frac{\dot{Z}_{\beta j}}{\sigma_2^2} (\ln x_2 - Z_{2}^T \beta)
\]

when \( \delta_y = 1 \)

\[
U(\gamma_j) = \frac{\partial \ell_1}{\partial \gamma_j} = \frac{\partial \ell_2}{\partial \gamma_j} = \frac{\dot{Z}_{\gamma j}}{\sigma_3^2} (\ln y - Z_{\gamma}^T \gamma)
\]

\[
U(\sigma_1) = \frac{\partial \ell_1}{\partial \sigma_1} = \frac{\partial \ell_2}{\partial \sigma_1} = \frac{(\ln x_1 - Z_{1}^T \alpha)^2}{\sigma_1^3} - \frac{1}{\sigma_1}
\]

\[
U(\sigma_2) = \frac{\partial \ell_1}{\partial \sigma_2} = \frac{\partial \ell_2}{\partial \sigma_2} = \frac{(\ln x_2 - Z_{2}^T \beta)^2}{\sigma_2^3} - \frac{1}{\sigma_2}
\]

when \( \delta_y = 1 \)

\[
U(\sigma_3) = \frac{\partial \ell_1}{\partial \sigma_3} = \frac{(\ln y - Z_{\gamma}^T \gamma)^2}{\sigma_3^3} - \frac{1}{\sigma_3}
\]

when \( \delta_y = 0 \)

\[
U(\gamma_j) = \frac{\partial \ell_2}{\partial \gamma_j} = \frac{\dot{Z}_{\gamma j}}{\sigma_3} \left( 1 - \Phi \left( \frac{\ln y - Z_{\gamma}^T \gamma}{\sigma_3} \right) \right)^{-1} \phi \left( \frac{\ln y - Z_{\gamma}^T \gamma}{\sigma_3} \right)
\]

when \( \delta_y = 0 \)

\[
U(\sigma_3) = \frac{\partial \ell_2}{\partial \sigma_3} = \left( 1 - \Phi \left( \frac{\ln y - Z_{\gamma}^T \gamma}{\sigma_3} \right) \right)^{-1} \phi \left( \frac{\ln y - Z_{\gamma}^T \gamma}{\sigma_3} \right) \frac{\ln y - Z_{\gamma}^T \gamma}{\sigma_3^2}
\]

Where \( \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \)
The components of the information matrix are:

\[
B(\alpha_{jk}) = -\frac{\partial^2 \ell_1}{\partial \alpha_j \partial \alpha_k} = \frac{\tilde{Z}_{\alpha j} \tilde{Z}_{\alpha k}}{\sigma^2_1}
\]

\[
B(\beta_{jk}) = -\frac{\partial^2 \ell_1}{\partial \beta_j \partial \beta_k} = \frac{\tilde{Z}_{\beta j} \tilde{Z}_{\beta k}}{\sigma^2_2}
\]

When \(\delta_y = 1\)

\[
B(\gamma_{jk}) = -\frac{\partial^2 \ell_1}{\partial \gamma_j \partial \gamma_k} = \frac{\tilde{Z}_{\gamma j} \tilde{Z}_{\gamma k}}{\sigma^2_2}
\]

When \(\delta_y = 0\)

\[
B(\gamma_{jk}) = -\frac{\partial^2 \ell_1}{\partial \gamma_j \partial \gamma_k} = \frac{\tilde{Z}_{\gamma j} \tilde{Z}_{\gamma k} \phi^2(V) - V \phi(V)(1 - \Phi(V))}{\sigma^2_3(1 - \Phi(V))^2}
\]

\[
B(\alpha_{j\sigma_1}) = -\frac{\partial^2 \ell_1}{\partial \alpha_j \partial \sigma_1} = \frac{2\tilde{Z}_{\alpha j}(\ln x_1 - Z^T_\alpha \alpha)}{\sigma^3_1}
\]

\[
B(\beta_{j\sigma_2}) = -\frac{\partial^2 \ell_1}{\partial \beta_j \partial \sigma_2} = \frac{2\tilde{Z}_{\beta j}(\ln x_2 - Z^T_\beta \beta)}{\sigma^3_2}
\]

When \(\delta_y = 1\)

\[
B(\gamma_{j\sigma_3}) = -\frac{\partial^2 \ell_1}{\partial \gamma_j \partial \sigma_3} = \frac{2\tilde{Z}_{\gamma j}(\ln y - Z^T_\gamma \gamma)}{\sigma^3_3}
\]

When \(\delta_y = 0\)

\[
B(\gamma_{j\sigma_3}) = -\frac{\partial^2 \ell_1}{\partial \gamma_j \partial \sigma_3} = \frac{3(\ln y - Z^T_\gamma \gamma)^2}{\sigma^4_3} - \frac{1}{\sigma^2_3}
\]

\[
B(\sigma_1) = -\frac{\partial^2 \ell_1}{\partial \sigma_1^2} = \frac{3(\ln x_1 - Z^T_\alpha \alpha)^2}{\sigma^4_1} - \frac{1}{\sigma^2_1}
\]

\[
B(\sigma_2) = -\frac{\partial^2 \ell_1}{\partial \sigma_2^2} = \frac{3(\ln x_2 - Z^T_\beta \beta)^2}{\sigma^4_2} - \frac{1}{\sigma^2_2}
\]

When \(\delta_y = 1\)

\[
B(\sigma_3) = -\frac{\partial^2 \ell_1}{\partial \sigma_3^2} = \frac{3(\ln y - Z^T_\gamma \gamma)^2}{\sigma^4_3} - \frac{1}{\sigma^2_3}
\]

When \(\delta_y = 0\)

\[
B(\sigma_3) = -\frac{\partial^2 \ell_2}{\partial \sigma_3^2} = \frac{V \phi(V)}{\sigma^2_3(1 - \Phi(V))} - \frac{1}{\sigma^2_3}
\]

Where \(\delta_y = 0\) and \(V = \frac{\ln y - Z^T_\gamma \gamma}{\sigma_3}, \frac{dV}{d\sigma_3} = \frac{-(\ln y - Z^T_\gamma \gamma)}{\sigma^2_3}\) and \(d^2 \phi(V)/dV^2 = -V \phi(V)\)

### 4.2 Quadrature Implementation

When computing the E-step two integrals must be estimated using quadrature. The first integral is the normalizing constant \(N\) in the denominator of \(h\). The second integral is the
target function \( Q(\theta|\theta^{(t)}) \). While I tried to choose the quadrature rule to fit the functional form of the integrand because I believe this will result in a more accurate approximation, I discovered that using a different quadrature rule for the normalizing constant \( N \) than for expectation of \( Q(\theta|\theta^{(t)}) \) results in calculated weights that do not sum to 1 and lead to biased estimates. Therefore, I used Laguerre quadrature with \( \alpha=0 \) for all numerical integration.

The choice of a scaling factor \( \lambda \) is a difficult one as the accuracy of the results depend on the accuracy of the numerical integration. Choosing \( \lambda_{X_1} = \lambda_{X_2} = \lambda_Y = c \) where \( c \) is a constant may be appropriate if all three followed the same marginal distribution but in application they may not. Choosing a constant \( \lambda \) may offer adequate coverage of one variable while having poor coverage of another variable when the marginals are very different. To remedy this, I choose \( \lambda_{X_1}, \lambda_{X_2}, \) and \( \lambda_Y \) to depend on the means of the marginal distributions of \( X_1, X_2, \) and \( Y \) respectively. Since \( \theta \) is updated with every iteration of the EM algorithm, I also update the scaling factors with every iteration.

### 4.3 Empirical Distribution of the Minimum

To derive \( \frac{\partial f(x_1,x_2,y|\theta_2)}{\partial \theta} \) where \( \theta = (\alpha_0, \beta_0, \beta_1, \gamma_0, \gamma_1, \gamma_2) \), first denote \( f_{X_1}(x_1|\theta) = e^{-\alpha_0}e^{-\alpha_0 x_1} \), \( f_{X_2}(x_2|\theta) = e^{-(\beta_0+\beta_1 x_1)}e^{-(\beta_0+\beta_1 x_1)x_2} \), and \( f_Y(y|\theta) = e^{-(\gamma_0+\gamma_1 x_1+\gamma_2 x_2)}e^{-(\gamma_0+\gamma_1 x_1+\gamma_2 x_2)y} \).
The following partial derivatives are necessary Equation 3.6:

\[
\frac{\partial f(x_1, x_2, y|\hat{\theta}_2)}{\partial \alpha_0} = e^{-x_1 e^{-\hat{\beta}_0 - 2\hat{\alpha}_0}} (x_1 - e^{\hat{\beta}_0}) f_{X_2}(x_2|\hat{\theta}_2) f_Y(y|\hat{\theta}_2)
\]

\[
\frac{\partial f(x_1, x_2, y|\hat{\theta}_2)}{\partial \beta_0} = e^{-x_2 e^{-\hat{\beta}_0 + \hat{\beta}_1 x_1}} - 2\beta_0 (x_2 - e^{\hat{\beta}_0 + \hat{\beta}_1 x_1}) f_{X_1}(x_1|\hat{\theta}_2) f_Y(y|\hat{\theta}_2)
\]

\[
\frac{\partial f(x_1, x_2, y|\hat{\theta}_2)}{\partial \beta_1} = e^{-x_2 e^{-\hat{\beta}_0 + \hat{\beta}_1 x_1}} - 2\hat{\beta}_0 (x_2 - e^{\hat{\beta}_0 + \hat{\beta}_1 x_1}) f_{X_1}(x_1|\hat{\theta}_2) f_Y(y|\hat{\theta}_2)
\]

\[
\frac{\partial f(x_1, x_2, y|\hat{\theta}_2)}{\partial \gamma_0} = e^{-y e^{-\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2}} - 2\gamma_0 (y - e^{\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2}) f_{X_1}(x_1|\hat{\theta}_2)
\]

\[
\times f_{X_2}(x_2|\hat{\theta}_2)
\]

\[
\frac{\partial f(x_1, x_2, y|\hat{\theta}_2)}{\partial \gamma_1} = e^{-y e^{-\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2}} - 2\gamma_0 (y - e^{\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2})
\]

\[
\times f_{X_1}(x_1|\hat{\theta}_2) f_{X_2}(x_2|\hat{\theta}_2)
\]

\[
\frac{\partial f(x_1, x_2, y|\hat{\theta}_2)}{\partial \gamma_2} = e^{-y e^{-\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2}} - 2\gamma_0 (y - e^{\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2})
\]

\[
\times f_{X_1}(x_1|\hat{\theta}_2) f_{X_2}(x_2|\hat{\theta}_2)
\]

4.4 Simulations

All simulations and analyses results were performed using SAS/BASE®, SAS/STAT® and SAS/IML® software, Versions 9.2 of the SAS System for Windows. Copyright ©2002–2008 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks of SAS Institute Inc., Cary, North Carolina, USA.
To best fit our motivating example, for Equation 4.2 and Equation 4.1 assume \( Z_p \) is a binary treatment indicator \( Z \) that can take on values 0 or 1. Also, assume that \( x_2 \) depends on the log-transformed \( x_1 \) and similarly \( y \) depend on the log-transformed \( x_1 \) and \( x_2 \) as below:

\[
\begin{align*}
\tilde{Z}_\alpha^T \alpha &= \alpha_0 + Z\alpha_1 \\
\tilde{Z}_\beta^T \beta &= \beta_0 + \log(x_1 + 1)\beta_1 + Z\beta_2 \\
\tilde{Z}_\gamma^T \gamma &= \gamma_0 + \log(x_1 + 1)\gamma_1 + \log(x_2 + 1)\gamma_2 + Z\gamma_3
\end{align*}
\]

(4.3)

4.4.1 Simulation Results

First, to test the estimating procedure I generate 1000 datasets each of \( n=1000 \) using the RAND function that follows exponential distribution Equation 4.2 such that: \( \alpha_0=1.5, \alpha_1=-0.1, \beta_0=2.0, \beta_1=5, \beta_2=-.1, \gamma_0=3.0, \gamma_1=.2, \gamma_2=.1, \) and \( \gamma_3=-.1 \). The variable \( Z \sim \text{Bernoulli}(0.5) \). I refer to this henceforth as Model A. This produces means for \( X_1=2.96, X_2=7.41, \) and \( Y=18.69 \). Administrative censoring of \( C_{X_1}=30, C_{X_2}=40, \) and \( C_Y=40 \) produces censoring of 11%, 29%, and 24% for \( X_1, X_2, \) and \( Y \) respectively. It induces Spearman’s partial rank-order correlation coefficients of \( \rho_{X_1,X_2}=.44, \rho_{X_1,Y}=.25, \) and \( \rho_{X_2,Y}=.21 \). The results below were generated using Laguerre quadrature with \( \alpha=0, 20 \) quadrature nodes and scaling factors calculated according to the method in Section 4.2. A relative convergence criteria of 1E-3 was used for this and all other estimating procedures. Quadrature nodes and weights were obtained with R:statmod package (36). Results are shown in Table I.
TABLE I

<table>
<thead>
<tr>
<th>Distribution parameter values</th>
<th>Mean parameter estimates</th>
<th>SD of parameter estimate</th>
<th>Mean standard error</th>
<th>95% CI coverage probability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>X1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0 = 1.50$</td>
<td>1.499</td>
<td>0.046</td>
<td>0.047</td>
<td>0.961</td>
</tr>
<tr>
<td>$\alpha_1 = -0.10$</td>
<td>-0.098</td>
<td>0.065</td>
<td>0.066</td>
<td>0.950</td>
</tr>
<tr>
<td>**X2</td>
<td>X1**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0 = 2.00$</td>
<td>2.001</td>
<td>0.087</td>
<td>0.089</td>
<td>0.950</td>
</tr>
<tr>
<td>$\beta_1 = 0.50$</td>
<td>0.499</td>
<td>0.051</td>
<td>0.052</td>
<td>0.950</td>
</tr>
<tr>
<td>$\beta_2 = -0.10$</td>
<td>-0.102</td>
<td>0.075</td>
<td>0.076</td>
<td>0.953</td>
</tr>
<tr>
<td>**Y</td>
<td>X1, X2**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_0 = 3.00$</td>
<td>3.001</td>
<td>0.113</td>
<td>0.115</td>
<td>0.947</td>
</tr>
<tr>
<td>$\gamma_1 = 0.20$</td>
<td>0.201</td>
<td>0.055</td>
<td>0.055</td>
<td>0.945</td>
</tr>
<tr>
<td>$\gamma_2 = 0.10$</td>
<td>0.099</td>
<td>0.042</td>
<td>0.042</td>
<td>0.954</td>
</tr>
<tr>
<td>$\gamma_3 = -0.10$</td>
<td>-0.099</td>
<td>0.073</td>
<td>0.075</td>
<td>0.953</td>
</tr>
</tbody>
</table>

For the second simulation (Model B) I chose $\theta$ to generate data similar to the actual VADT data: $\alpha_0=9.85$, $\alpha_1=.065$, $\beta_0=7.0$, $\beta_1=.3$, $\beta_2=.08$, $\gamma_0=5.68$, $\gamma_1=.42$, $\gamma_2=.17$, and $\gamma_3=-.28$. The variable $Z \sim$ Bernoulli(.5). This produces means for $X_1=13,590$, $X_2=12,742$, and $Y=41,543$. Administrative censoring of $C_{X_1}=2,555$, $C_{X_2}=2,555$, and $C_Y=2,555$ produces censoring of 88%, 87%, and 95% for $X_1$, $X_2$, and $Y$ respectively. It induces Spearman’s partial rank-order correlation coefficients of $\rho_{X_1,X_2}=.29$, $\rho_{X_1,Y}=.42$, and $\rho_{X_2,Y}=.28$. The results below were again generated using Laguerre quadrature with $\alpha=0$, 20 quadrature nodes and scaling factors calculated according to the method in Section 4.2. Results are shown in Table II.


