JOINT MODELING OF LONGITUDINAL AND SURVIVAL DATA

by

ALIA ALKHATHAMI

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial
fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

IN

PROBABILITY AND STATISTICS

OTTAWA-CARLETON INSTITUTE FOR MATHEMATICS AND STATISTICS

DEPARTMENT OF SCHOOL OF MATHEMATICS AND STATISTICS

CARLETON UNIVERSITY

OTTAWA, ONTARIO

©Copyright

ALIA ALKHATHAMI, 2021

To the Memories of My Mother

To Myself & My Family

Abstract

An important target of many biomedical and clinical research paradigms is to identify biomarkers, including risk scores, with strong prognostic capabilities. Biomarker evaluations are usually utilized to predict the progression of the disease under study. In such clinical studies, one major research objective is to identify immune response biomarkers measured longitudinally that may be associated with the risk of death, infection, or any other interesting event. Joint modeling of longitudinal and survival data has become incredibly beneficial for analyzing such clinical trial data. In this thesis, we propose and explore innovative joint models and methods for analyzing longitudinal response and time-to-event data, and also address the computational challenges associated with the likelihood inference. Specifically, we consider two approaches to obtain the estimates of the model parameters, the maximum likelihood (ML) method and approximate likelihood method based on the so-called h-likelihood. In both cases, we consider a joint model that assumes a latent process based on random effects to describe the association among longitudinal and survival data. We use the linear mixed-effects model for longitudinal measurements and Weibull frailty model for survival outcomes. The two submodels are linked through shared random effects, where the longitudinal and survival processes are conditionally independent given the random effects. We also investigate the impact of the association parameter on the regression estimators in a separate analysis of each outcome. In the case of strong associations, the separate analysis provides biased parameter estimates with poor coverage probabilities of the confidence interval. In this work, we also investigate the effect of a misspecified random

effects distribution on the parameter estimates in joint models. We assess the robustness properties of the maximum likelihood estimators obtained under misspecified random effects distributions, i.e., when the random effect follows a non-normal distribution, such as the Chi-square or Gamma distribution. Our empirical study shows that the use of the t-distribution for random effects gives more robust estimates of model parameters, as compared to the commonly used normal distribution. As an application, we analyze a large clinical dataset of sepsis patients obtained from a longitudinal study. We performed an extensive analysis of the data, referred to as the Genetic and Inflammatory Markets of Sepsis (GenIMS) data, using joint models and investigate the association between the survival outcome of a sepsis patient and a longitudinal D-dimer biomarker, as well as other demographic variables. The joint analysis appears to provide estimates with smaller standard errors, as compared to the separate analysis of longitudinal and survival outcomes.

Acknowledgements

Throughout the writing of this dissertation I have received a great deal of support and assistance from special people. Therefore, I would like to take some time to thank all the people without whom this work would never have been possible. Those people literally have contributed to the research in their own particular way, and for that they deserve special thanks from me.

To the greatest extend, I sincerely thank the almighty God for His graces, strength, sustenance, and His faithfulness and love from the beginning of my academic life up to this doctoral level. Truly, His benevolence has made me excel and succeed in all my academic pursuits. I am grateful to God for all my accomplishments, especially this work.

Mainly, My unalloyed appreciation also goes to my amiable, ever supportive and humble supervisor at the Department of Mathematics and Statistics at Carleton University Dr. Sanjoy Sinha for his guidance, encouragement, support, and deep consideration throughout my dissertation work. I have been greatly impressed by his inspiration and immense knowledge, which help me move forward. I have learned from him not only in academic research but also in future career development. I extremely appreciate his time and patience with me. He always responded and contributed my work and read the drafts of each page of my work more quickly than I could have hoped. It was a great privilege to work under his guidance, which I hope to continue even long after this thesis is presented.

I also want to take this opportunity to thank the Dean, director, faculty, and staff members of the Department of Mathematics and Statistics for all the opportunities it has afforded

me over the years, including learning and teaching opportunities that helped me grow and gain the knowledge you see applied in this work, as well as the financial support. The assistance I received from our department really allowed me to continue my studies and research, and for this, I am incredibly grateful. Special thanks go to my friends for their support, encouragement, and for all the care and fun I have received. I could not have completed this dissertation without receiving stimulating discussions from them as well as their happy distractions to rest my mind outside of my research.

Finally and most importantly, I would like to thank all my family members for their patience, encouragement, understanding, and unconditional love and support, especially my dear father and siblings. A very special thanks to my two brothers, Abdullah and Mesfer, who shared with me the best moments of laughter and joy while I am completed this thesis. I am totally indebted for giving me constant motivations and taking care of me during my tough times in this journey.

Statement of Originality

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices.

Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I request a permission to copy or to make other use of material in my thesis in whole or parts.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office and that this thesis has not been submitted for a higher degree to any other university or institution.

Alia Alkhathami

May, 2021

List of Contents

Abstract	ii
Acknowledgements	iv
Statement of Originality	vi
List of Tables	x
List of Figures	xi
List of Nomenclatures	xiii
List of Acronyms	xv
1 Introduction	1
1.1 Review of Longitudinal Data	1
1.2 Review of Survival (Time-to-event) Data	4
1.3 Joint Modelling of Longitudinal and Survival Data	9
1.4 Research Objectives and Thesis Outline	11
1.5 Organization of Thesis	13
2 Analysis of Longitudinal Data	14
2.1 Introduction	14
2.2 Notation	15

2.3	General Linear Models	15
2.4	Linear Mixed Effects Model	17
2.4.1	Inference for Linear Mixed Effects Model	19
2.4.2	Covariance Matrix Structure	21
2.4.2.1	Random-intercept model	21
2.4.2.2	Random-Intercept and Random-Slope Model	22
2.5	Generalized Linear Mixed Model	24
2.6	Simulation Study	26
2.6.1	Overview of Evaluation Criteria	26
2.6.2	Generating Longitudinal Data	27
2.6.3	Simulation Results and Discussion	28
2.7	Application on HIV/AIDS Study	29
2.7.1	Background	29
2.7.2	AIDS Dataset Description	30
2.7.3	AIDS Data Analysis	32
3	Analysis of Survival Data	36
3.1	Introduction	36
3.2	Notation	37
3.2.1	Non-parametric Methods	37
3.2.2	Semi-parametric Model	38
3.2.3	Parametric Model	40
3.2.3.1	Likelihood Function for Right-Censored Data	41
3.2.4	The Weibull Proportional Hazards Model	42
3.2.4.1	Likelihood Inference and Computational Method	43
3.2.5	The Accelerated Failure Time (AFT) Model	44
3.2.5.1	The Weibull Accelerated Failure Time Model	46
3.2.6	The Extended Cox Model with Time-Dependent Covariates	48

3.3	The Frailty Model	50
3.3.1	The Weibull Proportional Hazards Model with Shared Frailty	51
3.4	Simulation Study	52
3.4.1	Simulation Results and Discussion	53
3.5	Application on HIV/AIDS Study	55
3.5.1	Cox Proportional Hazard Model	56
3.5.2	Weibull Proportional Hazard Model	57
3.5.3	Weibull Frailty Model	60
4	Joint Likelihood Analysis of Longitudinal and Survival Data	62
4.1	Introduction	62
4.2	Notation	64
4.3	The Basic Joint Model	64
4.3.1	The Survival Submodel	66
4.3.2	The Longitudinal Submodel	71
4.3.3	Joint Model with Shared Random Effects	72
4.4	Review of Estimation Methods for Joint Models	77
4.5	Simulation Study	79
4.5.1	Generating longitudinal and survival data	79
4.5.2	Simulation Results and Discussion	80
4.6	Application on HIV/AIDS Data	85
4.6.1	Longitudinal submodel	85
4.6.2	Survival submodel	87
4.6.3	Joint modeling	89
5	Joint Analysis Based On H-likelihood	92
5.1	Introduction	92
5.2	Approximate Inference Based on the H-likelihood	93

5.3	Joint Analysis Based on H-Likelihood	94
5.4	Simulation Studies	98
5.4.1	Simulation Results	99
6	Joint Analysis under Misspecified Random Effects	104
6.1	Introduction	104
6.2	The Joint Model and Estimation	105
6.2.1	Two Submodels	105
6.2.2	The ML Estimation	106
6.3	Simulation Study	107
6.3.1	Generating longitudinal and survival data	107
6.3.2	Simulation Results	108
7	Application: GenIMS Study	113
7.1	Introduction	113
7.2	The Joint Models	119
7.3	ML Estimation	121
7.4	Results	123
8	Conclusions and Future Work	128
Appendix A	Estimation Procedure Based On H-likelihood Method	130
A.1	The (Joint) h-likelihood function	130
A.1.1	Estimation for $\theta = (\beta', \gamma', \varphi')'$	131
A.1.2	Estimation for ξ	136
Appendix B	R Codes of Simulation Study for Joint Model	144
	Bibliography	171

List of Figures

1.1	CD4 cell count measurements over time in AIDS study.(a) All subjects. (b) Six randomly selected subjects.	3
1.2	The histogram of all observed survival times (left panel), and Survival times of five randomly selected patients (right panel).	7
2.1	Longitudinal trajectories for square root CD4 cell counts for ten randomly selected subjects in the AIDS dataset.	32
3.1	Kaplan-Meier survival curves for ddC (zalcitabine) and ddl (didanosine) from 0-24 months.	56
3.2	Weibull regression diagnostic for covariates (drug, gender, prevOI, AZT). . .	58
3.3	Residual analysis of the survival model with Weibull distribution for AIDS data.	60
4.1	Graphical general representation of a joint model for a longitudinal process and an event process. Ovals represent latent variables while squares represent observed variables.	65
4.2	Spaghetti plot of CD4 cell counts of randomly selected 50 AISD patients. . .	86
7.1	Subject disposition for the entire Genetic and Inflammatory Markers of Sepsis (GenIMS) cohort. CAP indicates community-acquired pneumonia [31]. . . .	115
7.2	Logarithm of D-dimer levels of fifteen randomly selected sepsis patients over a week from GenIMS study.	116

7.3	The boxplot of overall D-dimer levels versus four baseline covariates age, steroid use, diabetes, and cardiac disease from GenIMS study.	117
7.4	Kaplan-Meier estimates of the survival probability for two groups of D-dimer levels: $\log(\text{D-dimer}) \leq 6.5$ and $\log(\text{D-dimer}) > 6.5$	118
7.5	Kaplan-Meier estimates of the survivor function for different groups of sepsis patients from GenIMS study. The estimates appears to be different among the groups.	119
7.6	Residual analysis of the survival submodel for GenIMS data. Top-left panel exhibits Q-Q plot of Cox-Snell residuals. Other three panels show scatterplots of Cox-Snell residuals against three covariates Age, Steroid use, and Diabetes.	125
7.7	Analysis of standardized residuals $r_{ij} = e_{ij}/\hat{\sigma}_\epsilon = (y_{ij} - \hat{y}_{ij})/\hat{\sigma}_\epsilon$ obtained from the linear mixed model for D-dimer level. Top-left panel shows scatterplot of r_{ij} against time (Day 1-7). Remaining five panels show Q-Q plots of the standardized residuals for five follow-up times (Day 1-5).	127

List of Nomenclatures

$\mathbf{I}_{n_i}$	$(n_i \times n_i)$ identity matrix
$\mathbf{y}_i$	$(n_i \times 1)$ vector of responses or repeated measurements for i^{th} individual
$\boldsymbol{\beta}$	$(p \times 1)$ vector of fixed effects parameters
$\boldsymbol{\varepsilon}_i$	$(n_i \times 1)$ vector of measurement or sampling errors
$\mathbf{D}$	unstructured variance-covariance matrix of random effect $\mathbf{u}_i$
$\mathbf{R}_i$	unstructured covariance matrix of random error $\boldsymbol{\varepsilon}_i$
$\mathbf{u}_i$	$(q \times 1)$ vector of random effect for i^{th} individual
$\mathbf{w}_i$	$(p \times 1)$ vector of covariates for the i^{th} individual in Cox regression/ Weibull PH models
$\mathbf{X}_i$	$(n_i \times p)$ design matrix corresponding to fixed effects $\boldsymbol{\beta}$
$\mathbf{Z}_i$	$(n_i \times q)$ design matrix corresponding to random effects $\mathbf{u}_i$
δ_i	Event indicator, denoted by $\delta_i = \mathbb{1}(s_i \leq C_i)$ with $\delta_i = 0$ when event time is right censored and $\delta_i = 1$ otherwise
$\exp(\lambda)$	the exponential density with parameter λ
γ	Corresponding vector of regression coefficients in Cox regression/Weibull PH models

C_i	Censoring time for i^{th} subject
$H(t)$	cumulative hazard function
$h(t)$	hazard function
$h_0(t)$	baseline hazard function at time t
n	Total number of individuals
n_i	Total number of measurements for i^{th} individual
$S(t)$	survival function at time t
s_i	Continuous random variable for event time/survival time/failure time for i^{th} individual, also indicates true event time
t_{ij}	Time of j^{th} measurement for i^{th} individual or subject
t_i	Observed survival time for i^{th} subject
u_i	frailty or random effect for individual or cluster i
V_i	Covariance matrix of y_i individual
$W(\lambda, \tau)$	the Weibull density with scale parameter λ and shape parameter τ
y_{ij}	Response variable for i^{th} individual at j^{th} measurement, $i = 1, 2, \dots, n; j = 1, 2, \dots, n_i$

List of Acronyms

AFT accelerated failure time.