For the third simulation (Model C) I chose $\theta$ to generate data similar to Model A but with Weibull distribution Equation 4.2 instead of exponential: $\alpha_0=1.5$, $\alpha_1=-.1$, $\sigma_1=1.2$,
\( \beta_0 = 2.0, \beta_1 = 0.5, \beta_2 = -0.1, \sigma_2 = 1.2, \gamma_0 = 3.0, \gamma_1 = 0.2, \gamma_2 = -0.1, \gamma_3 = -0.1, \) and \( \sigma_3 = 1.2. \) The variable \( Z \sim \text{Bernoulli}(0.5). \) This produces means for \( X_1 = 3379, \) \( X_2 = 10.67, \) and \( Y = 23.67. \) Administrative censoring of \( C_{X_1} = 40, \) \( C_{X_2} = 40, \) and \( C_Y = 40 \) produces censoring of 7\%, 28\%, and 27\% for \( X_1, \) \( X_2, \) and \( Y \) respectively. It induces Spearman’s partial rank-order correlation coefficients of \( \rho_{X_1, X_2} = 0.31, \rho_{X_1, Y} = 0.16, \) and \( \rho_{X_2, Y} = 0.13. \) The results below were again generated using Laguerre quadrature with \( \alpha = 0, \) 20 quadrature nodes, and scaling factors calculated according to the method in Section 4.2. Results are shown in Table III.

### Table III

<table>
<thead>
<tr>
<th>Module</th>
<th>Description</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>95% CI coverage probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_1 )</td>
<td>( \alpha_0 = 1.50 )</td>
<td>1.500</td>
<td>0.038</td>
<td>0.039</td>
<td>0.957</td>
</tr>
<tr>
<td></td>
<td>( \alpha_1 = -0.10 )</td>
<td>-0.103</td>
<td>0.054</td>
<td>0.054</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td>( \sigma_1 = 1.20 )</td>
<td>1.202</td>
<td>0.031</td>
<td>0.031</td>
<td>0.949</td>
</tr>
<tr>
<td>( X_2</td>
<td>X_1 )</td>
<td>( \beta_0 = 2.00 )</td>
<td>1.999</td>
<td>0.081</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 = 0.50 )</td>
<td>0.499</td>
<td>0.048</td>
<td>0.048</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td>( \beta_2 = -0.10 )</td>
<td>-0.098</td>
<td>0.061</td>
<td>0.061</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td>( \sigma_2 = 1.20 )</td>
<td>1.205</td>
<td>0.036</td>
<td>0.035</td>
<td>0.941</td>
</tr>
<tr>
<td>( Y</td>
<td>X_1, X_2 )</td>
<td>( \gamma_0 = 3.00 )</td>
<td>2.996</td>
<td>0.108</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>( \gamma_1 = 0.20 )</td>
<td>0.201</td>
<td>0.052</td>
<td>0.051</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>( \gamma_2 = 0.10 )</td>
<td>0.102</td>
<td>0.039</td>
<td>0.039</td>
<td>0.945</td>
</tr>
<tr>
<td></td>
<td>( \gamma_3 = -0.10 )</td>
<td>-0.102</td>
<td>0.060</td>
<td>0.062</td>
<td>0.949</td>
</tr>
<tr>
<td></td>
<td>( \sigma_3 = 1.20 )</td>
<td>1.204</td>
<td>0.040</td>
<td>0.039</td>
<td>0.945</td>
</tr>
</tbody>
</table>
For the fourth simulation (Model D) I chose $\theta$ to generate data similar to VADT with Weibull distribution Equation 4.2: $\alpha_0=9.44$, $\alpha_1=-.06$, $\sigma_1=1.30$, $\beta_0=6.95$, $\beta_1=.27$, $\beta_2=.09$, $\sigma_2=1.44$, $\gamma_0=6.04$, $\gamma_1=.35$, $\gamma_2=.14$, $\gamma_3=-.24$, and $\sigma_3=1.20$. The variable $Z \sim \text{Bernoulli}(0.5)$. This produces means for $X_1=4235$, $X_2=4251$, and $Y=4723$. Administrative censoring of $C_{X_1}=5000$, $C_{X_2}=5000$, and $C_Y=5000$ produces censoring of 77%, 77%, and 88% for $X_1$, $X_2$, and $Y$ respectively. It induces Spearman’s partial rank-order correlation coefficients of $\rho_{X_1X_2}=.29$, $\rho_{X_1Y}=.34$, and $\rho_{X_2Y}=.21$. The results below were again generated using Laguerre quadrature with $\alpha=0$, 20 quadrature nodes, and scaling factors calculated according to the method in Section 4.2. Results are shown in Table IV.
### TABLE IV

<table>
<thead>
<tr>
<th>Distribution parameter values</th>
<th>Mean parameter estimate</th>
<th>SD of parameter estimates</th>
<th>Mean standard error</th>
<th>95% CI coverage probability</th>
</tr>
</thead>
</table>

**X1**
- $\alpha_0 = 9.44$  
  Mean: 9.435  
  SD: 0.063
- $\alpha_1 = 0.06$  
  Mean: 0.059  
  SD: 0.069
- $\sigma_1 = 1.30$  
  Mean: 1.310  
  SD: 0.058

**X2|X1**
- $\beta_0 = 6.95$  
  Mean: 6.952  
  SD: 0.272
- $\beta_1 = 0.27$  
  Mean: 0.269  
  SD: 0.032
- $\beta_2 = 0.09$  
  Mean: 0.089  
  SD: 0.065
- $\sigma_2 = 1.44$  
  Mean: 1.455  
  SD: 0.065

**Y|X1, X2**
- $\gamma_0 = 6.04$  
  Mean: 6.046  
  SD: 0.667
- $\gamma_1 = 0.35$  
  Mean: 0.347  
  SD: 0.061
- $\gamma_2 = 0.14$  
  Mean: 0.142  
  SD: 0.076
- $\gamma_3 = -0.24$  
  Mean: -0.234  
  SD: 0.113
- $\sigma_3 = 1.20$  
  Mean: 1.211  
  SD: 0.078

#### 4.4.2 Survival Curves

Survival curves of complication-free survival time, defined as $\min(X_1, X_2, Y)$, are obtained via a quadrature approximation of Equation 3.4, again with $\alpha=0$, 20 quadrature nodes and scaling factor as in Section 4.2. The curves for Models A and B are shown in Figure 2 and Figure 3 respectively.
Figure 2. Model A: Complication-free Survival.
Figure 3. Model B: Complication-free Survival.
Similarly, by approximating Equation 3.7 I obtain the marginal survival curves in Figures 4 and 5 for Models A and B respectively.

Figure 4. Model A: Marginal Y Survival.
In order to demonstrate that the ISD test statistic in Equation 3.5 is distributed weighted Chi-squared where the weights are equal to the eigenvalues of $B$ I have conducted a simulation study using both Models A and B. The simulation procedure is as follows:

1. Generate 1000 datasets ($j = 1 \ldots 1000$) under the null hypothesis: $\alpha_1 = \beta_2 = \gamma_3 = 0$.

2. Calculate the test statistic $T_j = \int_0^\infty \left[ g(t|\hat{\theta}_P, \hat{\theta}_Z) - g(t|\hat{\theta}_P, \hat{\theta}_Z = 0) \right]^2 dt$ for each simulated dataset.
3. Calculate the eigenvalues of $B_j$ for each dataset and generate a weighted $\chi^2$ random variable $W_j$ for each dataset: $W_j = \lambda_1 A_1 + \lambda_2 A_2 + \lambda_3 A_3$, where $A_1$, $A_2$, $A_3$ are i.i.d. $\chi^2_1$.

4. Compare the empirical distribution of 1000 simulated $T$ and $W$.

Using the same simulated datasets for Model A in Section 4.4.1 I obtain Figure 6 of the empirical cumulative distribution functions (CDF)s of $T$ and $U$. The Kolmogorov-Smirnov two-sample test measures the maximum deviation of the empirical CDF within the groups from the pooled empirical CDF. In this simulation of Model A (n=1000) the asymptotic test statistic is 0.69 (p=0.72), not significant. I repeated the simulation under Model B (n=2000). Due to the higher censoring rate in Model B, the large-sample properties are not realized with n=1000 records per dataset, so I increased the sample size to n=2000. Figure 7 shows the plots of the empirical CDFs and the Kolmogorov-Smirnov asymptotic test statistic is 0.76 (p=0.61). The type-1 error rate for the Model B simulation was estimated to be 0.0486815. I conclude from that under the null-hypothesis the distribution of the test statistic $T$ is converging to a weighted $\chi^2$ random variable whose weights are the eigenvalues of $B$. This result can be used to quickly calculate a p-value for a given $T$-statistic.
Figure 6. Model A: Empirical CDF of T.
Figure 7. Model B: Empirical CDF of T.
CHAPTER 5

PARAMETRIC CONDITIONAL MODEL: VETERANS AFFAIRS

DIABETES TRIAL ANALYSIS

In this chapter I will present the results of the application of the above method to the VADT data. In Section 5.1 I describe the process for choosing a parametric distribution. In Section 5.2 I give some relevant details about the VADT data. In Section 5.3 I give the results of fitting the exponential version of the above method to the VADT data. In Section 5.4 I list the results of the Weibull version, including survival curves. In Section 5.5 I discuss the results of the analysis and the advantages and disadvantages of the method applied to the VADT data.

5.1 Parametric Distribution

In the presence of competing risks it is difficult to choose a parametric model for each of the three components in Equation 3.2. The generalized gamma distribution includes commonly used parametric distributions such as exponential, Weibull, and log-normal as special cases. It also can model increasing, decreasing, U-shaped, and inverted-U shaped (16). Table V shows the results of likelihood ratio tests of the nested generalized gamma/Weibull, and Weibull/exponential distributions for the marginal distributions of MI, CHF, stroke, and CV-death. In each case there is no evidence against using the simpler model. First, I use the exponential model for all three conditional distributions.
5.2 Description of the Data

There are \( n = 1,768 \) patients in this dataset (23 of the 1,791 had no follow-up visits). I model time until first CHF, MI, and CV-death corresponding to \( X_1, X_2, \) and \( Y \) respectively in the model specified in Equation 4.3. The events are rare, with 91%, 92%, and 96% censoring for CHF, MI, and CV-death. Survival times are measured in days. The maximum follow-up time was 2,693 days. For these data, all patients were censored on the same administrative censoring calendar date.

### Table V

<table>
<thead>
<tr>
<th>Event</th>
<th>Model</th>
<th>-2 Log-likelihood</th>
<th>( \chi^2(1) )</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>Gamma</td>
<td>1270.02</td>
<td>0.02</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>Lognormal</td>
<td>1270.04</td>
<td>2.98</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Weibull</td>
<td>1273.00</td>
<td>0.67</td>
<td>0.41</td>
</tr>
<tr>
<td>CHF</td>
<td>Gamma</td>
<td>1380.25</td>
<td>0.55</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>Lognormal</td>
<td>1380.80</td>
<td>1.27</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>1382.11</td>
<td>0.60</td>
<td>0.44</td>
</tr>
<tr>
<td>Stroke</td>
<td>Gamma</td>
<td>688.47</td>
<td>1.11</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>Weibull</td>
<td>689.57</td>
<td>1.80</td>
<td>0.18</td>
</tr>
<tr>
<td>CV-death</td>
<td>Gamma</td>
<td>669.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lognormal</td>
<td>669.65</td>
<td>0.20</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>Weibull</td>
<td>669.65</td>
<td>0.20</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>671.52</td>
<td>1.88</td>
<td>0.17</td>
</tr>
</tbody>
</table>
5.3 Exponential Model

For CHF and MI, time-to-first event is used while subsequent multiple events are ignored. For the quadrature routines Laguerre quadrature with $\alpha=0$ with 20 for the integration. The results are shown in Table VI.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>Standard error</th>
<th>95% Wald confidence interval</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$ CHF - Intercept</td>
<td>9.86</td>
<td>0.11</td>
<td>(9.66,10.1)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>$\alpha_1$ CHF - Treatment</td>
<td>0.11</td>
<td>0.15</td>
<td>(-0.19,0.40)</td>
<td>0.49</td>
</tr>
<tr>
<td>$\beta_0$ MI - Intercept</td>
<td>6.15</td>
<td>0.42</td>
<td>(5.34,6.96)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>$\beta_1$ MI - log(CHF+1)</td>
<td>0.42</td>
<td>0.05</td>
<td>(0.33,0.51)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>$\beta_2$ MI - Treatment</td>
<td>0.18</td>
<td>0.17</td>
<td>(-0.15,0.50)</td>
<td>0.29</td>
</tr>
<tr>
<td>$\gamma_0$ CV-death - Intercept</td>
<td>5.90</td>
<td>0.83</td>
<td>(4.28,7.53)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>$\gamma_1$ CV-death - log(CHF+1)</td>
<td>0.41</td>
<td>0.09</td>
<td>(0.24,0.58)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>$\gamma_2$ CV-death - log(MI+1)</td>
<td>0.15</td>
<td>0.11</td>
<td>(-0.05,0.36)</td>
<td>0.15</td>
</tr>
<tr>
<td>$\gamma_3$ CV-death - Treatment</td>
<td>-0.31</td>
<td>0.25</td>
<td>(-0.79,0.18)</td>
<td>0.21</td>
</tr>
</tbody>
</table>

I note that the parameters $\beta_1$ and $\gamma_1$ are statistically significant, indicating that the event times for MI and CV-death are correlated with event times for CHF but not with each other. The positive sign of the MLE of both $\beta_1$ and $\gamma_1$ means that an increase in the CHF-time signifies a decrease in the hazard for MI, and therefore an increase in expected MI-time. While none of the effects of treatment are statistically significant, $\gamma_3$, effect of treatment on
CV-death given $\log(\text{CHF}+1)$ and $\log(\text{MI}+1)$, is opposite in sign of either $\alpha_1$ or $\beta_2$. One possible explanation for such a mechanism is that the treatment may decrease the frequency of MI events, while simultaneously increasing the severity of MI events, leading an increase in the hazard for CV-death. A fatal MI and nonfatal MI are essentially the same event with different outcomes.