AIC Akaike Information Criterion.

AIDS Acquired Immunodeficiency Syndrome.

Ave.Length Average Length of 95% Confidence Interval(CI).

CP Coverage Probability.

d.f. Degree of freedom.

GenIMS Genetic and Inflammatory Marker of Sepsis.

GLMM Generalized Linear Mixed Model.

HIV Human Immunodeficiency Virus.

i.i.d Identically Independently Distributed.

LME Linear Mixed Effects.

MLE Maximum Likelihood Estimator/Estimation/Estimate.

MSE Mean squared errors.

PH Proportional hazards.

Chapter 1

Introduction

1.1 Review of Longitudinal Data

Longitudinal data refer to repeated or multiple measurements of one or more response variables for a given individual taken during a long follow-up period of time. For example, a researcher wants to find out the increase in body fatness in girls after menarche (in the age group of 10-15 years old). For this, the researcher observes the percentage body fat of girls over that period of time to collect meaningful data. By obtaining repeated measurements from the same individual, longitudinal studies allow us to estimate how the response variable of individuals changes over time. Longitudinal data are also related to clustered data in which repeated measurements are taken for different clusters, where these measures may be correlated within each cluster. In fact, longitudinal data are a special case of clustered data in which the clusters are composed of repeated measurements obtained from a single individual at multiple periods of time rather than measurements of all individuals within-group (or cluster). The distinctive feature of longitudinal data is that the repeated observations from a given individual are generally correlated, but the observations between individuals are typically independent, which motivates most statistical methods to analyze these within-individual correlations (Fitzmaurice and Laird, 2011 [18]; Wu, 2010 [87]). Many medical

studies are longitudinal in nature. For example, in hyperlipidemia clinical trials, cholesterol levels among women over 45 who walk daily for a period of 15 years are measured regularly throughout the trial. The measurements within a woman are expected to be correlated, as the current value of cholesterol level is more likely to depend on the previous value. However, the measures of cholesterol level between two women are usually assumed independent if these participants are randomly selected. Thus the main objective of longitudinal studies is to assess within-individual changes in the response variables over a period of time. Unlike longitudinal studies, a cross-sectional study collects data on a single occasion with all variables remaining the same throughout the study. This type of study allows us to only obtain estimates of between-individual differences in the response.

A markable example of longitudinal data is immunologic markers known as CD4 cell counts that were repeatedly measured over time in an AIDS study designed to evaluate two anti-HIV. The CD4 cells are white blood cells that fight infection, so the more we have these cells, the better for our immune system. As Human Immunodeficiency Virus (HIV) infection, a person is diagnosed with Immune Deficiency Syndrome (AIDS) when the CD4 count drops below 200. Usually, the CD4 cell count increases when the HIV virus is controlled with effective HIV treatment. Thus CD4 cell measures were recorded over time to evaluate the disease progression of an HIV-infected. Figure 1.1 shows CD4 trajectories of all patients and six randomly selected patients with HIV-infection during antiretroviral treatment (didanosine, zalcitabine). In this clinical trial, HIV-infected patients (i.e., two CD4 cell counts ≤ 300) were randomly selected to receive either 500 mg didanosine (ddI) daily or 2.25 mg zalcitabine (ddC) per day. Then patients were followed-up for clinical disease progression and survival. A more detailed description of this clinical study can be found in (Goldman, Carlin and Crane, 1996 [19]). This figure shows some features of longitudinal data that have received much attention in the literature: (i) the number of repeated measurements is different across patients; (ii) large variation among patients (i.e., between-subject variation); (iii) large variation in the repeated measurements within a given

patient; (v) missing data at some measurement times for some subjects.

Figure 1.1: CD4 cell count measurements over time in AIDS study.(a) All subjects. (b) Six randomly selected subjects.

The classical regression models for cross-sectional data, such as linear or generalized linear models, are not appropriate for longitudinal data analysis due to the within-individual variations and the between-individual variations that may influence directly the correlation among the repeated observations. These regression models for cross-sectional data are extended to the corresponding mixed effects models by introducing random effects in the models to incorporate the correlation in longitudinal data, such as linear mixed effects (**LME**) models, nonlinear mixed effects (**NLME**) models, and generalized linear mixed effects (**GLME**) models. We will briefly review **LME** and **GLME** models in Sections 2.4-2.5. Other common methods for analyzing complex longitudinal processes use nonparametric and semiparametric models.

1.2 Review of Survival (Time-to-event) Data

In practice, the main purpose of many studies is to analyze the time until the occurrence of a prespecified event of interest. In such cases, the response variable is the time until that event occurs, which is often called *survival time*, *event time*, or *failure time*. That is, the response variable is a non-negative discrete or continuous random variable, and represents the time from a well-defined origin to a well-defined event (Moore, 2016 [51]). Therefore, these data are called *survival data* or *time-to-event data*, which are times to an event of interest such as time to death, time to occurrence of a tumor, time to an accident, or time to infection of disease. As mentioned by Collett (2014) [9], these types of data do not only arise in clinical and epidemiologic studies but also in a great number of applied fields. For instance, in sociology, we might be interested in the time of a first divorce or marriage, in marketing the customer lifetime to purchase products, in economics the time until the acceptance of jobs by unemployed, and in industry time to failure of a machine. There are some special characteristics that distinguish the analysis of survival data from other areas in statistics (Collett, 2014 [9]; Wu and Liu, 2012 [88]; Rizopoulos, 2012 [62]). The first feature is that the distribution of the failure times is skewed, and often skewed to the right due to the fact that survival times should be positive, so survival data usually do not follow symmetric distributions (e.g., normal distribution). The second characteristic is that event times for all subjects usually do not have equal follow-up times as subjects may enter the study at different times. Another important feature of survival times is that they may be censored, which occurs when the event time of interest is not completely observed for all participants in a study. In general, there are four reasons that might cause censoring to occur for some subjects in the study: lost to follow-up, drop out, competing risk, or termination of the study (Collett, 2014 [9]). According to Rizopoulos (2012) [62], the analysis of censored data typically depends on the nature of the censoring mechanism that describes how censoring occurs. The first classification censoring mechanism is related to the position of censoring times and true survival times. More specifically, we have three main types of censoring:

right, left, and interval. If the events of some subjects occur beyond the end of the study, then the data are right-censored (unobserved) since their event times occur after the last time point they have been observed. Some examples of right censoring are: the study participant has not yet experienced the relevant outcome, such as relapse or death, by the end of the study; the study participant is lost to follow-up during the study period, or a study participant drops out from the study due to the occurrence of competing event (e.g., death by an adverse drug reaction or by an accident). If the event is observed, but the exact event time is unknown, then the time is left-censored. For example, the age at which children learn a task, some already knew before the entry of the study; or the time at which a teenager begins to drink alcohol. Interval-censored data occurs when the event is observed, but participants come in and out of observation, so the exact event time (unobserved) is within a certain known specified time interval. For instance, when an HIV-infected patient is tested for the onset of AIDS and found positive, then when exactly this patient developed AIDS is only known to be at some point in between the pre-last and last tests. The second classification censoring mechanism is with regard to whether or not the probability of a subject being censored depends on the failure procedure. In particular, there are three assumptions about censoring often considered for survival data: *(i) independent censoring*: when we assume that the survival time T and the censoring time C are independent. This means that the actual survival time of a participant, T , is independent of any mechanism that causes that participant's survival time to be censored at time C , where $C < T$. *(ii) random censoring*: when the censored time for a subject is representative of all the study subjects who remained at risk at time t with respect to their survival experience. That is, the failure rate for subjects who remained in the risk set is assumed to be equal to the failure rate for subjects who are censored. *(iii) non-informative censoring*: whether censoring is non-informative or informative depends on the distribution of the time-to-event random variable T and time-to-censorship random variable C . In other words, censoring in survival analysis is "non-informative" when participants are lost to follow-up or drop out of the study due to

reasons unrelated to the study. In this thesis, we assume that censoring is non-informative, which is the most popular assumption in the literature.

As an illustrative real-life example of survival data, we consider bladder cancer data obtained from a clinical trial conducted by the Veterans Administration Co-operative Urological Research Group (VACURG) (Byar, 1977 [7]). All patients entering the trial had superficial bladder tumors, and the time of recurrence tumors of a patient measured from the removal of old tumors at a given visit until the recurrence of new tumors. In this example, we consider the dataset based on the first four tumor recurrence times from patients randomly assigned to one group of the two treatments: placebo and thiotepa (Wei, Lin, and Weissfeld, 1989 [84]). The objective of this clinical study is to evaluate the effectiveness of thiotepa in reducing the recurrences of tumours for patients with bladder cancer. In this study, the event time is the number of months from a given visit until the j th tumor recurrence for the i th patient. The event of interest here is the number of recurrences of tumours. The dataset contains 85 patients, 29 out of 47 patients among the placebo group have experienced at least one time of tumor recurrence by the end of the study, so their failure times are observed, while 18 patients did not experience any tumor recurrence before the end of the study, so their survival times are right-censored. In contrast, 18 out of 38 patients from the thiotepa group have experienced the event of interest, and 20 patients measured with a right-censored event. Figure 1.2 shows the survival times of five randomly selected patients, where two event times are right-censored and the other event times are observed, and the histogram of the distribution of all the observed survival times.

Figure 1.2: The histogram of all observed survival times (left panel), and Survival times of five randomly selected patients (right panel).

In the analysis of survival data, the hazard and survival functions are two primary ways of modeling or specifying the survival times distribution. The *survival function* defines the probability of surviving up to a point t . That is,

$$S(t) = Pr(T \geq t), \quad 0 < t < \infty \quad (1.1)$$

where T is the survival time. This function must be nonincreasing as t increases, and it is often defined in terms of the *hazard function*, which is instantaneous failure rate or hazard rate. The *hazard function* defines as the probability that, given a subject has survived time t , he or she fails in the next time interval $[t, t + \Delta t)$, divided by the length of that interval. That is,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}; \quad t \geq 0 \quad (1.2)$$

Since $h(\cdot)$ is called instantaneous risk function, the survival function can be expressed in

term of the hazard function (or risk function) as

$$S(t) = \exp\left\{-H(t)\right\} = \exp\left\{-\int_0^t h(s)ds\right\},$$

where $H(t)$ is known as the cumulative hazard (or cumulative risk) function that describe the accumulated risk up to time t . It is easy to observe that the relationships between survival function, hazard function, and cumulative hazard function are

$$h(t) = -\frac{\partial}{\partial t} \log S(t), \quad H(t) = -\log S(t), \quad f(t) = h(t)S(t), \quad (1.3)$$

where $f(t)$ is the probability density function of T . Therefore, a model based on one of the above three functions can also be expressed in terms of another function.

Due to the special features of survival data or event data, special statistical models and methods are required. The non-parametric model is widely used in the analysis of a single sample survival data or in the comparison of two or more groups of survival times. The well-known two non-parametric procedures for comparing groups of event times are the *log-rank test* and the *Wilcoxon test* (Collett, 2014 [9]). Also, the semi-parametric model is another method that is useful in the survival analysis. Both non-parametric and semi-parametric models do not require any assumptions about the distribution of survival data. Parametric models are also used for analyzing the event times, and these models assume that survival times follow some known probability distributions (Rizopoulos, 2012 [62]; Collett, 2014 [9]; Wu and Liu, 1963 [87]). Both semi-parametric and parametric survival models are used to examine the relationship between survival time and one or more covariates (e.g., age, gender, race, etc.). In this thesis, we focus on the Weibull proportional hazard model, which is a parametric model that assumes a Weibull distribution for the survival data. In the following sections, we give a brief review of commonly used non-parametric, semi-parametric, and parametric survival models. Comprehensive details of survival models and methods can be found in Collett (2014) [9] and Lawless (2011) [35].