The Wald test statistic of $H_0 : \alpha_1 = \beta_2 = \gamma_3 = 0$ is $\chi^2(d.f. = 3)$ (p=0.37). The ISD test statistic is 32393.2 (p=0.43). I interpret this as there is no evidence of difference between the estimated complication-free or marginal survival curves shown in Figures 8 and 9.

Figure 8. VADT: Exponential Model, Complication-free Survival.
5.4 Weibull Model

Since the results of the exponential model were not significant, I tried fitting the more general Weibull model, the results of which are shown below. As people age and diabetes advances, I would expect a patient’s risk profile to change. To be more specific, I would expect a monotonically increasing hazard function with respect to time. The Weibull distribution, unlike the exponential distribution, allows for a monotonically increasing hazard function. However, none of the three shape parameters are significantly different from 1.0, so it makes sense to
abandon the Weibull model in favor of the parsimonious exponential model. The Wald-test statistic of $H_0 : \alpha_1 = \beta_2 = \gamma_3 = 0$ is 3.01 is $\chi^2$(d.f.=3) (p=0.39).

5.5 Discussion of Analysis and Method

For comparison purposes I fit a Cox-PH model on the complication-free survival time and the likelihood ratio test statistic for no group difference was 1.49 on $\chi^2$(d.f.=1) (p=0.22). I also performed a log-rank test, the statistic of which was also 1.49 on $\chi^2$(d.f.=1) (p=0.22). While the simple Cox-PH model based on a single composite endpoint appears to be more powerful in detecting treatment difference, it is more difficult to interpret because we do not
know exactly how the risk profile of the three events have changed with treatment. It is also important to note the significant correlation between CHF and MI and between CHF and CV-death in both Weibull and exponential models. This finding could be used as evidence in a larger argument in support of using CHF as a surrogate outcome for CV-death. It is also interesting to note the difference in the sign of the parameter estimate for intensive treatment effect on CV-death versus intensive treatment effect on CHF or MI. This is consistent with the fact that there were more CV-deaths counted in the intensive treatment group than in the standard group, while the other outcomes favored standard treatment.

To summarize, the above method has several advantages: (1) It can model three types of events simultaneously, including one terminal event. (2) It explicitly models the interdependence of different types of events. (3) It allows for unique treatment effect parameters for different types of events. (4) It allows for different censoring times for different types of events.

The above method does have several disadvantages. A major shortcoming is that in the current implementation it only models the time until first event of a given type. Table VIII shows the frequency of patients who had specific numbers of a given type of event. For the CHF and invasive coronary revascularization (ICR) events, there are many patients who had more than one event. The above method discards any information beyond the first event for a given patient. The above method could be modified to accommodate recurrent events by adding additional conditional models for the $k$th ordered event. Adding additional conditional models beyond the third would increase the dimensions of the quadrature routine and increase computation time exponentially with the number of dimensions. Given the resources available
to me, it wouldn’t be practical to fit more than three models: time-to-first CHF, time-to-second
CHF, and time to CV-death. The additional information provided by adding the second CHF
would not be worth discarding information about the time-to-first MI.

### TABLE VIII

**FREQUENCY OF VADT PATIENTS WITH MULTIPLE EVENTS**

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Number of Events</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>&gt; 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>130</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>CHF</td>
<td>213</td>
<td>52</td>
<td>20</td>
<td>10</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Coronary Revasc.</td>
<td>176</td>
<td>43</td>
<td>12</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Inoperable CAD</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Amputation</td>
<td>22</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Peripheral Revasc.</td>
<td>27</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td>61</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Cerebral Revasc.</td>
<td>27</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Any</td>
<td>228</td>
<td>103</td>
<td>80</td>
<td>43</td>
<td>17</td>
<td>14</td>
<td>6</td>
<td>4</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Another shortcoming of the above method is its reliance on parametric models for the
hazard functions of the various types of events. Fitting parametric models requires unverifiable
assumptions about the underlying distribution of the data. In the above example, comparing
the results from our method to a Cox model for time-to-first event analysis in a test of overall
treatment effect, the Cox model appears to be more powerful. This is perhaps due to the lack
of fit of my chosen parametric models to the data. It may also be due to the fewer degrees
of freedom spent estimating a simpler univariate Cox model based on a composite outcome as opposed to a true multivariate model. A multivariate semi-parametric Cox regression model would free us from restrictive parametric assumptions.
In Section 6.1 I review a general model on which my proposed model is based. In Section 6.2 I describe the proposed model and estimation procedures in detail. Section 6.3 covers estimation of the variance of the parameter estimates while Section 6.4 covers inference and hypothesis testing.

### 6.1 Three-level Hierarchical Proportional-Hazards Model

Many methods for modeling recurrent events focus on a stratified approach, modeling each ordered recurrent event of a given event type separately (64). This method is cumbersome and uses many degrees of freedom, especially when the number of patients who experienced multiple events is small. A more parsimonious approach would be a Cox-type model with a frailty parameter. One such model is the multivariate Cox-type model with normal frailty used by Zhu et al. 2011 (77). We can model the $K + 1$ intensity processes corresponding to $K$ recurrent events and one terminal event:

\[
A_{ik}(t|Z_i; b_k) = \int_0^t e^{\alpha_k^T Z_i + b_k} dA_{0k}(s) \tag{6.1}
\]

\[
\Lambda_i(t|Z_i; b_i) = \int_0^t e^{\beta^T Z_i + \phi^T b_i} dA_0(s)
\]

82
where \( b = (b_1, ..., b_K) \sim \mathcal{N}(0, \Sigma) \) and \( \phi = (\phi_1, ..., \phi_K) \) are column vectors and

\[
\Sigma = \begin{bmatrix}
\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1K} \\
\sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2K} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{K1} & \sigma_{K2} & \cdots & \sigma_K^2
\end{bmatrix}
\]

The \( \alpha_k \)s and \( \beta \)s are regression parameters for the effects of treatment or other covariates on the \( k \)th intensity process and terminal event process respectively. In Equation 6.1, \( \phi \) is a column vector of unknown constants representing the degree of dependence of the terminal event on the underlying patient- and event-specific randoms effect of the recurrent events. In this model the event history of recurrent events at time \( t \) have no direct effect on the terminal event. The model is intuitive in that the effect of the risk for a patient to have multiple events \( b_i \) on the hazard of the terminal event \( \Lambda_i(t|Z_i; b_i) \) is represented by the parameter \( \phi \). Thus the hazard of the terminal event depends on the hazards of the recurrent events through the random effects \( b_i \) and the hazards of the recurrent events depend on each other through the covariance of the multivariate normal distribution of the random effects.

### 6.2 New Proposed Three-level Hierarchical Proportional-Hazards Model

First I will divide the types of recurrent events into two distinct types each with baseline hazard functions denoted type-1 and type-2. Type-1 events have \( \sigma_k \geq 0 \) and type-2 events has \( \sigma_k = 0 \). This corresponds to knowing \( \sigma_k = 0 \) for some types of events. This assumption allows for flexible models for the hazard function, and it also allows us to handle cases where the number
of repeated events are too low to reliably estimate $\sigma_k$. I will assume that $\sigma_k$ may not be zero for $k=1, \cdots, K$ and $\sigma_k=0$, $k=K+1, \cdots, L$.

### 6.2.1 Defining Time

In Model Equation 6.1, the hazard functions are functions of time $t$ from time of entry into the study for all patients until the end of study. The probability of an event at time $t$ defined as $P\{dN_{ik} = 1|a_i, b_{ik}, Z_{i}\}$ is independent of the event history $N_{ik}(t-)$ at time $t$. In the proposed application this may not be the case. When dealing with CV events that lead to hospitalization as in the VADT, the event rate may be lower immediately following the event due to the medical care received. Therefore, redefining the hazard function in terms of the gap-time $t^*_{ijk} = t^{(j)}_{ik} - t^{(j-1)}_{ik}$ for the $j$th event of the $i$th patient, $k$th event-type is logical. The result is a hazard function that may increase monotonically from $t=0$ until the occurrence of an event, after which the hazard function may “reset” to something lower than it was immediately before the event. Simulating the data in this manner produces data that more closely resemble VADT data in terms of the timing and total number of repeated events per patient. Therefore, the proposed model is defined in terms of the gap-times $t^*_{ijk}$, with the * notation left out for brevity.

\[
A_{ik}(t|Z_i; a_i, b_{ik}) = \int_0^t e^{\alpha_k^T Z_i + b_{ik} + a_i} dA_{0k}(s) \quad \forall k = 1, \ldots, K \\
A_{ik}(t|Z_i; a_i, b_{ik}) = \int_0^t e^{\alpha_k^T Z_i + a_i} dA_{0k}(s) \quad \forall k = K + 1, \ldots, L \\
\Lambda_i(t|Z_i; a_i, b_i) = \int_0^t e^{\beta^T Z_i + \sum_{k=1}^K \phi_k(b_{ik} + a_i)} d\Lambda_0(s)
\]
One drawback of the model in Equation 6.1 is that it will require estimation of the unstructured variance-covariance matrix. In applications such as the VADT analysis with a large number of recurrent event types $K$ and a relatively small number of recurrent events per patient, there is sparse information on all parameters of the unstructured variance-covariance matrix, resulting in increased convergence time and estimates that could be highly variable and unstable. A more practical model for the proposed application is the model in Equation 6.3 in which the unstructured variance-covariance matrix is replaced with a diagonal matrix and adding a random effect $a_i$ that is shared by an individual across different types of recurrent events, which follows $a_i \sim \mathcal{N}(0, \sigma_a^2)$, $b_i = (b_{i1}, \ldots, b_{iK}) \sim \mathcal{N}(0, \Sigma)$, $\phi = (\phi_1, \ldots, \phi_K)$ are column vectors, $a_i \perp b_i$ and

$$\Sigma = \begin{bmatrix}
\sigma_1^2 & 0 & \cdots & 0 \\
0 & \sigma_2^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma_K^2
\end{bmatrix}$$

For type-1 events, one may also think of the proposed configuration as a new vector defined as $b_i^* = b_i + a_i$, distributed $\mathcal{N}(0, \Sigma^*)$, interpreted as the combination of the event-specific random effects and the patient-specific random effect. This implies that correlation of event type $k$ with another event type $k'$ $\forall k = 1, \ldots, K$ is always positive given by $\rho_{kk'} = \frac{\sigma_a^2}{\sqrt{\sigma_k^2 + \sigma_a^2} \sqrt{\sigma_{k'}^2 + \sigma_a^2}}$. 


\[
\Sigma^r = \begin{bmatrix}
\sigma^2_1 + \sigma^2_a & \sigma^2_a & \cdots & \sigma^2_a \\
\sigma^2_a & \sigma^2_2 + \sigma^2_a & \cdots & \sigma^2_a \\
\vdots & \vdots & \ddots & \vdots \\
\sigma^2_a & \sigma^2_a & \cdots & \sigma^2_K + \sigma^2_a 
\end{bmatrix}
\]

6.2.2 Log-likelihood

The observed data log-likelihood of Model Equation 6.3 can be written

\[
\ell(O_i) = \sum_{i=1}^{n} \log \int_a \int_b \int_{t_{ijk}}^{t_{ij(k)}} f(O_i|b_i, a_i) f(b, a) db da
\]

\[
= \sum_{i=1}^{n} \log \int_a \int_b \prod_{k=1}^{K} \left[ \prod_{j=1}^{m_{ik}} \{ A_{0k}\{t_{ijk}\} e^{\alpha_k^T Z_i + b_{ik} + a_i} \} \Delta N_k(t_{ijk}) \right]
\]

\[
\times e^{-\int_0^{t_{ijk}} \alpha_k^T Z_i + b_{ik} + a_i \text{d} A_{0k}(s)} \int_{t_{ijk}}^{t_{ij(k)}} \prod_{k=1}^{K} \left[ \prod_{j=1}^{m_{ik}} \{ A_{0k}\{t_{ijk}\} e^{\alpha_k^T Z_i + a_i} \} \Delta N_k(t_{ijk}) \right]
\]

\[
\times e^{-\int_0^{t_{ijk}} \alpha_k^T Z_i + a_i \text{d} A_{0k}(s)} \int_{t_{ijk}}^{t_{ij(k)}} \prod_{k=1}^{K} \left[ \prod_{j=1}^{m_{ik}} \{ A_{0k}\{t_{ijk}\} e^{\alpha_k^T Z_i + a_i} \} \Delta N_k(t_{ijk}) \right]
\]

\[
\times \left[ \Lambda_0\{Y_i\} e^{\beta^T Z_i + \sum_{k=1}^{K} \phi_k (b_{ik} + a_i)} \right] \Delta_i - \int_0^{y_i} e^{\beta^T Z_i + \sum_{k=1}^{K} \phi_k (b_{ik} + a_i) \text{d} A_0(s)} \right] \} f(b_i, a_i) db da
\]

The observed data are \( O_i = \{Y_i, \Delta_i, \Delta N_{ik}(t_{ijk}), t_{ijk}; t_{ijk} \leq (Y_i-t_{i(j-1)k}), k = 1, ..., L, j = 1, ..., m_{ik} \}_{i=1}^{n} \) where \( t_{ijk} = \min(C^*_{ik}, T_{ijk}) \), \( T_{ijk} \) is the elapsed time of the \( j \)th event of the \( k \)th event type for the \( i \)th patient since the \((j-1)\)th event of type \( k \) of the \( i \)th patient, \( C^*_{ik} \) is the elapsed censoring time for the \( i \)th patient since the \((j-1)\)th event of type \( k \) for the \( i \)th patient, \( m_{ik} \) is the number of \( k \)th events for the \( i \)th patient, \( Y_i = \min(C_i, T_i) \) (\( C_i \) is the independent censoring time, \( T_i \) is the terminal event time for the \( i \)th patient), \( \Delta N_{ik}(t_{ijk}) \) is the number of
events of type $k$ at time $t_{ijk}$ for patient $i$, and $A_{0k}\{t\}$ and $\Lambda_0\{t\}$ are the first derivatives of $A_{0k}(t)$ and $\Lambda_0(t)$ and $f(b_i; a_i) = \mathcal{N}(0, \sigma^2_a) * \mathcal{N}(0, \Sigma)$.