1.3 Joint Modelling of Longitudinal and Survival Data

In clinical trials and longitudinal studies, it is increasingly common to collect repeated measurements on either a continuous or discrete response, times to events of interest, and additional covariate information on each participant simultaneously. Examples can be found in HIV clinical trials, where for each subject biomarkers, such as CD4 cell counts and viral load are both measured over time with time-to-AIDS or death, in which these two processes are usually correlated, (DeGruttola and Tu, 1994 [20]; Tsiatis et al., 1995 [80]). It is often of interest to investigate the *association* between the survival times and longitudinal trajectories of the CD4+ and viral load process in such AIDS studies. Yet another example is the prostate cancer study, where patients are followed-up over time during a period of study until death or metastasis (Yu et al., 2004 [91]; Proust-Lima and Taylor, 2009 [58]). In these cases, we must model the longitudinal process and the time-to-event process *jointly*, in order to avoid potential bias and to increase efficiency in the parameter estimation.

Depending on the focus of the analysis, joint modeling of longitudinal data and survival data are often attractive in the following three situations (Wu, 2010 [87]):

- (i) *Modelling Survival data with time-varying covariates, where the covariates may be missing or may be measured with errors.* In this situation, the main focus is on survival analysis and longitudinal models are only fitted for the time-varying covariates in order to address measurement errors and missing data.
- (ii) *Modelling longitudinal data with informative dropouts.* In this case, the main focus is on longitudinal analysis, and a survival model is fitted to model the time to dropout in order to reduce possible biases associated with informative dropouts.
- (iii) *Modeling both the longitudinal data and the survival data.* In this case, the main focus is on understanding the *association* between the two processes to reduce bias and to increase the efficiency in estimation. In most cases, the longitudinal model and survival

model are linked through shared parameters or shared variables, where the model is often called *shared parameter models*.

The joint modeling techniques have been developed in the past two decades as a powerful statistical tool to fit the two processes (longitudinal and survival) simultaneously via their shared information. The basic joint model setting consists of two submodels, one for the longitudinal and one for the survival outcome as introduced by Self and Pawitan (1992) [69], DeGruttola and Tu (1994) [20], Tsiatis et al (1995) [80], Faucett and Thomas (1996) [16], and Wulfsohn and Tsiatis (1997) [90]. Particularly, Self and Pawitan (1992) [69] considered a two-step estimation approach, where they condition on the survival information to compute expected values of the covariates. The partial likelihood was used to obtain estimates of the risk parameters and corresponding variances were derived to account for the uncertainty in the expected covariate values. DeGruttola and Tu (1994) [20] proposed a joint model in which the time-to-event is modeled parametrically, which leads to straightforward likelihood inference. Tsiatis et al. (1995) [80] proposed a two-step method in which a growth curve random components model with normal errors was assumed for the biomarker and a modified expectation maximization (EM) algorithm was used for the parameter estimation. A Cox regression model is used to obtain estimates of the survival parameters. Faucett and Thomas (1996) [16] generated the joint posterior distribution of all unknown parameters of the comprehensive model given the observed data by using the Markov Chain Monte Carlo (MCMC) method of Gibbs sampling. Wulfsohn and Tsiatis (1997) [90] used a full likelihood approach for the joint model based on a linear mixed model for the longitudinal process and a proportional hazards model for the survival process.

Henderson et al. (2000) [25] considered a more general model where two stationary Gaussian processes were postulated, including both random effects and serial correlation for the longitudinal measurements and survival times, respectively. Tsiatis and Davidian (2004) [79] proposed an interesting contradiction between the random effects and serial correlation assumption. Tsiatis and Davidian (2001) [78] and Song et al. (2002) [72] proposed

a conditional score approach and developed a set of unbiased estimating equations. They considered the model of Wulfsohn and Tsiatis (1997) [90], but the assumption of normality of the random effects was relaxed to a distribution with a smooth density. Brown and Ibrahim (2003) [60] presented a Bayesian semiparametric joint hierarchical model where a flexible specification of the subject-specific profiles is considered. In addition, Brown (2009) [5], Rizopoulos and Ghosh (2011) [64], and Rizopoulos (2012) [62] presented joint models assuming different association structures of the longitudinal and survival outcomes. Li et al. (2010) [40] presented a joint model with longitudinal ordinal measurements and competing risks, in which a partial proportional odds model for the longitudinal ordinal outcome is linked to the event times by latent random variables.

In comparison to the separate estimation procedure, joint modeling has greatly improved the model estimation and provided more reliable inferences of the longitudinal-survival relationship. In Chapter 4, we introduce joint models and discuss computational methods for the joint analysis of longitudinal and survival data. In the next few chapters, we describe various aspects of the joint modeling approach in more details.

1.4 Research Objectives and Thesis Outline

The main research objective of the thesis is to assess the association between the longitudinal measurement and survival outcome, where we consider jointly modelling the longitudinal and survival data for a valid inference. There has been active research on joint models for longitudinal and survival data in recent years, which includes Rizopoulos (2012) [62], Sattar and Sinha (2019) [66], Wu and Liu (2012) [88], Collett (2014) [9], Wulfsohn and Tsiatis (1997) [90], Lee and Nelder (1996) [36], Ha, Noh and Lee (2017) [23], Ha et al. (2001) [12], Ibrahim (2010) [28], Scharfstein et al. (1999) [67].

Although there is rich literature on joint modelling of longitudinal and survival data, existing models/methods still have some limitations. In this thesis, our main focus is on the

computational issues associated with the likelihood inference in the joint model research.

- In the first section of the thesis (Chapter 4), we study the standard likelihood approach for the statistical inference using joint models. We discuss exact maximum likelihood estimation of the model parameters, including inference for the regression coefficients and random effects. The maximum likelihood method provides estimates that are approximately unbiased and also provides expected coverage probabilities when estimating both longitudinal and survival submodels parameters. Simulations were carried out to examine the empirical properties of the likelihood estimators under both joint and separate linear mixed-effects and Weibull frailty models. The suggested technique is also applied to real data from a HIV/AIDS study.
- In the second section of the thesis (Chapter 5), we have studied an approximate method based on Laplace approximations to the hierarchical likelihood (h-likelihood) method. We carry out simulations to assess the performance of the approximate likelihood estimation in joint models using a linear mixed-effects submodel for longitudinal data and a frailty submodel for survival data. Our study shows that the h-likelihood method loses slight efficiency when approximating the maximum likelihood estimates in joint models.
- In the third section of the thesis (Chapter 6), we investigate the effect of misspecifying the random effects distribution in the parameter estimation of joint models with shared random effects. For the misspecification study, we assume that the frailty random effects u_i in the longitudinal and survival submodel follow a t-distribution with ν degrees of freedom. We investigate how the association parameter between the longitudinal and time-to-event outcomes influence the ML estimates of linear mixed effects and Weibull frailty submodels under both misspecified and correctly specified distribution of the frailty random effects u_i . We find that the use of the t-distribution is more robust than the commonly used normal distribution when u_i is misspecified.

- In the fourth section of the thesis (Chapter 7), we analyze a large clinical dataset of pneumonia patients obtained from the Genetic and Inflammatory Markers of Sepsis (GenIMS) study. An important focus of the data analysis is on the association between the longitudinal D-dimer biomarker and pneumonia-to-death outcome. As a result, an integrated analysis of correlated outcome processes via joint model is a desirable approach instead of a separate analysis for each outcome. We performed an extensive analysis of the GenIMS data to investigate important feature of the demographic variables, longitudinal biomarker and their association with the survival experience of pneumonia patients.

1.5 Organization of Thesis

The remainder of the thesis is organized as follows. Chapter 2 and 3 review some fundamental statistical models and methods for longitudinal data and survival time data, respectively, including simulation studies and application on real data from AIDS/HIV study. In Chapter 4, we present a joint modeling approach to analyzing longitudinal and survival data based on the exact likelihood method. In this approach, we study the role of the association parameter between longitudinal and time-to-event data to the joint analysis. The longitudinal and survival outcomes are modelled by using linear mixed-effects and Weibull frailty models, respectively. In Chapter 5, we develop and study an innovative method for jointly modeling longitudinal response and time-to-event data, called the "h-likelihood method". In this approach, we study the effect of this approximate method on the parameter estimation. In Chapter 6, we carry out simulations to study the impact of misspecifying the random effects distribution on the parameter estimation, in which the Gaussian assumption of random effects are not valid. In Chapter 7, we apply our proposed methods to real data from the GenIMS study and present the findings from this data analysis. Chapter 8 concludes this thesis with some discussion and future work.

Chapter 2

Analysis of Longitudinal Data

2.1 Introduction

In the past two decades, there have been many remarkable developments in statistical methodology for longitudinal data analysis. As mentioned in Section 1.1, there are two sources of variation in longitudinal data: one is the variation from within individual, and the other between individual variation. Modelling the variation within individual allows one to see the change in the longitudinal outcome over time during the follow-up time period, while modelling between-individual variation allows one to understand the differences between individuals (Diggle, 2008 [10]; Wu, 2010 [87]). Longitudinal studies vary in types of outcomes of interest, for example, when the outcomes are continuous and may be approximately normally distributed, the linear mixed-effects models (LMM) have been the dominant approach for the analysis of longitudinal data. For the discrete and non-Gaussian longitudinal data, instead, the method has been extended to generalized linear mixed-effects models (GLMM). In addition, Liang and Zeger (1986) [41] proposed the generalized estimating equations (GEE) approach for analysing discrete longitudinal data. In the following section, we will review briefly two specific types of mixed effects models for analyzing continuous and discrete longitudinal data, where the former type of data will be assumed in this thesis.

2.2 Notation

In this chapter, we will focus on the continuous longitudinal variable. We first define some notation that will be used throughout the chapter.

Let Y be a continuous longitudinal variable. Denote the observed values of Y by y_{ij} for individual i at time t_{ij} , $i = 1, 2, \dots, N; j = 1, 2, \dots, n_i$. In addition, matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ contain some other longitudinal data which may be used as covariates but not of primary interest.

For model parameters, we denote $\boldsymbol{\beta}$ as a vector of regression coefficients in a longitudinal model. We also denote $\mathbf{u}_i$ as a vector of random effects for individual i and $\boldsymbol{\varepsilon}_i$ as a vector of random errors.

2.3 General Linear Models

General linear models are perhaps the most widely used statistical models in which the mean of the response variable Y is assumed to be a linear function of parameters $\boldsymbol{\beta}$, which represent the effects of predictors X . Suppose that $\mathbf{y}_i$ is a vector of n_i observations of an individual i , $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})'$ and $\mathbf{X}_i$ is a matrix of corresponding covariates, then the model may be written as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \dots, N, \quad (2.1)$$

where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{in_i})'$ is a vector of random errors in the regression model.

The response vector $\mathbf{Y}_i$ is said to consist of two components: the systematic component or linear predictor ($\mathbf{X}_i \boldsymbol{\beta}$) which describes linear relationships between the response variable and the predictors, and the random component ($\boldsymbol{\varepsilon}_i$) which specifies the distribution of the error and hence the distribution of the response vector $\mathbf{Y}_i$. In general, the response vector $\mathbf{Y}_i$ is assumed to be distributed according to a multivariate normal distribution with the

mean $\boldsymbol{\mu}_i$ and covariance matrix $\mathbf{V}_i$, i.e., $\mathbf{Y}_i \sim MVN(\boldsymbol{\mu}_i, \mathbf{V}_i)$ where

$$\boldsymbol{\mu}_i = E(\mathbf{Y}_i | \mathbf{X}_i) = \mathbf{X}_i \boldsymbol{\beta}$$

and

$$\mathbf{V}_i = Cov(\mathbf{Y}_i | \mathbf{X}_i)$$

In longitudinal studies, the repeated observations from the same individual or subject are typically not independent. The dependence between these repeated observations may be captured by the covariance matrix $\mathbf{V}_i$, in which the off-diagonal elements would not be equal to zero. In general, the covariance matrix $\mathbf{V}_i$ embeds different association levels among observations from the i^{th} individual, which, in general, it may depend on covariates $\mathbf{X}_i$.

In most longitudinal studies, when the data vary greatly across individuals, we may allow each individual to have individual-specific parameters. These individual-specific parameters can be obtained by adding *random effects* or *individual effects* to the population parameters (Wu, 2010, [87]). In this case, the random effects in a model not only incorporate heterogeneity in the data but also incorporate correlation between the multiple measurements within each individual or cluster.

In the next sections and chapters, we provide a more detailed description of commonly used mixed effects models:

- Linear mixed effects models (LMEs), which can be obtained from linear regression models by introducing random effects;
- Generalized linear mixed models (GLMMs), which can be obtained from GLMs by introducing random effects;
- Frailty models, which can be obtained from survival models by introducing random effects.