The goal is to obtain

$$\hat{\theta} = \{\hat{\alpha}, \hat{\beta}, \hat{\phi}, \sigma^2_k, \sigma^2_a, \hat{\Lambda}_0\{t_{ijk}\}, \Lambda_0\{Y_i\} \forall l = 1, ..., L, k = 1, ..., K \text{ and } j = 1, ..., m_{il}\}.$$  

6.2.3 Expectation Conditional Maximization Algorithm

The observed data log-likelihood is difficult to maximize directly due to the integration over the random effects $b$ and $a$. The ECM algorithm (51), by treating the $b_i$ and $a_i$ as missing, is useful here. The augmented data log-likelihood function that includes $b_i$ and $a_i$ as the augmented data are

$$\ell(A_i) = \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{j=1}^{m_{ik}} \left[ \Delta N_{ik}(t_{ijk}) \{ \log A_{0k}\{t_{ijk}\} + \alpha_k^T Z_i + b_{ik} + a_i \} \right]$$

$$- e^{\alpha_k^T Z_i + b_{ik} + a_i} A_{0k}(t_{ijk}) \right]$$

$$+ \sum_{i=1}^{n} \sum_{k=K+1}^{L} \sum_{j=1}^{m_{ik}} \left[ \Delta N_{ik}(t_{ijk}) \{ \log A_{0k}\{t_{ijk}\} + \alpha_k^T Z_i + a_i \} \right]$$

$$- e^{\alpha_k^T Z_i + a_i} A_{0k}(t_{ijk}) \right]$$

$$+ \Delta_i \{ \log \Lambda_0\{Y_i\} + \beta^T Z_i + \sum_{k=1}^{K} \phi_k(b_{ik} + a_i) \}$$

$$- e^{\beta^T Z_i + \sum_{k=1}^{K} \phi_k(b_{ik} + a_i)} \Lambda_0(Y_i) + \log f(b_i; \sigma^2_1 ... \sigma^2_K) + \log f(a_i; \sigma^2_a)$$
Taking the conditional expectation with the current parameter estimates, the expected augmented data log-likelihood becomes

\[
E[\ell(A_i)] = \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{j=1}^{m_{ik}} \left[ \Delta N_{ik}(t_{ijk}) \{ \log A_{0k}(t_{ijk}) + \alpha_k^T Z_i + E[a_i] \right] \\
- n \sum_{i=1}^{n} \sum_{k=1}^{K+1} \sum_{j=1}^{m_{ik}} \left[ \Delta N_{ik}(t_{ijk}) \{ \log A_{0k}(t_{ijk}) + \alpha_k^T Z_i + E[a_i] \right] \\
- e^{\alpha_k^T Z_i} E[e^{b_{ik} + a_i}] A_{0k}(t_{ijk}) \\
+ \Delta_i \{ \log \Lambda_0(Y_i) + \beta^T Z_i + \sum_{k=1}^{K} \phi_k( E[b_{ik}] + E[a_i]) \} \\
- e^{\beta^T Z_i} E[e^{\sum_{k=1}^{K} \phi_k(b_{ik} + a_i)} \Delta_0(Y_i) - \sum_{k=1}^{K} \left\{ \frac{1}{2} \log \sigma_k^2 + \frac{E[b_{ik}^2]}{2\sigma_k^2} \right\} - \frac{1}{2} \log \sigma_a^2 - \frac{E[a_i^2]}{2\sigma_a^2} \\
\]

This is the objective function to be maximized in the ECM algorithm.

**6.2.4 Expectation Step**

In order to maximize the expected augmented data log-likelihood, it is necessary to evaluate

\[
E\{g(b_i)|O_i\} = \int g(b_i)f(O_i|b_i)f(b_i)db / \int f(O_i|b_i)f(b_i)db
\]

The part of \( f(O_i|b_i) \) that does not depend on random effects \( b_i \) and \( a_i \) can be removed from both the numerator and denominator leaving

\[
E\{g(b_i)|O_i\} = \int g(b_i)\Gamma_i(b)f(b_i)db / \int \Gamma_i(b)f(b_i)db
\]
Where

\[
\Gamma_i(b, a) = \prod_{k=1}^{K} \left[ \prod_{j=1}^{m_k} \{ e^{b_{ik}+a_i} \} \Delta N_k(t_{ijk}) e^{-\int_{0}^{t_{ijk}} e^{\alpha_k Z_i+b_{ik}+a_i} dA_{0k}(s)} \right]^{m_k}
\]

In Equation 6.6 in the absence of time-varying covariates the Reimann-Steiltjes integrals simplify to

\[
\int_{0}^{t_{ik}} e^{\alpha_k Z_i+b_{ik}+a_i} dA_{0k}(s) = e^{\alpha_k Z_i+b_{ik}+a_i} A_{0k}(t_{ik}) \forall k = 1, \ldots, K
\]

\[
\int_{0}^{t_{ik}} e^{\alpha_k Z_i+b_{ik}+a_i} dA_{0k}(s) = e^{\alpha_k Z_i+b_{ik}+a_i} A_{0k}(t_{ik}) \forall k = K + 1, \ldots, L
\]

\[
\int_{0}^{Y_i} e^{\beta Z_i+\phi b_i} d\Lambda_0(s) = e^{\beta Z_i+\phi b_i} A_0(Y_i)
\]

To evaluate one-dimensional integrals of which there exists no closed form solution one could use a Gauss-Hermine quadrature rule with \( M \) number of fixed nodes \( v_m \) and and weights \( w_m \).

\[
\int_{-\infty}^{+\infty} e^{-u^2} g(u) \, du \approx \sum_{m=1}^{M} w_m g(v_m)
\]

To extend this rule to multidimensional integrals the multivariate quadrature rule involves calculating the tensor product of the nodes and weights. Extending a rule with \( M \) nodes into \( k \) dimensions requires evaluation of the integrand \( M^k \) times, making extending the rule
into more than a few dimensions unfeasible. Another approach taken by Zhu et al. 2011 (77) is Monte-Carlo integration. Monte-Carlo integration shares the same curse of dimensionality as Gauss-Hermite quadrature, but with possibly many more function evaluations to achieve the same level of accuracy. In addition, there is potentially a significant amount of Monte-Carlo error that may cause problems with ascertaining convergence of the ECM algorithm.

A relatively new technique for approximating multidimensional integrals that has seen use in some econometric literature is sparse-grid quadrature. The technique is based on a general rule extending univariate operators into multiple dimensions introduced by Smolyak 1963 (68). It is valid in general for extension of any one-dimensional quadrature rule or combination of different types of quadrature rules. The advantage is that the number of function evaluations does not rise exponentially as it would with a standard multivariate quadrature rule but significantly slower. Use of sparse-grid quadrature has been illustrated in Heiss and Winschel 2008(37) in a micro-econometric latent dependent variable model. The authors have provided nodes and weights (http://www.sparse-grids.de/) for various combinations of number of dimensions dimensions and accuracy level.

The main difference between a standard multivariate quadrature rules and a sparse-grid rule is that the first is exact for a tensor product of univariate polynomials while the second is exact for a class of polynomials of a bounded total order. Borrowing the example from Heiss and Winschel 2008 in two dimensions sparse-grid is exact for polynomials involving the terms $x_1$, $x_2$, $x_1^2$, $x_2^2$, and $x_1x_2$. The standard multivariate quadrature rule is exact for polynomials including these terms and additionally $x_1^2x_2$, $x_2^2x_1$, and $x_1^2x_2^2$ that do little to help speed up
convergence of the approximation. The rule for dimension $K$ at accuracy level $M$ is a weighted sum of product rules with different combinations of accuracy levels for each dimension. The sum is bounded similarly to how the sum of exponents of a multivariate polynomial is bounded, leading to fewer number of nodes than the standard multivariate quadrature rule.

While determination of nodes and weights is different in sparse-grid, implementation is identical to the standard multivariate quadrature rule. A $(K+1)$ by $m$ dimension matrix (each row a unique combination of nodes is assigned a weight) is given and the integrand is evaluated at each combination of nodes and the approximate integral is taken as the weighted sum of these. For the current application a nested quadrature rule is used with Gaussian weight of the form $f(x) = \exp(x^T x/2)/\sqrt{2\pi}$. In order to improve the accuracy of the quadrature, it helps to improve coverage of the integrand by the nodes by scaling and shifting the standard nodes. In the integrand we have $f(b_i, a_i) = \mathcal{N}(0, \text{diag}(\sigma_1^2, \ldots, \sigma_K^2, \sigma_a^2))$ as part of the likelihood function. I first scale the nodes $v = [v_1, \ldots, v_{(K+1)m}]$ by $\hat{\Sigma} = \text{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_K^2, \hat{\sigma}_a^2)$ and shift by $[\hat{b}_i, \hat{a}_i]$:

$$v = v^* \hat{\Sigma}^{1/2} + [\hat{b}_i, \hat{a}_i]$$

It is convenient to use the distribution of the random effects as the weight: $f(b_i, a_i) = (2\pi)^{-\frac{(K+1)}{2}} \exp(-[b_i, a_i] \Sigma^{-1} [b_i, a_i]^T / 2)$. Making the following substitutions in order to absorb much of $f(b_i, a_i)$ into the weight: $[b_i, a_i] = [b^*_i, a^*_i] \hat{\Sigma}^{1/2} + [\hat{b}_i, \hat{a}_i]$ and $d[b_i, a_i] = d[b^*_i, a^*_i] \hat{\Sigma}^{1/2}$ means that
\[ \int_a^b g(b_i, a_i) \Gamma_i(b, a) f(b_i, a_i) \, db \, da = \int_a^b \int_{b_i}^{b_i^*} g^*(b_i^*, a_i^*) \Gamma_i^*(b_i^*, a_i^*) f^*(b_i^*, a_i^*) |\Sigma|^{1/2} \, db^* \, da^* \quad (6.7) \]

First substituting into \( f(b_i, a_i) \) and expanding

\[ f^*(b_i^*, a_i^*) = (2\pi)^{-\frac{(K+1)}{2}} \exp(*) \]

Where

\[ (*) = -\left( [b_i^*, a_i^*]\Sigma^{-1/2} + [\hat{b}_i, \hat{a}_i]\right) \Sigma^{-1}\left( [b_i^*, a_i^*]\Sigma^{-1/2} + [\hat{b}_i, \hat{a}_i]\right)^T / 2 \]

\[ = -\left( [b_i^*, a_i^*][b_i^*, a_i]^T + 2[b_i^*, a_i^*]\Sigma^{-1/2}[\hat{b}_i, \hat{a}_i] + [\hat{b}_i, \hat{a}_i]\Sigma^{-1}[\hat{b}_i, \hat{a}_i]^T\right) / 2 \]

And finally

\[ f^*(b_i^*, a_i^*) = (2\pi)^{-\frac{(K+1)}{2}} e^{-\frac{|b_i^*, a_i|^2}{2}} e^{-\frac{|b_i^*, a_i|^T}{2} \Sigma^{-1/2}[\hat{b}_i, \hat{a}_i]^T} e^{-\frac{|b_i^*, a_i|^T}{2} \Sigma^{-1}[\hat{b}_i, \hat{a}_i]} \]

The equation Equation 6.7 becomes:

\[ (2\pi)^{-\frac{(K+1)}{2}} e^{-\frac{|b_i^*, a_i|^2}{2}} \]

\[ \int_{a_i^*}^{b_i^*} \int_{b_i^*}^{b_i^*} g^*(b_i^*, a_i^*) \Gamma_i^*(b_i^*, a_i^*) e^{-\frac{|b_i^*, a_i|^2}{2}} e^{-\frac{|b_i^*, a_i|^T}{2} \Sigma^{-1/2}[\hat{b}_i, \hat{a}_i]^T} \, db^* \, da^* \quad (6.8) \]
Let the Gaussian weight of the nodes be $w$ for the distribution \((2\pi)^{-\frac{(K+1)}{2}}e^{-\frac{[b^* - a^*]^T\Sigma^{-1}[b^* - a^*]}{2}}\)

The integral can be approximated with sparse grid quadrature as

$$
\approx e^{-\frac{[b_i - \hat{a}_i]^T\Sigma^{-1}[b_i - \hat{a}_i]}{2}} \sum_{m=1}^{M} w_m e^{-\frac{[b^*_i - a^*_i]^T\Sigma^{-1/2}[b^*_i - a^*_i]}{2}} g^*(b^*_i, a^*_i) \Gamma^*_i(b^*_i, a^*_i)
$$

Because the constants to the left of the summation can be canceled out in Equation 6.5 the numerator can be approximated by

$$
\approx \sum_{m=1}^{M} w_m e^{-v\Sigma^{-1/2}(\hat{b}_i - \hat{a}_i)^T} g^*(v^*) \Gamma^*_i(v^*)
$$

and the denominator approximated by

$$
\approx \sum_{m=1}^{M} w_m e^{-v\Sigma^{-1/2}(\hat{b}_i - \hat{a}_i)^T} \Gamma^*_i(v^*)
$$

**6.2.5 Conditional Maximization Step**

One iteration of the Newton-Raphson maximization algorithm during the conditional maximization step is sufficient to maximize the log likelihood with respect to

$$
\theta = \{\alpha_k, \beta, \phi, \Sigma, A_{0k}\{t_{j,k}\}, \Lambda_0\{Y_i\} \forall k = 1, ..., L \text{ and } j = 1, ..., m_{ik} \}.
$$

As per the log-likelihood function each of the three sets of parameters: $\theta_1 = \{\alpha_k, \beta, \phi\}$, $\theta_2 = \{\sigma^2_k \forall k = 1, ..., K, \sigma^2_a\}$ and $\theta_3 = \{A_{0k}\{t_{i,j,k}\}\forall k = 1, ..., L, \Lambda_0\{Y_i\}\}$ each of which can be
maximized separately and conditionally on the other parameters remaining fixed as described by Meng and Rubin (37).

1. Obtain \( \hat{\theta}_1 = \{\hat{\alpha}_k, \hat{\beta}, \hat{\phi}\} \) by performing one Newton-Raphson step:
\[
\hat{\theta}_3^{t+1} = \hat{\theta}_3^t + U(\hat{\theta}_3)\mathcal{I}^{-1}(\hat{\theta}_3) \quad \text{where} \quad \mathcal{I}(\hat{\theta}_1) \text{ is the expected value of the observed information matrix and } U(\hat{\theta}_1) \text{ is the expected value of the gradient vector.}
\]

2. Obtain \( \hat{\theta}_2 = \{\hat{\sigma}_k^2 \forall k = 1, \ldots, K, \hat{\sigma}_a^2\} \) by maximizing:
\[
\frac{\partial}{\partial \sigma_k^2} E[log f(b; \sigma_1^2, \ldots, \sigma_K^2)] \text{ and } \frac{\partial}{\partial \sigma_a^2} E[log f(a; \sigma_a)] \text{. Since they are both normally distributed with mean 0 the closed form solutions for the MLE are used:}
\]
\[
\hat{\sigma}_k^2 = \frac{1}{n} \sum_{i=1}^{n} E[b_{ik}^2] \forall k = 1, \ldots, K
\]
\[
\hat{\sigma}_a^2 = \frac{1}{n} \sum_{i=1}^{n} E[a_i^2]
\]

3. Obtain \( \hat{\theta}_3 = \{\hat{A}_{0k}\{t_{i,j,k}\} \forall k = 1, \ldots, L, \hat{A}_0\{Y_i\}\} \) using the Breslow-Day estimate as follows:
\[
\forall k = 1, \ldots, K:
\]
\[
E\left[ \frac{\partial \ell(O_i)}{\partial A_{0k}\{t_{j,k}\}} \right] = \sum_{i=1}^{n} \left\{ R_i(t_{j,k}) \Delta N_{ik}(t_{j,k}) A_{0k}\{t_{j,k}\} \right\}^{-1} - \sum_{i=1}^{n} \left[ R_i(t_{j,k}) e^{\alpha_k^T Z_i E[\epsilon b_{ik} + a_i]} \right]
\]
\[
\hat{A}_{0k}\{t_{j,k}\} = \frac{\sum_{i=1}^{n} R_i(t_{j,k}) \Delta N_{ik}(t_{j,k})}{\sum_{i=1}^{n} R_i(t_{j,k}) e^{\alpha_k^T Z_i E[\epsilon b_{ik} + a_i]}}
\]
\[ \forall k = K + 1, \ldots, L: \]

\[
E \left[ \frac{\partial \ell(O_i)}{\partial A_{0k}\{t_{j,k}\}} \right] = \sum_{i=1}^{n} \left\{ R_i(t_{j,k}) \Delta N_{ik}(t_{j,k}) A_{0k}\{t_{j,k}\}^{-1} \right\} - \sum_{i=1}^{n} \left[ R_i(t_{j,k}) e^{\alpha_k^T Z_i E[e_{a_i}]} \right]
\]

\[
\hat{A}_{0k}\{t_{j,k}\} = \frac{\sum_{i=1}^{n} R_i(t_{j,k}) \Delta N_{ik}(t_{j,k})}{\sum_{i=1}^{n} R_i(t_{j,k}) e^{\alpha_k^T Z_i E[e_{a_i}]}}
\]

Similarly

\[
\hat{\Lambda}_0\{Y_j\} = \frac{\sum_{i=1}^{n} R_i(Y_j) \Delta_i}{\sum_{i=1}^{n} R_i(Y_j) e^{\beta^T Z_i E[\sum_{k=1}^{K} \phi_k(b_{ik} + a_i)]}}
\]

For calculating \( I(\theta_1) \) and \( U(\theta_1) \) the first and second derivatives are given below as a reference.
First derivatives:

∀k = 1, ..., K :
\[
E \left[ \frac{\partial \ell(O_i)}{\partial \alpha_{kp}} \right] = \sum_{i=1}^{n} \left[ \sum_{j=1}^{m_k} \{ R_i(t_{j,k}) \Delta N_{ik}(t_{j,k}) Z_{ip} \} - Z_{ip} e^{\alpha_k^T Z_i A_{0k}(X_{ik})} E[e^{b_k+a_i}] \right]
\]

∀k = K + 1, ..., L :
\[
E \left[ \frac{\partial \ell(O_i)}{\partial \alpha_{kp}} \right] = \sum_{i=1}^{n} \left[ \sum_{j=1}^{m_k} \{ R_i(t_{j,k}) \Delta N_{ik}(t_{j,k}) Z_{ip} \} - Z_{ip} e^{\alpha_k^T Z_i A_{0k}(X_{ik})} e[a_i] \right]
\]

∀k = 1, ..., K :
\[
E \left[ \frac{\partial \ell(O_i)}{\partial \delta_{k}} \right] = \sum_{i=1}^{n} \delta_i Z_{ip} - Z_{ip} e^{\beta_k^T Z_i A_{0}(Y_i)} E[e^{\sum_{k=1}^{K} \phi_k(b_k+a_i)}]
\]

∀k = K + 1, ..., L :
\[
E \left[ \frac{\partial \ell(O_i)}{\partial \delta_{k}} \right] = \sum_{i=1}^{n} \delta_i Z_{ip} - Z_{ip} e^{\beta_k^T Z_i A_{0}(Y_i)} E[e^{\sum_{k=1}^{K} \phi_k(b_k+a_i)}]
\]
Second derivatives:

∀k = 1, ..., K :
\[
E \left[ \frac{\partial^2 \ell(O_i)}{\partial \alpha_k \partial \alpha_k} \right] = \sum_{i=1}^{n} -Z_{ip}Z_{iq}e^{\alpha_k^Tz_i}A_{0k}(Y_i)E[b_k + a_i] \]

∀k = K + 1, ..., L :
\[
E \left[ \frac{\partial^2 \ell(O_i)}{\partial \beta_p \partial \beta_q} \right] = \sum_{i=1}^{n} -Z_{ip}Z_{iq}e^{\beta^Tz_i}A_{0}(Y_i)E[e^{\sum_{k=1}^{K} \phi_k(b_k + a_i)}] \]

∀k = 1, ..., K :
\[
E \left[ \frac{\partial^2 \ell(O_i)}{\partial \phi_k \partial \phi_k} \right] = \sum_{i=1}^{n} -e^{\beta^Tz_i}A_{0}(Y_i)E[(b_k + a_i)e^{\sum_{k=1}^{K} \phi_k(b_k + a_i)}] \]

∀k, j = 1, ..., K :
\[
E \left[ \frac{\partial^2 \ell(O_i)}{\partial \phi_k \partial \phi_j} \right] = \sum_{i=1}^{n} -e^{\beta^Tz_i}A_{0}(Y_i)E[(b_k + a_i)(b_{ij} + a_i)e^{\sum_{k=1}^{K} \phi_k(b_k + a_i)}] \]

∀k = 1, ..., L :
\[
E \left[ \frac{\partial^2 \ell(O_i)}{\partial A_{0k} \{t_{j,k}\}^2} \right] = \sum_{i=1}^{n} -R_i(t_{j,k}) \Delta N_{ik}(t_{j,k})A_{0k}\{t_{j,k}\}^{-2} \]
∀k = 1, ..., K:

\[
E \left[ \frac{\partial^2 \ell(O_i)}{\partial A_0 t_{i,k} \partial \alpha_{kp}} \right] = \sum_{i=1}^{n} \left[ -R_i(t_{j,k})Z_{ip}e^{\alpha_k^T Z_k}E[e^{b_k + a_i}] \right]
\]

∀k = K + 1, ..., L:

\[
E \left[ \frac{\partial^2 \ell(O_i)}{\partial A_0 t_{i,k} \partial \alpha_{kp}} \right] = \sum_{i=1}^{n} \left[ -R_i(t_{j,k})Z_{ip}e^{\alpha_k^T Z_k}E[e^{a_i}] \right]
\]

\[
E \left[ \frac{\partial^2 \ell(O_i)}{\partial \Lambda_0 Y_j} \right] = \sum_{i=1}^{n} \left[ -R_i(Y_j)\Delta_0 \Lambda_0 {Y_j}^{-2} \right]
\]

\[
E \left[ \frac{\partial^2 \ell(O_i)}{\partial \Lambda_0 Y_j \partial \beta_p} \right] = \sum_{i=1}^{n} \left[ -R_i(Y_j)Z_{ip}e^{\beta Y_k}Z_k E[e^{\sum_{k=1}^{K} \phi_k(b_{ik} + a_i)}] \right]
\]

∀k = 1, ..., K:

\[
E \left[ \frac{\partial^2 \ell(O_i)}{\partial \Lambda_0 Y_j \partial \phi_k} \right] = \sum_{i=1}^{n} \left[ -R_i(Y_j)e^{\beta Y_k}Z_k E[(b_{ki} + a_i)e^{\sum_{k=1}^{K} \phi_k(b_{ik} + a_i)}] \right]
\]

### 6.3 Estimating the Variance-covariance Matrix

Louis’s method (48) is often used to calculate the variance-covariance matrix of \( \hat{\theta} \) in the presence of missing data. The parameter vector \( \hat{\theta} \) includes the parameters \( \hat{\theta}_3 = \{A_0 t_{i,k} \forall k = 1, ..., L, \Lambda_0 \{Y_i\} \} \) of which there are as many as there are observed events. This makes the information matrix of \( \hat{\theta} \) potentially very large and the inversion of said matrix difficult and potentially inaccurate due rounding error limited by the computer.

For this reason and since inference about the set of parameters \( \{\hat{\theta}_1, \hat{\theta}_2\} \) is of primary interest only the variance-covariance sub-matrix of \( V a r(\{\hat{\theta}_1, \hat{\theta}_2\}) \) is required for hypothesis testing and demonstrating 95% coverage probability in simulation. The EM-aided numerical differenti-
ation algorithm as described by Chen and Little (12) is used to calculate the variance-covariance matrix.

1. Perturb the jth component of the MLE \( \{ \hat{\theta}_1, \hat{\theta}_2 \} \) by \( d = \frac{1}{n} \) and call it \( \{ \hat{\theta}_1, \hat{\theta}_2 \}_j \)

2. Let \( \hat{\theta}_{3j} \) be the maximizer of \( E[\ell | O_i] \) with \( \{ \theta_1, \theta_2 \} \) fixed at \( \{ \hat{\theta}_1, \hat{\theta}_2 \}_j \), using the ECM algorithm outlined above.

3. Calculate the jth row of the matrix \( D \) by \( \frac{1}{d} E \left[ \frac{\partial \ell(O_i)}{\partial \{ \theta_1, \theta_2 \}} \right] \) with \( \theta \) fixed at \( \{ \hat{\theta}_1, \hat{\theta}_2 \}_j \) and \( \theta_{3j} \) fixed at \( \hat{\theta}_{3j} \).

4. Repeat steps 1-3 to obtain each row of \( D \) for each component of \( \{ \hat{\theta}_1, \hat{\theta}_2 \} \)

5. In order to force symmetry of the information matrix calculate the information matrix \( I = \frac{1}{2} \left[ (D + D^T) / 2 \right] \)

6. Some applications may require increased accuracy. It may be beneficial to repeat steps 1-5 with \( d = -\frac{1}{n} \) and take the average of the two estimated \( I \)s.

### 6.4 Wald Test

Testing \( H_0 : g(t|Z = 1) = g(t|Z = 0) \Rightarrow H_0 : \theta_Z = 0 \) Find the sub-matrix \( B_Z(\theta_Z) \) of the full information matrix \( B(\theta) \). Calculate the test statistic \( W = n \hat{\theta}_Z^T B_Z(\hat{\theta}_Z)^{-1} \hat{\theta}_Z \overset{d}{\to} \chi^2_r \), which is distributed Chi-squared with \( r \) degrees of freedom corresponding to the number of parameters being tested.
CHAPTER 7

THREE-LEVEL HIERARCHICAL PROPORTIONAL-HAZARDS MODEL: VETERANS AFFAIRS DIABETES TRIAL ANALYSIS

In this chapter I present results of the three-level proportional-hazards model applied to VADT data. Section 7.1 covers descriptive statistics relevant to the choice of models for this analysis. In Section 7.2 I review model assumptions, notation, and present the results of the model fit. Section 7.4 is a discussion of the results as well the challenges encountered in fitting these models to the VADT data.

7.1 Descriptive Statistics

The methods of the VADT are described in detail in Chapter 1. Some descriptive statistics are given in Section 5.2 and in Table VIII. There are \( n = 1,791 \) patients available in this dataset and all of them are included in the analysis. The absolute number of events and number of recurrent events per patient of a given event type are important for fitting the model. If relatively few patients have recurrent events of a given type \( k \), the estimate for \( \sigma_k \rightarrow 0 \) and the model fails to converge in a reasonable number of iterations. In order to more directly compare the results of the fitted model with the results of the previous analysis of composite outcome, I only modeled primary events which are more severe than secondary events. Although the main paper (4) lists seven different types of events, the dataset includes nine. In the original analysis there is no grouping of the nine events. However, the primary paper refers to only seven events,
combining invasive coronary revascularization, invasive peripheral-revascularization (IPR) and invasive cerebrovascularization (ICR) referring to them all as ”surgery for vascular disease”. They also refer to inoperable CAD as a separate event type. I chose to combine only ICR and inoperable CAD because there were only 20 inoperable CAD events in total and they were grouped with ICR in the protocol. The number of events for each primary and secondary event type by treatment group are provided in Table IX below. Events per patient (EPP) is the average number of events per patient, among patients who experienced an event. Proportion of patients who had an event (PPE) is the proportion of patients who had one or more events.
### TABLE IX