2.4 Linear Mixed Effects Model

As explained in Section 1.1, in longitudinal studies, the repeated observations from individuals (or subjects) are measured over time, so an ordinary linear regression cannot be used for the analysis, as the assumption of independence does not hold (i.e., observations within each individual might be correlated). Therefore, linear mixed effects models are commonly used as an alternative approach to analyze continuous longitudinal data. The underlying assumption of the linear mixed effects model is that each individual in the population is supposed to have his/her own mean response trajectories over time. That is, the mean response is a linear function of predictor variables with regression coefficients that vary randomly from one individual to another (Diggle, 2002 [10]; Fitzmaurice and Laird, 2011 [18]). For example, suppose blood pressure measurements are taken for patients from both control and treatment groups in our study. By their very nature, each individual will have different average blood pressure. Therefore, to assess the effects of treatment, we use a model that would have a separate intercept for each patient, and we may also model the slope separately, as the response of each patient to the drug may vary from one subject to another (Molenberghs and Verbeke [49]). Linear mixed effects (LME) models can be viewed as an extension of the linear regression model by introducing random effects in the model. For this reason, it is considered as a mixed model, as it combines both fixed and random effects (Fitzmaurice, 2012 [17]; Wu, 1963 [87]; Rizopoulos, 2012 [62]). The distinctive feature of the LME is that we estimate parameters for the mean response of the population, β , and also predict the effect of each individual's response trajectories over time (Fitzmaurice, 2012 [17]). Another advantage of using the LME model is that it is not necessary to have balanced data; that is, the linear mixed effects model can be used to model different number of measurements for each individual, which can be also taken at a different set of times (Wu, 2010 [87], Rizopoulos, 2012 [62]; Fitzmaurice and Laird, 2011 [18])

In general, a linear mixed effects model is any model that satisfies (Laird, 1982 [33];

Molenberghs and Verbeke, 2000 [49]; Wu, 2010 [87]):

$$\begin{aligned} y_{ij} &= \beta_1 x_{1ij} + \cdots + \beta_p x_{pij} + u_{i1} z_{1ij} + \cdots + u_{iq} z_{qij} + \varepsilon_{ij}, \\ u_{ik} &\sim N(0, \sigma_k^2), \text{Cov}(u_{ik}, u_{ik'}) = \sigma_{kk'}^2, \\ \varepsilon_{ij} &\sim \text{ind. } N(0, \sigma_\varepsilon^2). \end{aligned} \tag{2.2}$$

Alternatively, in matrix form

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \dots, N, \\ \mathbf{u}_i &\sim N(\mathbf{0}, \mathbf{D}), \quad \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i), \end{aligned} \tag{2.3}$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})'$ is an n_i -dimensional vector of longitudinal reponses for subject i , $\boldsymbol{\beta}$ is a p -dimensional vector for fixed effects parameters, $\mathbf{X}_i$ is an $(n_i \times p)$ design matrix of fixed effects $\boldsymbol{\beta}$ and $\mathbf{u}_i$ is a q -dimensional vector of random effects parameters that represents the characterization of between-individual variation. Also, $\mathbf{Z}_i$ is an $(n_i \times q)$ design matrix of random effects $\mathbf{u}_i$, and $\boldsymbol{\varepsilon}_i$ is an n_i -dimensional vector of random errors that represents the characterization of within-individual variation. In model (2.3), we assume that random effects $\mathbf{u}_i$ and random error $\boldsymbol{\varepsilon}_i$ follow a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{D}$. The within-individual error $\boldsymbol{\varepsilon}_i$ is assumed to be independent, and $\mathbf{u}_i$ is assumed to be independent of the random errors $\boldsymbol{\varepsilon}_i$ (i.e., $\text{Cov}(\boldsymbol{\varepsilon}_i, \mathbf{u}_i) = \mathbf{0}$) (Fitzmaurice and Laird, 2011 [18]; Wu, 1963 [87]). As a result, we have

$$E \begin{bmatrix} \mathbf{u}_i \\ \boldsymbol{\varepsilon}_i \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad \text{Var} \begin{bmatrix} \mathbf{u}_i \\ \boldsymbol{\varepsilon}_i \end{bmatrix} = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_i \end{bmatrix}.$$

Note that the random effects $\mathbf{u}_i$ can be interpreted as deviations of individual i from the population mean, while the mean parameters $\boldsymbol{\beta}$ is interpreted as the same as in a linear regression model. We further assume that $\mathbf{R}_i$ represents the diagonal matrix $\sigma_\varepsilon^2 \mathbf{I}_{n_i}$, with $\mathbf{I}_{n_i}$ being an $n_i \times n_i$ identity matrix, where a more detailed review about the specification of the covariance matrix $\mathbf{R}_i$ is provided in Fitzmaurice and Laird (2011) [18] and Wu (2010) [87].

The covariance of $\mathbf{Y}_i$ can be written as

$$\begin{aligned} \text{Var}(\mathbf{Y}_i) = \mathbf{V}_i &= \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \mathbf{R}_i \\ &= \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \sigma_\varepsilon^2 \mathbf{I}_{n_i}. \end{aligned} \quad (2.4)$$

Then we have the conditional and marginal distributions of the response Y_i , given by Helen (2014) [6]

$$\begin{aligned} \mathbf{Y}_i | \mathbf{u}_i &\sim N(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i, \sigma_\varepsilon^2 \mathbf{I}_{n_i}), \\ \mathbf{Y}_i &\sim N(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{V}_i), \quad \mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \sigma_\varepsilon^2 \mathbf{I}_{n_i}. \end{aligned}$$

The linear mixed effect model (2.3) can be viewed as a hierarchical two stage model because such a mixed model allows the analysis of within-subject and between-subject sources of variation, where stage 1 specifies the within-subject variation, that is given by $\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i + \boldsymbol{\varepsilon}_i$, and stage 2 specifies the between-subject variation, that is given by $\mathbf{u}_i \sim N(\mathbf{0}, \mathbf{D})$ and $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$.

2.4.1 Inference for Linear Mixed Effects Model

The expectation-maximization (EM) algorithm of Laird (1982) [33] can be used to obtain the statistical inference for linear mixed effects model. Newton-Raphson and Fisher scoring algorithms for estimation and inference of model (2.3) is proposed by Lindstrom and Bates (1988) [44] and Stefanski and Carroll (1987) [74]. Until now, the statistical inference of an LME model is mostly based on the maximum likelihood and restricted maximum likelihood methods (Wu, 2010 [87]; Rizopoulos, 2012 [62]).

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_\varepsilon^2, \mathbf{D})$ denote all parameters in LME model (2.3). Then the likelihood

function for the observed data $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_N)'$ is given by

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{y}) &= \prod_{i=1}^N f(\mathbf{y}_i; \boldsymbol{\theta}) \\ &= \prod_{i=1}^N \int f(\mathbf{y}_i|\mathbf{u}_i; \boldsymbol{\beta}, \sigma_\varepsilon^2) f(\mathbf{u}_i|\mathbf{D}) d\mathbf{u}_i, \end{aligned} \quad (2.5)$$

where $f(\mathbf{y}_i|\mathbf{u}_i; \boldsymbol{\beta}, \sigma_\varepsilon^2)$ is the normal density with mean vector $\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i$ and covariance matrix $\sigma_\varepsilon^2 \mathbf{I}_{n_i}$, and $f(\mathbf{u}_i|\mathbf{D})$ is the normal density with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{D}$. Maximization of the likelihood function (2.5) is typically based on an iterative algorithm, such as the expectation maximization (EM) algorithm or the Newton-Raphson method (Lindstrom and Bates, 1988 [44]; Wu and Yu, 2016 [89]). Assuming that the longitudinal responses of subjects are independent conditionally on their random effects, the log-likelihood of the **LME** model is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^N \log \int f(\mathbf{y}_i|\mathbf{u}_i; \boldsymbol{\beta}, \sigma_\varepsilon^2) f(\mathbf{u}_i|\mathbf{D}) d\mathbf{u}_i. \quad (2.6)$$

For a known covariance matrix $\mathbf{V}_i$, the maximum likelihood estimator of the fixed effects parameter $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^N \mathbf{X}'_i \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}'_i \mathbf{V}_i^{-1} \mathbf{y}_i, \quad (2.7)$$

which is an unbiased estimate of $\boldsymbol{\beta}$.

The asymptotic distribution of $\hat{\boldsymbol{\beta}}$ is multivariate normal with the mean being the true value of $\boldsymbol{\beta}$ and the covariance given by

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^N \mathbf{X}'_i \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1}. \quad (2.8)$$

In case $\mathbf{V}_i$ is unknown, we use the estimate $\hat{\mathbf{V}}_i$ in equation (2.7) to find an estimate

of β . Restricted maximum likelihood (REML) method can be employed to estimate the covariance matrix V_i , where the likelihood does not contain β and it is defined only in terms of V_i . More intensive discussion on statistical inference with LME models based on likelihood methods can be found in Molenberghs and Verbeke (2000) [49], Fitzmaurice and Laird (2011) [18] and Rizopoulos (2012) [62].

2.4.2 Covariance Matrix Structure

2.4.2.1 Random-intercept model

The simplest case of a linear mixed effects model is the mixed model with a randomly varying subject effect. This model postulates that the longitudinal responses of all subjects have exactly the same evolution in time, but the subjects vary at the baseline level. This type of linear mixed model is called the random-intercepts model, where we fit a linear regression model with individual-specific intercepts but a constant slope. Basically, a linear regression model for cross-sectional data is extended by incorporating the random effect (or individual effect) into the intercept but not into the slope, yielding the following model (Fitzmaurice and Laird, 2011 [18]; Rizopoulos, 2012 [62]):

$$\begin{cases} y_{ij} = \beta_0 + \beta_1 t_{ij} + u_{i0} + \varepsilon_{ij}, \\ u_{i0} \sim N(0, \sigma_u^2), \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2), \quad i = 1, 2, \dots, N, j = 1, 2, \dots, n_i. \end{cases} \quad (2.9)$$

We can define the intercept for the i th subject (or individual-specific parameter) as $(\beta_0 + u_{i0})$, which differs randomly from one subject to another. That is, the response for the i th individual at the j th measurement time is assumed to vary from the population mean, $\beta_0 + \beta_1 t_{ij}$, by an individual effect, u_{i0} , and within-individual measurement error, ε_{ij} , and both u_{i0} and ε_{ij} are assumed to be random and independent (Fitzmaurice and Laird, 2011

[18]). In this model, the conditional and marginal means of Y_{ij} are given by

$$\begin{aligned} E(Y_{ij}|u_i) &= \mathbf{x}'_{ij}\boldsymbol{\beta} + u_{i0}, \\ E(Y_{ij}) &= \mathbf{x}'_{ij}\boldsymbol{\beta}, \end{aligned} \tag{2.10}$$

where $\mathbf{x}_{ij} = (1, t_{ij})'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)'$. Note that the former describes the mean response trajectory over time for each subject and the latter describes the mean response of the profile population. For model (2.9), the marginal variance is defined as $\sigma_u^2 + \sigma_\varepsilon^2$, which is constant over time, while the marginal covariance between any pair of responses $(y_{ij}, y_{ij'})$ is equal to σ_u^2 . Thus the marginal covariance structure for the longitudinal responses under the random-intercepts model has the form

$$Var(\mathbf{Y}_i) \equiv \mathbf{V}_i = \sigma_u^2 \mathbf{J}_{n_i} + \sigma_\varepsilon^2 \mathbf{I}_{n_i}, \tag{2.11}$$

where $\mathbf{I}_{n_i}$ is an $n_i \times n_i$ identity matrix. Given the marginal covariance, σ_u^2 , the correlation between the repeated measurements of any pair is given by

$$Cor(Y_{ij}, Y_{ik}) = \rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2}, \quad j \neq k, \quad j, k = 1, 2, \dots, n_i. \tag{2.12}$$

As the random subject effect, u_{i0} , induces correlation among the repeated measurements, we can conclude that there is no correlation (i.e., $\rho = 0$) when there is no random effect (i.e., $\sigma_u^2 = 0$). By contrast, the correlation will be high when the within-subject variation is much smaller than the between-subject variation (i.e., $\sigma_\varepsilon^2 \ll \sigma_u^2$) (Fitzmaurice and Laird, 2011 [18]). Thus the correlation between repeated measurements within-subject i as defined in (2.12) underlines an important aspect of mixed effect models.