NUMBER OF EVENTS BY TREATMENT GROUP

<table>
<thead>
<tr>
<th>Type of Event</th>
<th>Standard</th>
<th>Intensive</th>
<th>Total</th>
<th>EPP</th>
<th>PPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(MI) myocardial infarction</td>
<td>81</td>
<td>76</td>
<td>157</td>
<td>1.11</td>
<td>0.08</td>
</tr>
<tr>
<td>(CHF) congestive heart failure</td>
<td>128</td>
<td>122</td>
<td>250</td>
<td>1.61</td>
<td>0.17</td>
</tr>
<tr>
<td>(IR) Invasive revascularization</td>
<td>174</td>
<td>156</td>
<td>330</td>
<td>1.33</td>
<td>0.08</td>
</tr>
<tr>
<td>(Ang) Angina*</td>
<td>224</td>
<td>233</td>
<td>457</td>
<td>1.52</td>
<td>0.14</td>
</tr>
<tr>
<td>Combined Cardiology</td>
<td>607</td>
<td>587</td>
<td>1194</td>
<td>2.44</td>
<td>0.27</td>
</tr>
<tr>
<td>(Amp) Amputation</td>
<td>20</td>
<td>16</td>
<td>36</td>
<td>1.29</td>
<td>0.02</td>
</tr>
<tr>
<td>(IPR) Invasive peripheral revascularization</td>
<td>37</td>
<td>30</td>
<td>67</td>
<td>1.56</td>
<td>0.02</td>
</tr>
<tr>
<td>(IC) Intermittent claudication*</td>
<td>19</td>
<td>13</td>
<td>31</td>
<td>1.07</td>
<td>0.02</td>
</tr>
<tr>
<td>(CLI) Critical limb ischemia*</td>
<td>21</td>
<td>15</td>
<td>36</td>
<td>1.09</td>
<td>0.02</td>
</tr>
<tr>
<td>Combined Peripheral Vascular</td>
<td>67</td>
<td>63</td>
<td>130</td>
<td>1.24</td>
<td>0.06</td>
</tr>
<tr>
<td>(Str) Stroke</td>
<td>39</td>
<td>28</td>
<td>67</td>
<td>1.06</td>
<td>0.04</td>
</tr>
<tr>
<td>(ICR) Invasive cerebrovascularization</td>
<td>15</td>
<td>16</td>
<td>31</td>
<td>1.07</td>
<td>0.02</td>
</tr>
<tr>
<td>(TIA) Transient ischemic attack*</td>
<td>13</td>
<td>19</td>
<td>32</td>
<td>1.00</td>
<td>0.02</td>
</tr>
<tr>
<td>Combined Cerebrovascular</td>
<td>97</td>
<td>74</td>
<td>171</td>
<td>2.19</td>
<td>0.04</td>
</tr>
<tr>
<td>Combined Recurrent</td>
<td>771</td>
<td>724</td>
<td>1495</td>
<td>2.59</td>
<td>0.32</td>
</tr>
<tr>
<td>(CV-death) CV death</td>
<td>29</td>
<td>38</td>
<td>67</td>
<td>1.00</td>
<td>0.04</td>
</tr>
<tr>
<td>Total All Types</td>
<td>800</td>
<td>762</td>
<td>1562</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

#### 7.2 Model Assumptions and Notation

The units of time used in the model are days. There are no ties within a particular type of event (i.e., MI), therefore handling of ties is not an issue. The confidence intervals for regression parameters are based on the normal approximation: $\hat{\theta} \pm 1.96se(\theta)$. The confidence intervals for $\sigma_1$, $\sigma_2$ and $\sigma_a$ are based on the Satterthwaite approximation $(v\hat{\sigma}^2/\chi^2_{v,0.975}, v\hat{\sigma}^2/\chi^2_{v,0.025})$, where $v = 2(\hat{\sigma}^2/se(\hat{\sigma}^2)^2$ and $\chi^2_{v,\alpha}$ is the $\alpha$-quantile of the $\chi^2$ distribution with $v$ degrees of freedom.
$P$-values for all parameter are based on the Wald test with test statistic $\hat{\theta}^2/\text{var}(\hat{\theta}) \sim \chi^2_1$. Consequently, confidence intervals and $p$-values for $\sigma_1$, $\sigma_2$ and $\sigma_a$ are discordant.

I used sparse-grid quadrature with 15 nodes for the following models. For starting values I set all $\alpha_{jk}$ and $\beta_j$ equal to the equivalent parameter estimate in an independent (without random effects) event-type-specific Cox PH model fitted using SAS PROC PHREG. Starting values for remaining parameters were all $\phi_k=.50$, $\sigma_k=.50$, and $\sigma_a=.75$, the same starting values used in the simulation. Starting values for all $b_{ki}=0$ and $a_i=0$. Numerical differentiation (with $d=\frac{1}{1011}$ only) was used to estimate the standard errors. Convergence criteria was $1E-5$ defined as the square-root of the sum of squares of the gradient vector $\frac{\partial}{\partial \hat{\theta}}$ at each iteration. For the numerical differentiation variance calculation the limit was 200 iterations for each row of the estimated information matrix $\mathcal{I}$ with convergence criteria $c=1E-13$ at iteration number $(k+1)$ defined as the sum of the change in the estimated hazard parameters $\hat{\theta}_3$:

$$\text{crit} = \sum_{\hat{\theta}_3} ||\hat{\theta}_3^{(k+1)} - \hat{\theta}_3^{(k)}||$$

7.2.1 Full Model

In order to fit a model with an event-type-specific random effect $b_k$ it is important that there are enough observed events to estimate $\sigma_k$. In the model presented here I fit random effects $b_1$ and $b_2$ for CHF and stroke respectively. Other event-types were identified that could possibly support random effects are amputation (Amp) and IPR. However, computational issues in this particular dataset prevented me from modeling all four of these recurrent event-types
with their own random effects simultaneously. Other combinations including CHF/Amp and CHF/IPR were attempted but found to be non-convergent.

During the planning of the trial two baseline variables were anticipated to be important predictors of risk and therefore randomized treatment assignment was stratified on both. The first of these covariates is an indicator variable for a patient having experienced a macrovascular event at any point prior to randomization (PE:1=Yes, 0=No). According to the protocol, qualifying events include: MI, angina, CHF, amputation for ischemic gangrene, invasive vascular intervention, TIA, IC or stroke. The second covariate is another indicator variable for whether or not a patient was using insulin at time of randomization (Insulin:1=Yes, 0=No). I included both covariates in addition to treatment in this model. The results are shown in Table X.
## TABLE X

**FULL MODEL**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>Standard error</th>
<th>95% Wald confidence interval</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>*α11 CHF: PE</td>
<td>2.197</td>
<td>0.244</td>
<td>(1.718,2.675)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>*α21 CHF: Insulin</td>
<td>0.506</td>
<td>0.219</td>
<td>(0.076,0.935)</td>
<td>0.021</td>
</tr>
<tr>
<td>α31 CHF: Int</td>
<td>-0.082</td>
<td>0.211</td>
<td>(-0.495,0.331)</td>
<td>0.697</td>
</tr>
<tr>
<td>*α12 Stroke: PE</td>
<td>1.413</td>
<td>0.315</td>
<td>(0.797,2.030)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>α22 Stroke: Insulin</td>
<td>-0.011</td>
<td>0.296</td>
<td>(-0.592,0.570)</td>
<td>0.970</td>
</tr>
<tr>
<td>α32 Stroke: Int</td>
<td>-0.283</td>
<td>0.294</td>
<td>(-0.589,0.293)</td>
<td>0.335</td>
</tr>
<tr>
<td>*α13 Amp: PE</td>
<td>1.208</td>
<td>0.367</td>
<td>(0.489,1.927)</td>
<td>0.001</td>
</tr>
<tr>
<td>α23 Amp: Insulin</td>
<td>-0.165</td>
<td>0.348</td>
<td>(-0.848,0.518)</td>
<td>0.636</td>
</tr>
<tr>
<td>α33 Amp: Int</td>
<td>-0.231</td>
<td>0.349</td>
<td>(-0.915,0.452)</td>
<td>0.507</td>
</tr>
<tr>
<td>*α14 IPR: PE</td>
<td>2.110</td>
<td>0.358</td>
<td>(1.408,2.812)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>α24 IPR: Insulin</td>
<td>0.081</td>
<td>0.270</td>
<td>(-0.448,0.609)</td>
<td>0.765</td>
</tr>
<tr>
<td>α34 IPR: Int</td>
<td>-0.216</td>
<td>0.267</td>
<td>(-0.739,0.306)</td>
<td>0.417</td>
</tr>
<tr>
<td>*α15 IR: PE</td>
<td>1.265</td>
<td>0.164</td>
<td>(0.944,1.585)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>*α25 IR: Insulin</td>
<td>0.461</td>
<td>0.158</td>
<td>(0.152,0.770)</td>
<td>0.003</td>
</tr>
<tr>
<td>α35 IR: Int</td>
<td>-0.092</td>
<td>0.147</td>
<td>(-0.381,0.197)</td>
<td>0.531</td>
</tr>
<tr>
<td>*α16 MI: PE</td>
<td>0.779</td>
<td>0.189</td>
<td>(0.408,1.150)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>α26 MI: Insulin</td>
<td>0.357</td>
<td>0.193</td>
<td>(-0.020,0.735)</td>
<td>0.064</td>
</tr>
<tr>
<td>α36 MI: Int</td>
<td>-0.056</td>
<td>0.186</td>
<td>(-0.420,0.308)</td>
<td>0.762</td>
</tr>
<tr>
<td>*α17 ICR: PE</td>
<td>2.903</td>
<td>0.660</td>
<td>(1.609,4.196)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>α27 ICR: Insulin</td>
<td>-0.354</td>
<td>0.374</td>
<td>(-1.087,0.380)</td>
<td>0.345</td>
</tr>
<tr>
<td>α37 ICR: Int</td>
<td>0.035</td>
<td>0.379</td>
<td>(-0.707,0.778)</td>
<td>0.926</td>
</tr>
<tr>
<td>*β1 CV-death: PE</td>
<td>1.870</td>
<td>0.344</td>
<td>(1.195,2.545)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>β2 CV-death: Insulin</td>
<td>0.185</td>
<td>0.281</td>
<td>(-0.365,0.736)</td>
<td>0.510</td>
</tr>
<tr>
<td>β3 CV-death: Int</td>
<td>0.303</td>
<td>0.274</td>
<td>(-0.233,0.840)</td>
<td>0.268</td>
</tr>
<tr>
<td>*ϕ1 CV-death: CHF</td>
<td>0.381</td>
<td>0.166</td>
<td>(0.056,0.706)</td>
<td>0.022</td>
</tr>
<tr>
<td>*ϕ2 CV-death: Stroke</td>
<td>0.583</td>
<td>0.280</td>
<td>(0.034,1.133)</td>
<td>0.037</td>
</tr>
<tr>
<td>*σ1 CHF</td>
<td>2.295</td>
<td>0.465</td>
<td>(1.600,3.508)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>*σ2 Stroke</td>
<td>1.730</td>
<td>0.838</td>
<td>(0.805,6.021)</td>
<td>0.039</td>
</tr>
<tr>
<td>*σ3 patient</td>
<td>1.659</td>
<td>0.139</td>
<td>(1.418,1.968)</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

-Log-likelihood=12,699.62, *:indicates p-value < 0.05
7.2.2 Reduced Model

Removing the treatment covariate allows me to test the overall treatment effect with the likelihood ratio test. Results of the reduced model are shown in Table XI. In a likelihood ratio test of $H_0: \alpha_{31} = \alpha_{32} = \alpha_{33} = \alpha_{34} = \alpha_{35} = \alpha_{36} = \alpha_{37} = \beta_3 = 0$ the test statistic is 4.36 $\sim \chi^2(df = 8)$, $p$-value = 0.823.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>Standard error</th>
<th>95% Wald confidence interval</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_{11} ) CHF: PE</td>
<td>2.198</td>
<td>0.244</td>
<td>(1.720,2.676)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>( \alpha_{21} ) CHF: Insulin</td>
<td>0.504</td>
<td>0.219</td>
<td>(0.075,0.934)</td>
<td>0.021</td>
</tr>
<tr>
<td>( \alpha_{12} ) Stroke: PE</td>
<td>1.419</td>
<td>0.315</td>
<td>(0.803,2.036)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>( \alpha_{22} ) Stroke: Insulin</td>
<td>-0.007</td>
<td>0.296</td>
<td>(-0.588,0.574)</td>
<td>0.980</td>
</tr>
<tr>
<td>( \alpha_{13} ) Amp: PE</td>
<td>1.210</td>
<td>0.366</td>
<td>(0.493,1.928)</td>
<td>0.001</td>
</tr>
<tr>
<td>( \alpha_{23} ) Amp: Insulin</td>
<td>-0.157</td>
<td>0.348</td>
<td>(-0.839,0.524)</td>
<td>0.651</td>
</tr>
<tr>
<td>( \alpha_{14} ) IPR: PE</td>
<td>2.111</td>
<td>0.356</td>
<td>(1.412,2.809)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>( \alpha_{24} ) IPR: Insulin</td>
<td>0.089</td>
<td>0.269</td>
<td>(-0.438,0.616)</td>
<td>0.740</td>
</tr>
<tr>
<td>( \alpha_{15} ) IR: PE</td>
<td>1.266</td>
<td>0.161</td>
<td>(0.950,1.582)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>( \alpha_{25} ) IR: Insulin</td>
<td>0.462</td>
<td>0.156</td>
<td>(0.156,0.768)</td>
<td>0.003</td>
</tr>
<tr>
<td>( \alpha_{16} ) MI: PE</td>
<td>0.780</td>
<td>0.188</td>
<td>(0.412,1.148)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>( \alpha_{26} ) MI: Insulin</td>
<td>0.356</td>
<td>0.192</td>
<td>(-0.019,0.732)</td>
<td>0.063</td>
</tr>
<tr>
<td>( \alpha_{17} ) ICR: PE</td>
<td>2.905</td>
<td>0.655</td>
<td>(1.620,4.189)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>( \alpha_{27} ) ICR: Insulin</td>
<td>-0.361</td>
<td>0.373</td>
<td>(-1.092,0.371)</td>
<td>0.334</td>
</tr>
<tr>
<td>( \beta_1 ) CV-death: PE</td>
<td>1.861</td>
<td>0.340</td>
<td>(1.195,2.527)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>( \beta_2 ) CV-death: Insulin</td>
<td>0.186</td>
<td>0.280</td>
<td>(-0.362,0.734)</td>
<td>0.506</td>
</tr>
<tr>
<td>( \phi_1 ) CV-death: CHF</td>
<td>0.391</td>
<td>0.162</td>
<td>(0.073,0.710)</td>
<td>0.016</td>
</tr>
<tr>
<td>( \phi_2 ) CV-death: Stroke</td>
<td>0.565</td>
<td>0.267</td>
<td>(0.041,1.088)</td>
<td>0.035</td>
</tr>
<tr>
<td>( \sigma_1 ) CHF</td>
<td>2.299</td>
<td>0.465</td>
<td>(1.603,3.572)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>( \sigma_2 ) Stroke</td>
<td>1.746</td>
<td>0.841</td>
<td>(0.815,6.020)</td>
<td>0.039</td>
</tr>
<tr>
<td>( \sigma_\text{a} ) patient</td>
<td>1.660</td>
<td>0.129</td>
<td>(1.434,1.946)</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

-Log-likelihood=12,701.80, *: indicates p-value < 0.05
7.3 Supplementary Four-Dimensional Analysis