2.4.2.2 Random-Intercept and Random-Slope Model

Although the random-intercept model considered above is attractive for its simplicity, the assumption of constant correlation between the repeated measurements is unrealistic in prac-

tice (Rizopoulos, 2012 [62]). Intuitively, we would expect that the response trajectory for each subject differs not only in their baseline level, but also in terms of changes in their responses at subsequent time points. An extension of model (2.9) that allows the response to vary randomly among subjects at any occasion is the random-intercepts and random-slopes model. That is,

$$\begin{cases} y_{ij} = \beta_0 + \beta_1 t_{ij} + u_{i0} + u_{i1} t_{ij} + \varepsilon_{ij}, \\ (u_{i0}, u_{i1})' \sim N(0, \mathbf{D}), \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2), \quad i = 1, 2, \dots, N, j = 1, 2, \dots, n_i, \end{cases} \quad (2.13)$$

with y_{ij} being the response, time t_{ij} the covariate, and ε_{ij} the random error.

Using vector and matrix notation, model (2.13) can be expressed as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \dots, N, \quad (2.14)$$

where

$$\mathbf{X}_i = \mathbf{Z}_i = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \mathbf{u}_i = \begin{bmatrix} u_{i0} \\ u_{i1} \end{bmatrix}, \mathbf{D} = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}, \quad j = 1, 2, \dots, n_i.$$

Thus the marginal covariance structure for any pair of responses of the same subject is given by

$$\text{cov}(Y_{ij}, Y_{ij'}) = \begin{bmatrix} 1 & t_{ij} \end{bmatrix} \mathbf{D} \begin{bmatrix} 1 \\ t_{ij'} \end{bmatrix} + \sigma_\varepsilon^2 = d_{11} + (t_{ij} + t_{ij'})d_{12} + t_{ij}t_{ij'}d_{22} + \sigma_\varepsilon^2,$$

and the marginal variance $\text{Var}(Y_{ij}) = d_{11} + 2t_{ij}d_{12} + t_{ij}^2d_{22} + \sigma_\varepsilon^2$. Note that both the marginal mean and covariance of $\mathbf{Y}_i$ are functions of times of measurement t_{ij} . Thus the marginal

distribution of $\mathbf{Y}_i$ is given by

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \sigma_\varepsilon^2\mathbf{I}_{n_i}).$$

The main feature of the covariance pattern $Var(\mathbf{Y}_i) \equiv \mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \sigma_\varepsilon^2\mathbf{I}_{n_i}$ under the random-intercept and random-slope model is to distinguish between the different sources of within-individual and between-individual variation (Fitzmaurice and Laird, 2011 [18]; Wu, 2010 [87]).

2.5 Generalized Linear Mixed Model

Generalized linear mixed models (GLMMs) can be used to analyze longitudinal data. It can be viewed as an extension of the generalized linear model in which random effects are added to the linear predictor. GLMMs are the most frequently used random effects models for analysing discrete or non-normal repeated measurements $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})'$, where y_{ij} 's are assumed to follow a distribution in the exponential family.

$$f(y_{ij}|u_i) = \exp \left\{ \frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi) \right\}, \quad (2.15)$$

where θ_{ij} is the natural or canonical parameter, ϕ is the dispersion parameter, and a , b , and c are specific functions. It can also be shown that Y_{ij} has the mean and variance (Molenberghs and Verbeke, 2005 [50]), given by

$$E(Y_{ij}) = \mu_{ij} = \frac{\partial b(\theta_{ij})}{\partial \theta_{ij}}, \quad Var(Y_{ij}) = a(\phi) \frac{\partial^2 b(\theta_{ij})}{\partial \theta_{ij}^2}.$$

We assume the canonical parameter θ_{ij} is given by

$$\begin{aligned} \theta_{ij} &= g(\mu_{ij}) = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{u}_i, \\ \mathbf{u}_i &\sim N(\mathbf{0}, \mathbf{D}), \quad i = 1, 2, \dots, N. \end{aligned} \quad (2.16)$$

Similarly to LME models, the repeated measurements y_{ij} 's of individual i at time j are assumed conditionally independent given the random effects $\mathbf{u}_i$, while $\mathbf{X}_i$ and $\mathbf{Z}_i$ are design matrices, and $g(\cdot)$ is a known link function. GLMMs assume that the response is linked to a function of covariates with fixed regression coefficients and random coefficients rather than modeling the responses directly. GLMMs allow us to model response variables from different distributions besides Gaussian, such as binary and poisson responses. For instance, if the longitudinal data $y_{i1}, y_{i2}, \dots, y_{in_i}$ are binary outcomes, we use a logistic (logit) link function

$$g(y) = \log\left(\frac{y}{1-y}\right),$$

while we may consider a log link function $g(y) = \log(y)$ for count outcomes. As for continuous outcomes where we assume a normal distribution, the link function is simply the identity $g(y) = y$, and the above GLMM model becomes an LME model.

Again, similar to LME models, parameters in the generalized linear mixed effects model are estimated based on the likelihood method. The likelihood function can be written as

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{y}) &= \prod_{i=1}^N \int f(\mathbf{y}_i|\boldsymbol{\beta}, \mathbf{u}_i) f(\mathbf{u}_i|\mathbf{D}) d\mathbf{u}_i, \\ &= \prod_{i=1}^N \int \prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{\beta}, \mathbf{u}_i) f(\mathbf{u}_i|\mathbf{D}) d\mathbf{u}_i, \end{aligned} \quad (2.17)$$

where $f(y_{ij}|\boldsymbol{\beta}, \mathbf{u}_i)$ is a probability density function of the j^{th} response from the the i^{th} subject in the exponential family (2.15), while $f(\mathbf{u}_i|\mathbf{D})$ is the probability distribution of the i^{th} random effect. Compared to an LME model, the evaluation of the likelihood function is more complicated in GLMMs. We cannot derive analytic or closed-form solutions for any estimates in the GLMM, as the likelihood function are nonlinear in the random effects $\mathbf{u}_i$. The addition of the random effects also makes likelihood inference more challenging because with each random effect u_i , the dimensionality of the integral also increases, adding to the intractability of the likelihood function (2.17). A Monte Carlo EM algorithm may be used to

find the **MLEs**, but it can be computationally very intensive, especially when the dimension of the random effects is not low. In general, inference for GLMMs can be conducted either through approximations of the likelihood or with Bayesian methods.

Approximate methods, such as penalized quasi-likelihood or Laplace approximation, and integrating based on Gaussian-Hermite quadrature may be used to find the ML estimates (Youngjo, 2006 [38]). As for Bayesian methods, inference is conducted through approximating the posterior distribution of the parameters in the GLMM, which is usually accomplished with the Markov chain Monte Carlo (MCMC) algorithm. A more detailed review of these methods is discussed in McCulloch (2008) [48] and Wu (2010) [87].

2.6 Simulation Study

In this section, we carry out a series of simulations to study the empirical properties of regression coefficients and variance components in linear mixed models. In particular, we obtain empirical biases, mean squared errors (MSE) and coverage probability (CP) of the estimators for different combinations of sample sizes.

2.6.1 Overview of Evaluation Criteria

Bias: The bias of an estimator $\hat{\theta}$ of a parameter θ is defined as the difference between the expected value of $\hat{\theta}$ and the true value of the parameter θ . The bias is approximated as

$$Bias(\hat{\theta}) \cong \frac{1}{S} \sum_{s=1}^S \left\{ \hat{\theta}^{(s)} - \theta \right\}, \quad (2.18)$$

where $\hat{\theta}^{(s)}$ is the estimator of θ obtained at the s^{th} simulation ($s = 1, 2, \dots, S$).

Mean Squared Error (MSE): The mean squared error measures the average squared differences between the estimator $\hat{\theta}$ and the parameter θ . The MSE of an estimator $\hat{\theta}$

of a parameter θ is approximated as

$$MSE(\hat{\theta}) \cong \frac{1}{S} \sum_{s=1}^S \left(\hat{\theta}^{(s)} - \theta \right)^2. \quad (2.19)$$

Coverage Probability (CP): The coverage probability of an estimator $\hat{\theta}$ for a 95% confidence interval on θ is obtained from

$$CP(\hat{\theta}) = \frac{1}{S} \sum_{s=1}^S I\{|\hat{\theta}^{(s)} - \theta| \leq 1.96 \times SE(\hat{\theta}^{(s)})\}, \quad (2.20)$$

where $SE(\hat{\theta})$ is an estimate of the standard error of $\hat{\theta}$ and I is an indicator variable, where the 95% confidence interval for θ is obtained as

$$CI(\hat{\theta}) = \hat{\theta} \pm 1.96 \times SE(\hat{\theta}). \quad (2.21)$$

Average Length: The length of a confidence interval for θ is obtained as the difference between two confidence limits:

$$L.CI(\hat{\theta}) = \{\hat{\theta} + 1.96 \times SE(\hat{\theta})\} - \{\hat{\theta} - 1.96 \times SE(\hat{\theta})\}, \quad (2.22)$$

where the average length of a confidence interval (**Ave.Length**) is obtained by

$$Ave.L.CI(\hat{\theta}) = \frac{1}{S} \sum_{s=1}^S L.CI(\hat{\theta}^{(s)}). \quad (2.23)$$

2.6.2 Generating Longitudinal Data

Using simulation studies, we evaluate the maximum likelihood estimators in the linear mixed effect model for continuous longitudinal data $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})', i = 1, 2, \dots, N$. We evaluate the estimates by their biases, mean square errors (MSEs), coverage probability, and

average lengths of 95% confidence intervals, which are defined in Section 2.6.1.

The longitudinal outcomes were generated from the mixed model

$$y_{ij} = \beta_0 + \beta_1 time_j + \beta_2 time_j \times x_{1i} + u_{0i} + \epsilon_{ij}, \quad i = 1, \dots, N, j = 1, \dots, n_i. \quad (2.24)$$

We first generate the random effect, u_i , from a normal distribution with mean 0 and variance $\sigma_u^2 = 1.0$, while the random error, ϵ_{ij} , is generated from $N(0, \sigma_\epsilon^2 = 0.5)$. The covariate x_{1i} is generated as a binary treatment indicator (i.e., x_i is 0 for the control group and 1 for the treatment group), and $time_j$ is considered as the follow-up times (in month) at which the longitudinal measurements were taken. Throughout the simulations, the true values of the regression parameters and variance components were fixed at $(\beta_0, \beta_1, \beta_2) = (10.00, -0.25, 1.5)$ and $(\sigma_\epsilon^2, \sigma_u^2) = (0.5, 1.0)$, respectively. Each simulation run was based on 1000 replications of data sets, with each data set containing a combination of sample sizes, $N = \{100, 150, 240, 400\}$.

2.6.3 Simulation Results and Discussion

Table 2.1 presents empirical biases, mean squared errors (MSEs) and coverage probabilities (CPs) of the ML estimators of the regression parameters $(\beta_0, \beta_1, \beta_2)$ and variance components $(\sigma_\epsilon^2, \sigma_u^2)$ for the longitudinal model (2.24). It is obvious that the maximum likelihood method provides approximately unbiased estimates of the regression parameters and variance components under all simulation configurations considered, except the intercept parameter (β_0) and variance component σ_u^2 , for which the estimates appear to be slightly biased. The corresponding mean squared errors (MSEs) of the ML estimators are also small; except MSEs of (β_0) and (σ_u^2) are somewhat larger. Overall, the MSEs of the estimators tend to decrease as the sample size N increases. Moreover, we see that the coverage probabilities (CPs) of the ML method are generally close to the nominal 95% confidence level, whereas the CP values for σ_u^2 and β_2 are above the nominal level when the sample size $N=400$. We conclude that the ML method performs well in the sense that the ML estimate has negligible biases and

MSEs, and also provides accurate coverage probabilities.

Table 2.1: Empirical percentage relative biases, mean squared errors, and coverage probabilities of maximum likelihood (ML) estimators under linear mixed effect (LME) model. [True parameter values: $(\beta_0, \beta_1, \beta_2) = (10, -0.25, 1.5)$ and $(\sigma_\epsilon^2, \sigma_u^2) = (0.5, 1)$].

Parameter	Bias	MSE	CP	Parameter	Bias	MSE	CP
$N = 100$				$N = 150$			
β_0	0.033917	0.015859	94	β_0	0.017531	0.013171	88
β_1	0.001101	0.000130	94	β_1	0.002487	0.000105	92
β_2	0.001505	0.000045	92	β_2	0.002093	0.000028	94
σ_ϵ^2	-0.003376	0.000996	96	σ_ϵ^2	-0.000966	0.00075	95
σ_u^2	-0.029693	0.023595	93	σ_u^2	-0.012682	0.02131	90
$N = 240$				$N = 400$			
β_0	0.027397	0.006365	93	β_0	0.018203	0.003509	91
β_1	0.001600	0.000058	92	β_1	0.002710	0.000033	96
β_2	0.001177	0.000015	96	β_2	0.000781	0.000009	97
σ_ϵ^2	-0.003178	0.000499	92	σ_ϵ^2	0.002544	0.000318	91
σ_u^2	-0.023713	0.010502	95	σ_u^2	0.002169	0.006599	98

2.7 Application on HIV/AIDS Study

In this section, we describe a dataset from a randomized clinical trial involving **HIV/AIDS** patients. The main objective of this study is to compare the safety and efficacy of two antiretroviral drugs (didanosine and zalcitabine) in treating patients who had failed or were intolerant of Zidovudine (AZT) therapy. More details regarding this study can be found in Abrams (1994) [1].