Fitting the primary event data in more than three dimensions was difficult due to the sparseness of events. Therefore, in order to demonstrate the model at greater than three dimensions I combined primary events into three groups: cardiology, peripheral vascular, and cerebrovascular as per Table IX. I treated events in these groups as the same type of recurrent event and where there were ties I assigned a time of one day for events beyond the first that occurred on the same day. The results are shown in Table XII. There is no indication of treatment effect on any of the three event types, while PE is significant throughout. One should take caution while interpreting these results due to the grouping of events. It is included here for demonstration purposes. Another four-dimensional model similar to this one with the addition of secondary outcomes was also fitted with very similar results.
TABLE XII
PRIMARY OUTCOMES GROUPED MODEL

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>Standard error</th>
<th>95% Wald confidence interval</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_{11} ) Cardio. - Int</td>
<td>-0.094</td>
<td>0.111</td>
<td>(-0.311,0.122)</td>
<td>0.393</td>
</tr>
<tr>
<td>( \ast \alpha_{21} ) Cardio. - PE</td>
<td>1.267</td>
<td>0.119</td>
<td>(1.034,1.499)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>( \ast \alpha_{31} ) Cardio. - Insulin</td>
<td>0.429</td>
<td>0.115</td>
<td>(0.205,0.654)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>( \alpha_{12} ) Periph. - Int</td>
<td>-0.165</td>
<td>0.226</td>
<td>(-0.608,0.277)</td>
<td>0.464</td>
</tr>
<tr>
<td>( \ast \alpha_{22} ) Periph. - PE</td>
<td>1.614</td>
<td>0.256</td>
<td>(1.113,2.116)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>( \alpha_{32} ) Periph. - Insulin</td>
<td>-0.168</td>
<td>0.228</td>
<td>(-0.615,0.280)</td>
<td>0.462</td>
</tr>
<tr>
<td>( \alpha_{13} ) Cereb. - Int</td>
<td>-0.441</td>
<td>0.397</td>
<td>(-1.220,0.338)</td>
<td>0.267</td>
</tr>
<tr>
<td>( \ast \alpha_{23} ) Cereb. - PE</td>
<td>1.668</td>
<td>0.433</td>
<td>(0.820,2.516)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>( \alpha_{33} ) Cereb. - Insulin</td>
<td>-0.276</td>
<td>0.402</td>
<td>(-1.063,0.511)</td>
<td>0.491</td>
</tr>
<tr>
<td>( \beta_{1} ) CV Death - Int</td>
<td>0.319</td>
<td>0.290</td>
<td>(-0.249,0.887)</td>
<td>0.271</td>
</tr>
<tr>
<td>( \ast \beta_{2} ) CV Death - PE</td>
<td>1.897</td>
<td>0.384</td>
<td>(1.144,2.650)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>( \beta_{3} ) CV Death - Insulin</td>
<td>0.170</td>
<td>0.293</td>
<td>(-0.405,0.745)</td>
<td>0.563</td>
</tr>
<tr>
<td>( \phi_{1} ) Death((b_{1} \text{ Cardio.}))</td>
<td>0.536</td>
<td>0.342</td>
<td>(-0.135,1.207)</td>
<td>0.117</td>
</tr>
<tr>
<td>( \ast \phi_{2} ) Death((b_{2} \text{ Periph.}))</td>
<td>1.075</td>
<td>0.534</td>
<td>(0.027,2.122)</td>
<td>0.044</td>
</tr>
<tr>
<td>( \phi_{2} ) Death((b_{3} \text{ Cereb.}))</td>
<td>-0.183</td>
<td>0.201</td>
<td>(-0.577,0.210)</td>
<td>0.361</td>
</tr>
<tr>
<td>( \ast \sigma_{1} ) Cardio.</td>
<td>0.670</td>
<td>0.242</td>
<td>(0.368,1.583)</td>
<td>0.006</td>
</tr>
<tr>
<td>( \ast \sigma_{2} ) Periph.</td>
<td>0.839</td>
<td>0.483</td>
<td>(0.349,4.041)</td>
<td>0.082</td>
</tr>
<tr>
<td>( \ast \sigma_{3} ) Cereb.</td>
<td>5.468</td>
<td>1.390</td>
<td>(3.514,9.668)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>( \ast \sigma_{a} ) patient</td>
<td>0.830</td>
<td>0.209</td>
<td>(0.535,1.460)</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

-Log-likelihood=11,219.38, *:indicates p-value < 0.05

7.4 Discussion of Analysis

First, treatment effect is nonsignificant for all event types in the full model in Table X. The positive sign of the treatment effects, while nonsignificant, indicate a reduction in risk of all events but ICR and death for the intensive group versus the standard group, which is consistent with the event counts in Table IX. In both full and reduced models the prior event
at baseline variable is significant for all eight event types. Having received insulin at baseline is associated with an increased risk of CHF and IR in both models. In both full and reduced models, increased random effect for CHF and stroke are associated with increased risk of CV-death after adjusting for the patient-specific frailty $a_i$, with hazard ratios in the full model of $1.46 \ (p=.016)$ and $1.79 \ (p=.035)$ per unit standard deviation increase of the event-specific random effect respectively.

The computation time for these models are significant. Number of events, number of quadrature nodes, and number of dimensions are all drivers of computation time. The full model took 7.2 hours to converge in 2,163 iteration on an Intel(R) CPU: i7 2600K @3.4Ghz and 16 GB of RAM. The MEMLIB option in SAS was enabled, which allows storage of temporary datasets in memory rather than in a working directory on the hard drive. Multiple cores were not utilized within the algorithm. Use of multiple cores within the algorithm would require rewriting and redesigning the program in another environment that supported multiple core processing such as R. Multiple core processing may reduce computation time of a single model fit significantly on a PC with adequate memory. I chose SAS because I had access to a SAS Grid distributed computing system that allowed me to run multiple simulations in parallel. Distributing the simulations amongst multiple machines allowed for much shorter run time for the entire set as opposed to running simulations serially on a single machine optimized for multiple cores.

The number of nodes is important for obtaining successful convergence in this dataset. With three dimensions, fewer than 11 nodes could lead to divergence (often caused by negative
weights combined with influential observations, a quirk in the sparse-grid quadrature procedure). Too many nodes (>15) may lead to numerical overflow due to a double-exponential function in the likelihood combined with large values at the extreme ends of the node vector in the quadrature routine and patients with many multiple events. This problem becomes worse when the dimensions increase as the likelihoods for multiple dimensions are multiplied together leading to even larger numbers. Increasing the dimension generally requires increase in the number of nodes for successful convergence. Therefore, a delicate balance must be struck between increased dimensionality and numerical accuracy. In addition, increased dimensions lead to increased memory required to store the matrix used in quadrature. Memory-saving measures could be taken in the program (such as using for-loops instead of vectorized calculations) at the cost of increased computation time. With this dataset, I was not able to find a suitable compromise to model four-dimensions (adding Amp or IPR), although four-dimensions was not a problem when grouping the event-types into four distinct groups. The lesson here is that a minimum number of recurrent events for a given event type is required to fit the model with event-specific random effects, making the model unsuitable for very sparse data. Although the capability to fit both event-types with and without event-specific random effects increases its usefulness.

Recall the results of the original complication-free survival time model: for the intensive versus standard group relative reduction of 11.9% (hazard ratio, 0.88; 95% confidence interval, 0.74 to 1.05; p=.14) was realized, unadjusted for prior events or baseline insulin. While complication-free survival time is clinically important it is by definition short-sighted. When
looking beyond the first event with our model we do not see anything approaching significance of treatment effect for any of the four types of events. In the above analysis the overall test for treatment effect was $p=.823$. It may be intensive treatment does increase complication-free time, but that the first event is merely postponed and subsequent events are inevitable regardless of treatment group. This lack of evidence of long-term effect should be considered when weighing the risks associated with intensive glycemic control such as an increased cost and risk of hypoglycemic episodes.
CHAPTER 8

THREE-LEVEL HIERARCHICAL PROPORTIONAL-HAZARDS MODEL: SIMULATIONS

8.1 Data Generation

There are four scenarios, each having 500 simulations consisting of \( n = 500 \). For the ECM algorithm which estimates \( \hat{\theta} \) convergence limit of 10,000 iterations was never exceeded. The convergence criteria for the ECM algorithm and numerical differentiation as defined in Section 7.2 with criteria \( 1E - 5 \) and \( 1E - 13 \) respectively. In the numerical differentiation, each row of the information matrix was calculated twice, once for \( d = \frac{1}{500} \) and once again for \( d = -\frac{1}{500} \), taking the average of the two resulting rows.

I used a SAS Grid distributed computing system made available by the VA Informatics and Computing Infrastructure (VINCI). The system consists of six Windows Server 2008 machines each with eight Intel Xeon CPU X5677 processors @ 3.47GHz. Each simulation took an average of eight hours when all available resources are available and dedicated and SAS MEM-LIB option is on. Simulations were split into batches of 20–50, depending on server load, and submitted in parallel. Therefore the sample size and number of simulations was kept relatively small as it took several days to complete all 500 simulations for a given scenario.
I simulated four scenarios labeled A1, A2, B1, and B2. Scenarios A1 and A2 include a single covariate \( Z_1 \) which is distributed \( Z_1 \sim \text{Bernoulli}(0.5) \). Scenarios B1 and B2 include both \( Z_1 \sim \text{Bernoulli}(0.5) \) and \( Z_2 \sim \text{Uniform}(-0.5, 0.5) \).

Event times of type \( k \forall k = 1, ..., 4 \) were generated from the Weibull distribution \( f(t) \)

\[
f(t) = \frac{a}{b} t^{a-1} e^{-\frac{t^a}{b}}
\]

where \( a = 2 \) and \( b = e^{-\left(\alpha_k Z_i + b_k + a_i\right)/\xi_k} \) for recurrent events or \( b = e^{-\left(\beta T Z_i + \sum_{k=1}^{K} \phi_k (b_k + a_i)/\xi_3\right)} \) for terminal events. The parameters were chosen to produce a dataset that resembles the features of VADT data in which some types of recurrent events are more frequent than others and death is relatively rare. The chosen parameters for each scenario are shown in Table XIV.

Means and standard deviations of number of events for each type of event in each scenario are shown in Table XIII. We also simulated censoring at study end \( \tau = 1 \) and independent 10% independent censoring where censoring time \( C_i \sim \text{Uniform}(0, \tau) \).

<table>
<thead>
<tr>
<th>Event</th>
<th>Scenario A1</th>
<th>Scenario A2</th>
<th>Scenario B1</th>
<th>Scenario B2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>219.3 (17.3)</td>
<td>244.8 (10.2)</td>
<td>218.9 (15.7)</td>
<td>242.8 (17.5)</td>
</tr>
<tr>
<td>Type 2</td>
<td>652.6 (31.2)</td>
<td>515.8 (19.0)</td>
<td>652.6 (32.3)</td>
<td>514.3 (23.9)</td>
</tr>
<tr>
<td>Terminal</td>
<td>125.4 (9.8)</td>
<td>160.8 (10.2)</td>
<td>127.7 (9.6)</td>
<td>163.4 (9.9)</td>
</tr>
</tbody>
</table>
### TABLE XIV

SIMULATED PARAMETER VALUES BASED ON VADT

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Scenario A1</th>
<th>Scenario A2</th>
<th>Scenario B1</th>
<th>Scenario B2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ξ₁</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>α₁₁</td>
<td>-0.50</td>
<td>0</td>
<td>-0.50</td>
<td>0</td>
</tr>
<tr>
<td>α₂₁</td>
<td>–</td>
<td>–</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>ξ₂</td>
<td>1.30</td>
<td>1.30</td>
<td>1.30</td>
<td>1.30</td>
</tr>
<tr>
<td>α₁₂</td>
<td>0.50</td>
<td>0</td>
<td>0.50</td>
<td>0</td>
</tr>
<tr>
<td>α₂₂</td>
<td>–</td>
<td>–</td>
<td>-0.25</td>
<td>-.025</td>
</tr>
<tr>
<td>ξ₃</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>β₁</td>
<td>-0.75</td>
<td>0</td>
<td>-0.75</td>
<td>0</td>
</tr>
<tr>
<td>β₂</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>φ₁</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>φ₂</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>σ₁</td>
<td>1.50</td>
<td>1.50</td>
<td>1.50</td>
<td>1.50</td>
</tr>
<tr>
<td>σ₂</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>σₐ</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
</tbody>
</table>

---: The parameter was not included in the data generation mechanism

#### 8.2 Simulation Results

Results are shown in Table XV and Table XVI in which the mean difference between the estimate and the value is labeled Bias, the standard error of the estimate labeled SE, the mean of the estimated standard error labeled SEE, and coverage probability labeled CP. A confidence interval of a single parameter $\hat{\theta}_s$ covers $\theta_s$ if $\hat{\theta}_s - 1.96se(\hat{\theta}_s) < \theta_s < \hat{\theta}_s + 1.96se(\hat{\theta}_s)$. The confidence interval for $\sigma_1$, $\sigma_2$, and $\sigma_\alpha$ are based on the Satterthwaite approximation as given in Section 7.1.
### TABLE XV

**SIMULATION RESULTS SCENARIOS A1 AND A2**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Scenario A1</th>
<th></th>
<th>Scenario A2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{11}$</td>
<td>0.017 0.220 0.204 0.936</td>
<td></td>
<td>0.007 0.191 0.193 0.950</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.014 0.151 0.142 0.934</td>
<td></td>
<td>-0.006 0.149 0.147 0.950</td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.015 0.198 0.209 0.962</td>
<td></td>
<td>-0.007 0.177 0.181 0.954</td>
<td></td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.012 0.139 0.136 0.970</td>
<td></td>
<td>0.027 0.121 0.112 0.970</td>
<td></td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.049 0.140 0.135 0.942</td>
<td></td>
<td>0.039 0.131 0.126 0.950</td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>-0.083 0.376 0.422 0.986</td>
<td></td>
<td>-0.124 0.354 0.396 0.982</td>
<td></td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>-0.077 0.247 0.246 0.974</td>
<td></td>
<td>-0.057 0.248 0.256 0.976</td>
<td></td>
</tr>
<tr>
<td>$\sigma_a$</td>
<td>-0.066 0.155 0.160 0.982</td>
<td></td>
<td>-0.035 0.157 0.160 0.970</td>
<td></td>
</tr>
</tbody>
</table>