2.7.1 Background

The Human Immunodeficiency Virus (HIV) is one of the most global serious public health issues, with an estimated 38 million people living with HIV in 2019 (Haddad, 2019 [24]). HIV

is a virus that gradually attacks the human immune system and eventually develops to AIDS (acquired immunodeficiency syndrome) in the majority of cases if left without treatment. Current Anti-HIV can control the virus so that people with HIV can live long and healthy lives. A current treatment regimen used to treat patients with HIV/AIDS is Zidovudine (AZT) therapy, which prevents HIV from replicating and from destroying the immune system of an infected person and reduce the amount of viral load in the body (Wilson, 2013 [86]). Due to effective antiretroviral therapy, HIV infection has changed from a disease with high morbidity and mortality to a chronic disease in developed countries such as Canada (Hogg, 1999 [26]). In 2019, according to the World Health Organization (WHO), 68% of adults and 53% of children living with HIV globally were receiving lifelong antiretroviral therapy (ART). However, it has been found that patients can be intolerant to AZT or experience a disease progression despite it. Thus researchers provide alternative treatment or intervention methods such as antiretroviral drugs: Didanosine (ddl) and Zalcitabine (ddC). Both drugs are popular in use to treat patients with HIV infection who cannot tolerate zidovudine treatment or who have had experienced a "failure" (Abrams, 1994 [1]). In HIV studies, blood samples and viral loads often are repeatedly collected over time for each patient in order to measure immune responses, such as CD4 lymphocyte cell count. These repeated immune response measurements of CD4 cells are important biomarkers of HIV disease progression and the survival of HIV-infected patients (Lim, 2013 [42]; Kagan, 2015 [29]). CD4 cells (also known as CD4+ T cells) are white blood cells that fight infection. A person will receive a diagnosis of HIV when a CD4 count is lower than 200 cell/mm³. The repeated measurements of CD4 cells are the longitudinal data of interest.

2.7.2 AIDS Dataset Description

The AIDS study was constructed at 133 clinical sites in 16 units participating in the Community Programs for Clinical Research on AIDS (CPCRA). The study began in December of 1990 and completed enrollment of 467 patients the following September. The study con-

cluded as planned in September of 1992 (Abrams, 1994 [1]). Patients with HIV infection who were at least 13 years of age (except pregnant or breast-feeding women) were enrolled if they met the following standard: (i) they had an AIDS-defining condition or they had two CD4 counts of 300 cells or less per cubic millimeter, (ii) with either a positive serologic test for HIV or a clinician's working diagnosis of HIV infection, (iii) they had undergone zidovudine therapy that led to intolerance of the drug or progression of disease during therapy (Abrams, 1994 [1]). After their eligibility was confirmed, they were randomly assigned to receive two antiretroviral drugs, either Didanosine (ddl) or Zalcitabine (ddC). However, if drug intolerance developed at any time or the disease progressed anew after 3 months of treatment, a patient was allowed to switch the drug treatment once with a washout period of at least 3 days. The dataset consisted of 1408 observations on nine variables. Specifically, the data consisted of *three continuous* explanatory variables: CD4 (*the square root of the CD4 cells count*), Time (*the time to death or censoring*), and Obstime (*the observed time points at which CD4 measurements are taken*). The CD4 cells count were taken throughout a blood sample and recorded at baseline (0 months) and at multiple time points (2, 6, 12 and 18 months) during the trial. The data also consisted of other *six binary* explanatory variables: Gender (*female, male*), Drug (*ddl, ddC*), Death (*death=1, censored=0*), PrevOI (at study entry, AIDS=previous opportunistic infection of AIDS diagnosis, noAIDS=no previous AIDS diagnosis), and AZT (failure or intolerance). In total, 237 received ddC and 230 patients received ddl. There were 422 males and 45 females participated in the study. At the beginning of the study, there were 307 patients who had a diagnosis of AIDS and 160 patients who did not have a previous diagnosis of AIDS. A total of 292 patients were intolerant to AZT while 175 experienced a blue "failure" to AZT, which means the disease progressed despite AZT. Figure 2.1 presents longitudinal trajectories of the square root of the CD4 cell count for a random sample of ten patients.

Let y_{ij} represent the square root of the j^{th} CD4 counts on the i^{th} individual, $j = 1, 2, \dots, n_i; i = 1, 2, \dots, N$. Figure 2.1 shows longitudinal trajectories of the square-root

of CD4 cell counts for ten randomly selected AIDS patients.

Figure 2.1: Longitudinal trajectories for square root CD4 cell counts for ten randomly selected subjects in the AIDS dataset.

2.7.3 AIDS Data Analysis

In the HIV/AIDS dataset, the repeated measurements of CD4 cell counts are the longitudinal data of interest. We use a simple linear mixed-effect model given by

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + u_{i0} + \varepsilon_{ij}, \quad (2.25)$$

for $j = 1, 2, \dots, n_i; i = 1, 2, \dots, N$, where $N = 467$ were individuals followed over the period of the study (0-18 months), and t_{ij} denotes the observed time points for CD4 measurements. Also, ε_{ij} denotes the mutually independent measurement errors, β_0 is the intercept, β_1 is the time effect, and u_{i0} is the random intercept effect for the i^{th} patient. As mentioned earlier, this special type of linear mixed model is known as the random-intercept model, in which we assume all patients have exactly the same evolution in time but patients differ at baseline (Rizopoulos, 2012 [62]). The fitted model shows a significant negative effect of observation

time on CD4 count, which indicates that on average the CD4 cell counts declines in time. However, the model (2.25) is based on the strong assumption that patients have different CD4 counts at the baseline, but their decreasing rate is all the same with time.

Unlike the random-intercept model (2.25), the next model assumes that the change rates of CD4 counts with time are different for different subjects. Specificity, we fit a linear mixed-effects model with random intercept and slope for the observation time, given by

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + u_{i0} + u_{i1} t_{ij} + \varepsilon_{ij}. \quad (2.26)$$

The results of the fitted model (2.26) are shown in Table 2.2. The estimate of the time effect β_1 is obtained as -0.1501 , which describes the negative effect of observation time on CD4 cell count. That is, for every one month increase in observation time, the CD4 cell count decreases by 0.1501 cells per cubic millimeter of blood. Obstime seems a significant predictor of CD4 cell count as t-value= -9.88 and p-value <0.0001 . A decrease of CD4 cell counts is very harmful to a patient's autoimmune response in which the body loses the ability to fight diseases.

Table 2.2: Summary of Fitted Model (2.26) with Random Intercept and Slope

LME Model	AIC	BIC	logLik		
Random Int. and Slp.	7141.282	7172.76	-3564.641		
Random Effect	StdDev	Corr			
Intercept	4.5065	(Intr)			
obstime	0.1729	-0.152			
Residuals	1.7508				
Fixed Effect	Estimate	Std. Error	DF	T-Value	P-value
Intercept	7.1890	0.2222	937	32.3605	0
obstime	-0.1501	0.0152	937	-9.8844	0

In the next step, we examined the treatment effect on CD4 cell count by controlling other covariates in the model. To achieve this, we fit a linear mixed-effects model with a random intercept and random slope for the observation time t_{ij} and include all four covariates

"Drug", "Gender", "Prev OI" and "AZT" , that is,

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 Drug_i + \beta_3 Gender_i + \beta_4 PrevOI_i + \beta_5 AZT_i + u_{i0} + u_{i1} t_{ij} + \varepsilon_{ij}. \quad (2.27)$$

The results in Table 2.3 show that "Prev OI" and observation time (t_{ij}) are significant predictors of the CD4 cell counts. In comparison to the fitted model (2.25) and (2.26), the AIC value is smaller for model (2.27) with the added covariates ($AIC = 7020.004$), as compared to the model with only a random intercept ($AIC = 7176.633$) and the model with a random intercept and slope ($AIC = 7141.282$). The regression estimate for "Prev OI" is negative, which indicates that the CD4 cells count is less for patients who have a previous diagnosis of AIDS, as compared to those who did not have a previous diagnosis of AIDS at the study entry.

Table 2.3: Summary of Fitted Model (2.27) with Random Intercept and Slope, and with Covariates

Full LME Model	AIC	BIC	logLik		
with Covariates	7020.004	7072.439	-3500.002		
Random Effect	StdDev	Corr			
Intercept	4.0029	(Intr)			
obstime	0.1726	-0.18			
Residuals	1.7496				
Fixed Effect	Estimate	Std. Error	DF	T-Value	P-value
Intercept	10.3869	0.6886	937	15.0836	0.0000
obstime	-0.1524	0.1514	937	-10.0686	0.0000
Drug	0.4544	0.3803	462	1.1948	0.2328
Gender	-0.3154	0.6527	462	-0.4833	0.6291
PrevOI	-4.62561	0.4787	462	-9.6633	0.0000
AZT	-0.2570	0.4725	462	-0.5439	0.5868

In agreement with our previous models, the effect of observation time (obstime) is significant (p-value = 0) and has a negative estimate. We conclude, as the observation time increases by month, the CD4 cell count decreases by 0.1524. In addition, the treatment effect was estimates as 0.4544 and found to be insignificant (p-value=0.2328). Nevertheless, the CD4 cell count was higher in the ddl group, as compared to the ddC group. Meanwhile,

males had lower CD4 cell counts than women by 0.3154. Finally, patients who had experienced a disease progression despite AZT therapy had lower CD4 cell counts, as compared to patients who were intolerant to AZT.

Chapter 3

Analysis of Survival Data

3.1 Introduction

Survival analysis arises in many applied fields such as medicine, economics, engineering, epidemiology, and biology. As illustrated in Section 1.2, survival time is defined as a time to event of interest such as death, relapse of disease, unemployment, and completion of a task. We also know that the main characteristic of a survival time is that may be censored for some individual due to the end of a study or withdrawal during the period of the study. In this case, we do not have exact survival times for those who are still alive at the end of the study, or who are lost to follow-up during the period of the study. In this dissertation, we consider the most popular censoring mechanism (right censoring), in which the lower limit of the exact survival time is observed.

Many statistical methods have been developed for analyzing survival data. To start with, non-parametric statistical inference is often used to estimate the survivor function and compare survival curves between two or more groups. For example, the Kaplan-Meier (KM) estimator (Kaplan and Meier, 1958 [30]) for a survivor function and log-rank test (Peto, 1972 [54]) for comparison of survivor functions are derived by a nonparametric approach. However, if the appropriate distribution for survival data is pre-specified or assumed, the parametric

approach is more applicable. When the main interest is the association of survival time with various risk factors, the most popular model is the Cox regression (Cox, 1972 [8]) model based on a semiparametric approach, since the baseline hazard function is unspecified, while the effect of predictors on the hazard rate is parametrically specified. Other parametric approaches are also available under the assumed survival distributions, such as the Weibull proportional hazard model, accelerated failure time (AFT) model, and proportional odds model. In situations where survival times are repeated or not independent within individuals, the inclusion of a random effect in the survival model is essential; often referred to as the frailty model. In the following sections, we present a brief review of commonly used nonparametric, semiparametric, and parametric survival models and methods.

3.2 Notation

Let s_i be the true survival time of individual i , subject to right censoring. Let c_i be the censoring time for individual i . Here we assume that the censoring mechanism is non-informative or random. Let δ_i be the event indicator, denoted by $\delta_i = \mathbb{1}(s_i \leq c_i)$, with $\delta_i = 0$ when the event time is right censored and $\delta_i = 1$ otherwise. Let $t_i = \min(s_i, c_i)$ be the observed time. We assume the conditional independence of the random variables $\{t_i, c_i\}$ given the covariate information. Also, let $\mathbf{w}_i = (w_{i1}, \dots, w_{ip})$ be a $p \times 1$ vector of covariates from subject i . Finally, we denote the observed survival data by $\{(t_i, \delta_i), i = 1, 2, \dots, N\}$. For model parameters, we denote $\boldsymbol{\gamma}$ as a vector of regression coefficients in a survival model.