### TABLE XVI

**SIMULATION RESULTS SCENARIOS B1 AND B2**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Scenario B1</th>
<th></th>
<th>Scenario B2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{11}$</td>
<td>0.029 0.208 0.207 0.930</td>
<td></td>
<td>0.000 0.199 0.196 0.956</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>-0.030 0.360 0.349 0.942</td>
<td></td>
<td>-0.008 0.341 0.341 0.940</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.019 0.143 0.143 0.948</td>
<td></td>
<td>0.009 0.149 0.147 0.956</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>-0.030 0.257 0.245 0.932</td>
<td></td>
<td>-0.012 0.270 0.257 0.936</td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.013 0.206 0.207 0.956</td>
<td></td>
<td>0.003 0.185 0.179 0.950</td>
<td></td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.006 0.358 0.350 0.938</td>
<td></td>
<td>0.012 0.314 0.317 0.956</td>
<td></td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.018 0.133 0.138 0.982</td>
<td></td>
<td>0.022 0.122 0.119 0.972</td>
<td></td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.032 0.136 0.133 0.962</td>
<td></td>
<td>0.038 0.138 0.128 0.946</td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>-0.108 0.411 0.418 0.974</td>
<td></td>
<td>-0.098 0.389 0.402 0.980</td>
<td></td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>-0.058 0.262 0.248 0.952</td>
<td></td>
<td>-0.064 0.265 0.257 0.966</td>
<td></td>
</tr>
<tr>
<td>$\sigma_a$</td>
<td>-0.069 0.160 0.158 0.988</td>
<td></td>
<td>-0.042 0.157 0.160 0.964</td>
<td></td>
</tr>
</tbody>
</table>
8.3 Discussion of Simulation Results

These results indicate that the estimation procedure outlined above produces unbiased estimates for regression parameters $\alpha$, $\beta$, and $\phi$ with adequate coverage probability. However, for $\sigma_1$, $\sigma_2$, and $\sigma_a$ there is a persistent negative bias. Negative bias of variance parameter estimates in generalized linear mixed models using Gaussian quadrature is well documented (10),(11). The coverage probability for these parameters are also slightly larger than 95%. The biases are robust to increases in sample size and convergence criteria as shown in Table XVII; the results from 100 simulations Scenario A1 with $n$ increased to 2,000 and stricter convergence criteria decreased to 1E-6. A further modification to Scenario C1 is made, setting $\sigma_1 = \sigma_2 = \sigma_a = 0.50$, as Scenario C2. In Scenario C2, the bias of the estimates for $\sigma_1$, $\sigma_2$, and $\sigma_a$ seems to have improved. This seems to indicate that the algorithm has difficulty estimating the variance of the random effects when they are relatively large, perhaps related to large realized random effects (outliers) that in turn generate a large number of recurrent events relative to the rest of the sample.

When the number of recurrent events is large, the integrand of the type found in the formula for $E[u_i^2]$, which is bimodal, is characterized by two modes relatively far apart. As such, the quadrature nodes may provide poor coverage of the integrand, thereby underestimating the integral. Shifting the nodes by $\hat{a}_i$ may make the problem worse. I tried different kinds of transformations such as shifting by the mode of the integrand $\hat{\mu}$ and scaling by $-\frac{\partial^2}{\partial a^2} \log g(a)|_{a=\hat{\mu}}$, but there was a significant increase in computation time and a lack of stability often leading to non-convergence. Mixing of quadrature routines combined with different shifting/scaling
TABLE XVII

SIMULATION RESULTS MODELS C1 AND C2

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Scenario C1</th>
<th></th>
<th></th>
<th></th>
<th>Scenario C2</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>SE</td>
<td>SEE</td>
<td>CP</td>
<td>Bias</td>
<td>SE</td>
<td>SEE</td>
<td>CP</td>
</tr>
<tr>
<td>α_{11}</td>
<td>-0.005</td>
<td>0.096</td>
<td>0.102</td>
<td>0.960</td>
<td>-0.001</td>
<td>0.103</td>
<td>0.094</td>
<td>0.956</td>
</tr>
<tr>
<td>α_{12}</td>
<td>0.000</td>
<td>0.072</td>
<td>0.071</td>
<td>0.960</td>
<td>-0.004</td>
<td>0.065</td>
<td>0.064</td>
<td>0.950</td>
</tr>
<tr>
<td>β_{1}</td>
<td>-0.002</td>
<td>0.098</td>
<td>0.102</td>
<td>0.960</td>
<td>0.008</td>
<td>0.099</td>
<td>0.102</td>
<td>0.954</td>
</tr>
<tr>
<td>φ_{1}</td>
<td>0.016</td>
<td>0.059</td>
<td>0.063</td>
<td>0.970</td>
<td>0.014</td>
<td>0.139</td>
<td>0.143</td>
<td>0.970</td>
</tr>
<tr>
<td>φ_{2}</td>
<td>0.026</td>
<td>0.057</td>
<td>0.064</td>
<td>0.970</td>
<td>-0.006</td>
<td>0.110</td>
<td>0.111</td>
<td>0.950</td>
</tr>
<tr>
<td>σ_{1}</td>
<td>-0.060</td>
<td>0.196</td>
<td>0.210</td>
<td>0.990</td>
<td>0.035</td>
<td>0.193</td>
<td>0.177</td>
<td>0.900</td>
</tr>
<tr>
<td>σ_{2}</td>
<td>-0.085</td>
<td>0.141</td>
<td>0.123</td>
<td>0.860</td>
<td>-0.012</td>
<td>0.092</td>
<td>0.093</td>
<td>0.950</td>
</tr>
<tr>
<td>σ_{a}</td>
<td>-0.051</td>
<td>0.090</td>
<td>0.083</td>
<td>0.990</td>
<td>0.002</td>
<td>0.072</td>
<td>0.067</td>
<td>0.960</td>
</tr>
</tbody>
</table>

formulas also resulted in non-convergence (most likely due to the presence of bimodal functions). More research is required to find an ideal method or combination of methods that accurately approximates all of the different functional forms of integrals necessary for this application.

The number of dimensions and nodes are the driving force behind the computation time for this model. Twelve nodes was chosen because it allowed all simulations to reach convergence without any errors. I found that with 10 nodes there were some errors such as negative estimates of the hazard parameters. This was found to be more likely to occur in simulations containing a patient with a large number of repeated events by chance. This problem appeared to be caused by using an exponential model for data generation, which allowed for a single patient to have multiple events in quick succession. Fitting the VADT data under this model and attempting to reproduce the data using the estimated parameters with the exponential data generation mechanism resulted in some patients having a large number (10–20) of recurrent
events, the occurrence of which is unrealistic and not present in the VADT. This lead me to rethink the data generation mechanism and switch to Weibull-generated recurrent events in a gap-time model. Weibull allows for an increasing hazard function, effectively resetting the hazard to zero immediately after a recurrent event. This made several recurrent events in quick succession unlikely and led to simulated data that resembled the VADT data much more closely. The result was a lower number of maximum events per patient realized with the same variance parameter for a given random effect distribution. This assumption is reasonable for the VADT application since experience of a primary outcome event often resulted in hospitalization and treatment, leading to a decreased risk of subsequent events immediately following the previous event.

Increasing the number of nodes is costly in terms of computation time but increasing the number of dimensions or types of events being modeled is much more costly, even with the savings afforded by using sparse-grid quadrature. Estimation of the VADT data with four dimensions and three covariates took more than one day and increasing the number of dimensions to five would require multiple days. As mentioned earlier, it is anticipated that multi-core processing would help reduce the computation time.
APPENDIX

Memorandum

WARC HINES, JR. VA HOSPITAL/NORTH CHICAGO VAMC

INSTITUTIONAL REVIEW BOARD

DATE: March 17, 2008

TO: Nicholas V. Emanuele, M.D., 111A

CC: Kathleen Bakke

FROM: Tony Sroczynski, PharmD, CPhA

PIB Coordinator (161)

RE: PROMISE II, RRS # 06-061; CSP # 465: Glycemic Control and Complications in Diabetes Mellitus Type 2 (6-2-03 - CLOSED TO ACCRUAL)

Dear Dr. Emanuele:

Your request for continuing review of the study listed above was reviewed at the 3/17/2008 meeting of the Institutional Review Board.

The requested continuation involves no changes to the protocol or consent form. This is to confirm your request for continuation & approval. You must obtain a signed consent prior to conducting any research procedures. You are also reminded of applicable HIPAA requirements.

You are granted permission to continue your study as described effective immediately. The study is next subject to continuing review on or before 3/16/2009, unless closed before that date. It is your responsibility as the principal investigator to submit the completed report at a minimum of 4 weeks prior to the expiration of the study for review and processing purposes. However, the PIP staff will send a courtesy reminder approximately 45 days prior to expiration.

Sincerely,

The Human Research Protection Plan is available on the Hines VA (North Chicago VAMC) website. Please make sure you are familiar with your responsibilities.

VA regulations require annual human subjects protection training. It is your responsibility to ensure training is current for all personnel on this project.

A subject is considered enrolled if he/she has signed a consent and must be reported at the time of continuing review. For database studies, enrollment relates to the number of individual subject records reviewed and must be reported at the time of continuing review.

- The witness is witnessing only the subjects signature and is not to be a member of the research team.
- An Enrollment, follow-up and completion Progress Note is required for all subjects enrolled, except for database studies. Using a progress note titled "Researcher" Progress Note will enable easy identification of such a note. No progress notes are required for database studies.
- A Research Flag is entered for any subject enrolled in a study that has more than 1 visit. Please contact Larry Brand for the Research Flag menu.
- When a subject has completed the study, please use the research flag menu to discontinue him/her in order to remove the Flag Alert from the file. All historical information remains available.
- Signed Consent forms must be forwarded to the Pharmacy (if applicable), Medical Records and the Human Studies Office.
- Copies of VA Form 10-0012 must be filed in the medical record if the study involves an investigational drug.
- All Adverse Events (related or unrelated), Unanticipated Problems and Protocol Deviations are to be reported to the IRB. Any serious AE, Unanticipated Problems must be reported within 48 hours of PI identification.
- All Applications, Investigator’s Manuals, Check-lists and Reporting Forms are available electronically: http://www.hnclzmed.va.gov/files/research.

As with the initial approval, changes to the study must be submitted and approved prior to implementation. Contact me at 708-202-5594 or email: Tony.Sroczynski@va.gov if you have any questions or require further information.

Sincerely,

APPENDIX
APPENDIX (continued)

Principal Investigator: Nicholas Emanuele, M.D. #17
Study Title: CFSP#01: “GLYCEMIC CONTROL AND COMPLICATIONS IN DIABETES MELLITUS TYPE 2”

HINES VAH / NC VAAM INSTITUTIONAL REVIEW BOARD (IRB)
REVIEW AND DETERMINATION OF CONTINUING APPROVAL or CLOSURE

The information contained in this Continuing Review Packet and project file were reviewed by a qualified IRB member. Review Comments and recommendations were brought forward and discussed by the IRB and the status determined as described below:

Approved on 7/06/06, for Period from 7/06/06 to 7/06/07
Category of Review: ☑ Full Board ☐ Expedited (If Expedited; Review Category: )

Reported as Closed or Terminated: ☐ Disapproved on

(Site letter for reason)

In Summary, the IRB had determined that the following have been met as required by the regulations and Research Handbook 1040:

1. Minimization of Risks: Risks, both physical and non-physical, to human subjects are minimized by using procedures that are consistent with sound research design; that do not unnecessarily expose subjects to risk and, whenever appropriate, using procedures already being performed on the subjects for diagnostic or treatment purposes.

2. Reasonable Risk/Benefit Ratio: Risks, both physical and non-physical, to human subjects are reasonable in relation to any anticipated benefits (the risk benefit ratio) to subjects, and the importance of the knowledge that may reasonably be expected to result.

3. Equitable Selection of Subjects: The selection of is equitable, the IRB has taken into account the purpose of the research and the research setting.

4. Review and Approval of the Informed Consent Form (Approval of waiver/alteration or waiver of documentation, if applicable)

5. IRB approval of the wording of the consent to be documented through the use of a stamp on each page of the VA Form 10-088 that indicates the date of the most recent IRB approval of the document. If the consent form is amended during the protocol approval period, the form will bear the approved date of the amendment rather than the date of the approved protocol.

6. The IRB has reviewed the required language for a valid authorization to release health information

7. Securing Informed Consent and Documentation of the Informed Consent Process. Informed consent will be sought from each prospective subject or, if approved, the subject's authorized representative. A person knowledgeable about the consenting process and the research to be conducted will obtain the informed consent.

8. Monitoring Safety: The research plan makes adequate provisions for monitoring the data collected to ensure the safety of subjects. The plan includes appropriate procedures for reporting AEs.

9. Privacy and Confidentiality: Adequate provisions will be taken to protect the privacy of subjects and to maintain the confidentiality of individually-identifiable data. Such provisions consider the requirements of Standards for Privacy of Individually Identifiable Health Information (HIPAA Privacy Rule), 45 CFR parts 160 and 164, and other laws protecting the rights of veterans' information, including Privacy Act of 1974, 5 U.S.C. 552a; VA Claims Confidentiality Statute, 38 U.S.C. 1301; Confidentiality of Drug Abuse, Alcoholism and Alcohol Abuse, Substance Use with Human Immunodeficiency Virus (HIV), and Sickle Cell Anemia Medical Records, 38 USC 7317, and Confidentiality of Health Care Information in VA's Health Care Information Systems, 38 USC 7323.

10. Protection of Vulnerable Subjects. Additional safeguards have been included in the study to prevent the welfare of vulnerable subjects.

11. Conflict of Interest: Adequate steps to manage, reduce or eliminate potential or real conflicts of interest (financial, role [investigator/patient relationship], and/or institutional) have been taken. All VA investigators must be aware of and comply with VA policies and procedures regarding conflict of interest.

12. Investigator's Educational Requirements and Certification: The PI and all other investigators of this proposed research activity must meet all current educational requirements for the protection of human research subjects as mandated by the facility's Assurance, VA OER, funding institutions, and applicable OHRP requirements. The IRB has determined that the investigator is qualified through education, training, and experience to conduct the research.

Chair, IRB

Date

Rev. 8-10-07
10
CITED LITERATURE


of integrating longitudinal data and survival analysis.” Statistics in Medicine 31 (19):
1903–1917.

Henderson. 2003. “Design of the cooperative study on glycemic control and
complications in diabetes mellitus type 2 Veterans Affairs Diabetes Trial.” Journal of


censored intermediate events.” Journal of the American Statistical Association 96 (454):
449–457.


VITA

Derrick Gregory Kaufman

Education

• University of Illinois at Chicago, Chicago, IL M.S. in Biostatistics Awarded May 2008

• University of Illinois at Chicago, Chicago, IL B.S. in Bioengineering Awarded May 2005

Work Experience

• Cooperative Studies Program Coordinating Center, Hines VA, Hines, IL Clinical Study Biostatistician Jun 2009 – Present

• Zurich Financial Services, Schaumburg, IL Actuarial Analyst Dec 2007 – June 2009

Publications