3.2.1 Non-parametric Methods

As previously mentioned, the survivor and hazard functions play important roles in survival analysis. The first step for analyzing survival data is to display a numerical or graphical summary of survival times through estimates of the survivor or hazard function. Particularly, for a sample of survival times, where none of the observations are censored, the survivor

function $S(t)$ in (1.1) can be estimated by the empirical survivor function

$$\hat{S}(t) = \frac{\text{Number of participants with survival time} \geq t}{\text{Total number of participants}}. \quad (3.1)$$

Given observed data $t_{(1)} < t_{(2)} < \dots < t_{(r)}$, in which there is censoring, the survivor function $S(t)$ can be estimated by the following nonparametric estimator $\hat{S}(t)$, called the Kaplan-Meier estimator,

$$\hat{S}(t) = \prod_{j=1}^k \left(\frac{n_j - d_j}{n_j} \right), \quad \text{for } t_{(k)} \leq t < t_{(k+1)}, \quad k = 1, 2, \dots, r, \quad (3.2)$$

where n_j is the number of individuals who have not experienced any events before time $t_{(j)}$, and d_j the number of individuals who experience the event at time $t_{(j)}$. If one wishes to compare the survival times obtained from two groups of individuals, such as a treatment group and a control group, two commonly used tests are the (weighted) log-rank and Wilcoxon tests. A detailed discussion of these tests can be found in Collett (2014) [9]. The Kaplan-Meier estimate of the survivor function, the log-rank test and Wilcoxon test for survival differences are referred to as *non-parametric* or *distribution free methods*, since they do not require any assumptions about the distribution of survival times (Collett, 2014 [9]).

3.2.2 Semi-parametric Model

The non-parametric methods described above are used to deal with the simplest case of **i.i.d** survival data. However, through modeling approaches to the analysis of survival data, we can explore how the survival times of a group of individuals or subjects depend on the values of one or more explanatory variables. For instance, individuals in epidemiological studies may differ by demographic variables such as age, gender, level of education, blood pressure, smoking habits, physical activity level, nutrition, heart rate, and so forth. Some of these or all covariates may have an impact on the time that an individual survives. These risk

factors can be fixed variables or they may change over time, in which case, they are called time-dependent variables. The basic model used for this type of survival data was proposed by Cox (1972, [13]) known as *proportional hazards (PH) model* or *Cox regression model*.

This survival model is often used to describe the hazard function at a particular time that depends on *p-explanatory variables*, represented by the vector $\mathbf{w}_i = (w_1, \dots, w_p)'$. In particular, it allows the baseline hazard function, $h_0(t)$, for an individual for whom the values of the vector $\mathbf{w}_i$ of all explanatory variables are zero, to be multiplied by $\exp(\mathbf{w}_i' \boldsymbol{\gamma})$, which is a function of the vector of explanatory variables for the i^{th} individual, called the *relative hazard*. The main idea behind it is the separation of the age or time effect in the baseline hazard function on one side and the effect of the covariates in an exponential term on the other side, that is,

$$h_i(t) = h_0(t) \exp(\mathbf{w}_i' \boldsymbol{\gamma}), \quad t \geq 0. \quad (3.3)$$

Here $h_0(t)$ is an unspecified and non-negative baseline hazard function, (i.e., the hazard function for a theoretical subject with $\mathbf{w}_i = (w_1, \dots, w_p)' = \mathbf{0}$), and $\boldsymbol{\gamma}$ denotes the vector of regression coefficients. This model combines non-parametric functional form in the baseline hazard (often a step-wise function) with parametric function in the covariate part, therefore, it has a semiparametric character. It can be re-expressed in the form

$$\log \left\{ \frac{h_i(t)}{h_0(t)} \right\} = w_{1i} \gamma_1 + w_{2i} \gamma_2 + \dots + w_{pi} \gamma_p.$$

Thus the proportional hazards model may also be regarded as a linear model for the logarithm of the hazard ratio. In the Cox-proportional hazards model, the hazard of an individual with some covariates is proportional to a baseline function of time, where the baseline hazard function has no specified form. In this case, no particular form of a probability distribution is assumed for the survival times. As a result, the PH survival model (3.3) is referred to as a *semi-parametric* model. As for statistical inference for the Cox estimator, Cox (1975) suggests using a partial likelihood approach in the case of arbitrary baseline hazard. The

partial likelihood function is given by

$$L_p(\gamma) = \prod_{i=1}^N \left\{ \frac{\exp(\mathbf{w}'_i \gamma)}{\sum_{l \in R(t_i)} \exp(\mathbf{w}'_l \gamma)} \right\}^{\delta_i}, \quad (3.4)$$

where $R(t_i)$ is the set of individuals who are at risk for experiencing the event at a time t_i .

The corresponding log-likelihood function is obtained as

$$\log L_p(\gamma) = \sum_{i=1}^N \delta_i \left\{ \mathbf{w}'_i \gamma - \log \sum_{l \in R(t_i)} \exp(\mathbf{w}'_l \gamma) \right\}. \quad (3.5)$$

In particular, the maximum partial likelihood estimators of γ are obtained by solving the estimation equation

$$\frac{\partial \log L_p(\gamma)}{\partial \gamma} = \sum_{i=1}^N \delta_i \left\{ \mathbf{w}_i - \frac{\sum_{l \in R(t_i)} \mathbf{w}_l \exp(\mathbf{w}'_l \gamma)}{\sum_{l \in R(t_i)} \exp(\mathbf{w}'_l \gamma)} \right\} = \mathbf{0}. \quad (3.6)$$

It can be shown that the maximum likelihood estimator $\hat{\gamma}$ is unbiased, efficient, and asymptotically normally distributed, so if $N \rightarrow \infty$ then $\hat{\gamma} \sim N(\gamma, [E\{I(\gamma)\}]^{-1})$, where $I(\gamma)$ is the Fisher information matrix. Detailed discussion of Cox model and its extended form can be found in Therneau and Grambsch (2000) [76] and Collett (2014) [9].

3.2.3 Parametric Model

As discussed earlier, there is no need to assume a particular form of the probability distribution for the survival times when we use the Cox regression model. However, in some applications, it may be reasonable to assume that the survival data follow a parametric distribution such as Exponential, Weibull, Gompertz, Log-logistics, Log-normal, or Gamma distribution (Wienke, 2010 [85]). In this case, a parametric PH model can make more precise inferences than a Cox PH model. Models in which the survival times are parametrically specified are known as *parametric survival models* such as the proportional hazards model

(PH), accelerated failure time model (AFT), and the proportional odds model.

3.2.3.1 Likelihood Function for Right-Censored Data

Consider a random sample of n right censored time-to-event data, $\{(t_i, \delta_i), i = 1, 2, \dots, N\}$. Let $S(t_i|\boldsymbol{\theta})$ denote the parametric form of the survivor function for the i^{th} subject, and let the survival time t_i has the probability density function, $f(t_i|\boldsymbol{\theta})$, and the cumulative distribution function $F(t_i|\boldsymbol{\theta})$. To make inference about unknown parameters $\boldsymbol{\theta}$, the information from censored data and event data is combined through the overall likelihood function

$$L(\boldsymbol{\theta}) = \prod_{i=1}^N \{f(t_i|\boldsymbol{\theta})\}^{\delta_i} \{S(t_i|\boldsymbol{\theta})\}^{1-\delta_i}. \quad (3.7)$$

An alternative expression for the above likelihood function can be obtained by considering the relationships among the survivor function, hazard function and density function, given by

$$h(t_i|\boldsymbol{\theta}) = \frac{f(t_i|\boldsymbol{\theta})}{S(t_i|\boldsymbol{\theta})} \Rightarrow h(t_i|\boldsymbol{\theta}) = f(t_i|\boldsymbol{\theta})S(t_i|\boldsymbol{\theta}).$$

We can also show that the cumulative hazard is obtained as

$$H(t_i|\boldsymbol{\theta}) = -\log\{S(t_i|\boldsymbol{\theta})\} \Rightarrow S(t_i|\boldsymbol{\theta}) = \exp\{-H(t_i|\boldsymbol{\theta})\} = \exp\left\{-\int_0^{t_i} h(s|\boldsymbol{\theta})ds\right\}.$$

So the likelihood function can be written in terms of the hazard function as

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^N \{h(t_i|\boldsymbol{\theta})S(t_i|\boldsymbol{\theta})\}^{\delta_i} \{S(t_i|\boldsymbol{\theta})\}^{1-\delta_i} \\ &= \prod_{i=1}^N \{h(t_i|\boldsymbol{\theta})\}^{\delta_i} \exp\{-H(t_i|\boldsymbol{\theta})\} \\ &= \prod_{i=1}^N \{h(t_i|\boldsymbol{\theta})\}^{\delta_i} \exp\left\{-\int_0^{t_i} h(s|\boldsymbol{\theta})ds\right\}. \end{aligned} \quad (3.8)$$

The corresponding log-likelihood is given by

$$\ell(\boldsymbol{\theta}|t_i, \delta_i) = \sum_{i=1}^N \delta_i \log h(t_i|\boldsymbol{\theta}) + \sum_{i=1}^N \int_0^{t_i} h(s|\boldsymbol{\theta}) ds. \quad (3.9)$$

We can obtain the maximum likelihood estimates of $\boldsymbol{\theta}$ by solving the score function $\frac{\partial \ell(t_i, \delta_i|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$. Below we briefly discuss the properties and theoretical basis of a *Weibull PH model* with right censored data and statistical inference for this model based on the likelihood method.

3.2.4 The Weibull Proportional Hazards Model

In modelling survival data, the parametric model based on Weibull distribution is widely used, similar to the normal distribution used in linear regression, for the following reasons: (i) its hazard function can take a variety of forms, which offers much flexibility in modeling survival data; (ii) since survival data are often skewed, descriptive statistics such as the median and percentiles can be easily obtained in order to summarize data rather than the mean and standard deviation; (iii) it has both the proportional hazards and accelerated failure time properties (to be described later in Section 3.2.5). The Weibull proportional hazard model is considered as a parametric survival regression model, denoted by $W(\lambda \exp(\mathbf{w}_i' \boldsymbol{\gamma}), \tau)$. The hazard function is given by

$$h_i(t) = h_0(t) \exp(\mathbf{w}_i' \boldsymbol{\gamma}), \quad i = 1, 2, \dots, N, \quad (3.10)$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_p)'$ is a vector of regression coefficients, $\mathbf{w}_i = (w_{1i}, w_{2i}, \dots, w_{pi})'$ is a vector of covariates, and $h_0(t)$ is the baseline hazard function. If the survival time has a Weibull distribution, $W(\lambda, \tau)$, with the scale parameter λ and shape parameter τ , then its hazard function is a parametric function such that $h_0(t) = \lambda \tau t^{\tau-1}$. Based on the Weibull

proportional hazard model (3.10), the hazard function becomes

$$\begin{aligned} h_i(t) &= \lambda \tau t^{\tau-1} \exp(\mathbf{w}_i' \boldsymbol{\gamma}) \\ &= \tau t^{\tau-1} [\lambda \exp(\mathbf{w}_i' \boldsymbol{\gamma})], \quad i = 1, 2, \dots, N, \end{aligned} \quad (3.11)$$

and the corresponding survivor function is given by $S_i(t) = \exp\{-\lambda t^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma})\}$. In this setting, the survival time has a Weibull distribution with the scale parameter $\lambda \exp(\mathbf{w}_i' \boldsymbol{\gamma})$ and shape parameter τ . The model has a similar form as Cox regression model; the only difference is that the form of baseline hazard function is not specified and the shape of this function is determined by the actual data (Wu and Yu, 2016 [89]; Collett, 2014 [9]).

3.2.4.1 Likelihood Inference and Computational Method

As before, consider survival data $\{(t_i, \delta_i), i = 1, 2, \dots, N\}$. The proportional hazards function is given by

$$h_i(t) = \tau t^{\tau-1} [\lambda \exp(\mathbf{w}_i' \boldsymbol{\gamma})], \quad i = 1, 2, \dots, N, \quad (3.12)$$

where $h_0(t) = \lambda \tau t^{\tau-1}$ and $\mathbf{w}_i$ are the baseline hazard and covariate vectors for the i th subject, respectively. The likelihood function is given by

$$\begin{aligned} L(\lambda, \tau, \boldsymbol{\gamma}) &= \prod_{i=1}^N \left[h_i(t_i) \right]^{\delta_i} S_i(t_i) \\ &= \prod_{i=1}^N \left[\lambda \tau t_i^{\tau-1} \exp(\mathbf{w}_i' \boldsymbol{\gamma}) \right]^{\delta_i} \exp(-\lambda t_i^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma})) \\ &= \prod_{i=1}^N \left[\tau t_i^{\tau-1} \exp(\mathbf{w}_i' \boldsymbol{\gamma}) \right]^{\delta_i} \exp(-t_i^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma})), \quad (\lambda = e^{\gamma_0}). \end{aligned} \quad (3.13)$$

The corresponding log-likelihood is given by

$$\log L(\lambda, \tau, \boldsymbol{\gamma}) = \sum_{i=1}^N \delta_i \log(\tau) + (\tau - 1) \sum_{i=1}^N \delta_i \log(t_i) + \sum_{i=1}^N \delta_i \mathbf{w}_i' \boldsymbol{\gamma} - \sum_{i=1}^N t_i^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma}).$$

The most natural estimation method for unknown parameters $\theta = (\lambda, \tau, \gamma')'$ is the Newton-Raphson method, which leads to the iterative equations

$$\theta^{(m+1)} = \theta^{(m)} + \left[-\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'} \right]^{-1} \left[\frac{\partial \log L(\theta)}{\partial \theta} \right] \Big|_{\theta^{(m)}},$$

for $m = 0, 1, 2, \dots$, starting with some initial value $\theta^{(0)}$ of θ . In the case of the Weibull proportional hazards model, the score function is obtained from the log-likelihood function as

$$U(\theta) = \begin{bmatrix} \frac{\partial \log L(\gamma)}{\partial \gamma} \\ \frac{\partial \log L(\tau)}{\partial \tau} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N \delta_i \mathbf{w}_i - t_i^\tau \mathbf{w}_i \exp(\mathbf{w}_i' \gamma) \\ \sum_{i=1}^N \frac{\delta_i}{\tau} + \sum_{i=1}^N \delta_i \log(t_i) - \sum_{i=1}^N t_i^\tau \log(t_i) \exp(\mathbf{w}_i' \gamma) \end{bmatrix},$$

and the Fisher Information (Hessian matrix) as

$$I(\theta) = \begin{bmatrix} -\frac{\partial^2 \ell}{\partial \gamma \partial \gamma'} & -\frac{\partial^2 \ell}{\partial \gamma \partial \tau'} \\ -\frac{\partial^2 \ell}{\partial \tau \partial \gamma'} & -\frac{\partial^2 \ell}{\partial \tau \partial \tau'} \end{bmatrix}.$$

For the Weibull distribution, we have

$$I(\theta) = \begin{bmatrix} \sum_{i=1}^N t_i^\tau \mathbf{w}_i \exp(\mathbf{w}_i' \gamma) \mathbf{w}_i & \sum_{i=1}^N t_i^\tau \log(t_i) \mathbf{w}_i \exp(\mathbf{w}_i' \gamma) \\ \sum_{i=1}^N t_i^\tau \log(t_i) \mathbf{w}_i \exp(\mathbf{w}_i' \gamma) & \sum_{i=1}^N \frac{\delta_i}{\tau^2} + \sum_{i=1}^N t_i^\tau \{\log(t_i)\}^2 \exp(\mathbf{w}_i' \gamma) \end{bmatrix}.$$

3.2.5 The Accelerated Failure Time (AFT) Model

As discussed earlier, we assume that the hazard ratio $h_i(t)/h_0(t)$ is assumed constant over time for both the Cox and Weibull proportional hazards models. In practice, however, the proportional hazards assumption may not be valid, so the Cox and/or Weibull PH models may not be adequate in such a situation. An alternative survival model is called the accelerated failure time (AFT) model, in which the proportional hazards assumption is not required. The general accelerated failure time model for the hazard function of the i th subject

at time t , $h_i(t)$, can be written as

$$h_i(t) = \exp(-\mathbf{w}_i' \boldsymbol{\alpha}) h_0(t \exp(-\mathbf{w}_i' \boldsymbol{\alpha})), \quad i = 1, 2, \dots, N, \quad (3.14)$$

where $\mathbf{w}_i' \boldsymbol{\alpha} = \eta_i$ is a linear component of the model called *the acceleration factor*, in which $\mathbf{w}_i$ is a vector of p -predictor variables for the i th subject. The corresponding survivor function for the i^{th} subject is given by

$$S_i(t) = S_0(t \exp(-\mathbf{w}_i' \boldsymbol{\alpha})). \quad (3.15)$$

For a parametric AFT model, the following alternative log-linear representation is widely used

$$\log t_i = \mu + \mathbf{w}_i' \boldsymbol{\alpha} + \sigma \varepsilon_i, \quad (3.16)$$

where μ and σ are the intercept and scale parameters, respectively, and the term ε_i is a random error. Parametric distributions for ε_i in log-linear formulation of the AFT model (3.16) leads to different distributions of survival times t_i , so we have particular parametric AFT models. In practice, the most commonly used parametric AFT models are based on Weibull, log-logistic, and lognormal distributions for survival times. The following table presents a summary of the survivor hazard functions of the distribution of ε_i , which lead to these three commonly used parametric AFT models for survival times t_i (Collett, 2014 [9]).

Table 3.1: Summary of Parametric Accelerated Failure Models

Distribution of t_i	Distribution of ε_i	$S_{\varepsilon_i}(\varepsilon)$	$h_{\varepsilon_i}(\varepsilon)$
Weibull	Gumbel	$\exp(-e^\varepsilon)$	e^ε
Log-logistic	logistic	$(1 + e^\varepsilon)^{-1}$	$(1 + e^{-\varepsilon})^{-1}$
Lognormal	Standard Normal	$1 - \Phi_{(\varepsilon)}$	$\frac{\exp(-\varepsilon^2/2)}{1 - \Phi_{(\varepsilon)} \sqrt{(2\pi)}}$

From the survivor and hazard functions of ε_i , the survivor and hazard functions of t_i can

be obtained as

$$S_i(t) = S_{\varepsilon_i} \left(\frac{\log t - \mu - \mathbf{w}_i' \boldsymbol{\alpha}}{\sigma} \right), \quad h_i(t) = \frac{1}{\sigma t} h_{\varepsilon_i} \left(\frac{\log t - \mu - \mathbf{w}_i' \boldsymbol{\alpha}}{\sigma} \right). \quad (3.17)$$

The log-linear form of the AFT model (3.16) is similar to the general linear model in regression analysis. Inference for the parametric AFT model can be based on the likelihood method. The likelihood is given by

$$L(\boldsymbol{\alpha}, \sigma) = \prod_{i=1}^N \{f_i(t_i)\}^{\delta_i} \{S_i(t_i)\}^{1-\delta_i}, \quad (3.18)$$

where $f_i(t_i)$ and $S_i(t_i)$ are the density function and survivor function for individual i , respectively. The likelihood score equation for estimating $(\mu, \boldsymbol{\alpha}, \sigma)$ is given by

$$\frac{\partial \log L(\mu, \boldsymbol{\alpha}, \sigma)}{\partial (\mu, \boldsymbol{\alpha}, \sigma)} = \mathbf{0}, \quad (3.19)$$

which may be solved iteratively using the Newton-Raphson method. The likelihood inference then proceeds in the usual way, in which standard asymptotic properties for the MLE holds under some regularity conditions. As the Weibull distribution can be parametrized as either a PH model or an AFT model, and it is, indeed, the only probability distribution that possesses this property, the results of fitting a Weibull model can therefore be interpreted in either framework. In the following section, we provide a brief review of how the Weibull distribution of lifetimes can be adapted in the AFT model.

3.2.5.1 The Weibull Accelerated Failure Time Model

When the survival times have Weibull distribution $W(\lambda, \tau)$ with scale parameter λ and shape parameter τ , in which the baseline hazard function is $h_0(t) = \lambda \tau t^{\tau-1}$, the hazard function

for the Weibull AFT model becomes

$$h_i(t) = \exp(-\mathbf{w}_i' \boldsymbol{\alpha}) \lambda \tau (\exp(-\mathbf{w}_i' \boldsymbol{\alpha}) t)^{\tau-1} = \lambda \tau \exp(-\tau \mathbf{w}_i' \boldsymbol{\alpha}) t^{\tau-1}. \quad (3.20)$$

In terms of the log-linear AFT model given in equation (3.16), the survivor function may be written as is defined as

$$S_i(t) = \exp \left\{ - \exp \left(\frac{\log t - \mu - \mathbf{w}_i' \boldsymbol{\alpha}}{\sigma} \right) \right\}, \quad (3.21)$$

which can be rewritten in the form

$$S_i(t) = \exp(-\lambda_i t^{1/\sigma}), \quad (3.22)$$

where $\lambda_i = \exp\{-(\mu + \mathbf{w}_i' \boldsymbol{\alpha})/\sigma\}$ is the scale parameter of the Weibull distribution and $1/\sigma$ is the shape parameter.

In contrast, if survival times have a Weibull distribution, the survivor function under the proportional hazards model, as defined earlier in Section 3.2.4, is given by

$$S_i(t) = \exp(-\lambda t^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma})), \quad (3.23)$$

where λ, τ are the parameters of the Weibull baseline hazard function $h_0(t) = \lambda \tau t^{\tau-1}$. Therefore, there is a direct correspondence between the two models (3.21) and (3.23), given by

$$\lambda = \exp(\mu/\sigma), \quad \tau = 1/\sigma, \quad \gamma_j = -\alpha_j/\sigma, \quad \text{for } j = 1, 2, \dots, p.$$

An alternative way to represent the Weibull PH model (3.11) in terms of the log-linear model (3.16) is $\log t_i = \frac{1}{\tau}(-\log \lambda - \mathbf{w}_i' \boldsymbol{\gamma} + \epsilon_i)$, where ϵ_i follows a Gumbel distribution. For further reading about the parametric PH and AFT models, see Collett (2014) [9].

3.2.6 The Extended Cox Model with Time-Dependent Covariates

So far, we have considered survival models with baseline covariates, such as sex, age at baseline, randomized treatment, etc. However, in some studies, it may also be of interest to investigate whether time-dependent covariates are associated with the risk of an event. In this situation, the effect of a time-dependent covariate becomes a function of time (Collett, 2014 [9]). That is, the log-hazard ratio in the Cox proportional hazards model is no longer constant over time, and so we no longer have the classic PH model. Before we introduce the extended Cox model (also known as the Andersen-Gill model), we should distinguish between two types of variables or covariates that change over time, namely, *external* or *exogenous* covariates and *internal* or *endogenous* covariates. The endogenous covariates require special treatment compared to exogenous ones. The exogenous time-dependent covariates are predictable process, meaning that its values will be known in advance at any future time. A simple example is the age of a subject, where once the age of a subject is known at origin time, the subject's age will be known exactly at any future time. On the other hand, endogenous time-dependent covariates are related to a particular individual in a study, and its values can only be measured while the subject is still alive. Such data typically arise when repeated measurements are taken on an individual. Examples include biomarkers and clinical parameters, such as CD4 cell counts for HIV-infected patients, the prothrombin index for patients with liver cirrhosis, and aortic gradient level for patients with aortic stenosis (Rizopoulos, 2012 [62]).

The Cox model (3.10) may be extended to handle *exogenous* time-dependent covariates using the counting process formulation, investigated in detail by Andersen and Gill (1982) [2] and Therneau and Grambsch (2000) [77]. The main idea behind the extended Cox model lies in the consideration of a counting process model, where a subject is assigned to the risk set for an event as long as this subject is under observation at the occurrence of event time. In counting process notation, the event process for the i^{th} subject is written as $\{N_i(t), R_i(t)\}$, where $N_i(t)$ denotes the number of events for the i^{th} subject by time t , and $R_i(t)$ is a left

continuous risk process, so that (Rizopoulos, 2012 [62])

$$R_i(t) = \begin{cases} 1, & \text{if subject } i \text{ is at risk at time } t \\ 0, & \text{if subject } i \text{ is not at risk at time } t \end{cases}$$

To introduce the extended Cox model, let $y_i(t)$ denote the covariate vector at time t for subject i , and $\mathcal{Y}_i(t) = \{y_i(s), 0 \leq s < t\}$ denote the intermittently observed (and therefore incomplete) covariate history for the i^{th} subject up to time t . Then the extended Cox model is written as

$$h_i(t|\mathcal{Y}_i(t), \mathbf{w}_i) = h_0(t)R_i(t) \exp(\mathbf{w}_i'\boldsymbol{\gamma} + \alpha y_i(t)), \quad (3.24)$$

where $\mathbf{w}_i$ is a vector of baseline covariates and the parameter α is the regression coefficient that accounts for the effect on the hazard of the time-dependent covariate at time t , $y_i(t)$, on the hazard function.

Estimation of $\boldsymbol{\gamma}$ and α is again based on the partial log-likelihood function of a time-dependent relative risk model that can be written as (Rizopoulos, 2012 [62])

$$\ell_p(\boldsymbol{\gamma}, \alpha) = \sum_{i=1}^N \int_0^\infty \left\{ R_i(t) \exp(\mathbf{w}_i'\boldsymbol{\gamma} + \alpha y_i(t)) - \log \left[\sum_j \exp(\mathbf{w}_j'\boldsymbol{\gamma} + \alpha y_j(t)) \right] \right\} dN_i(t). \quad (3.25)$$

The maximization of this partial log-likelihood function leads to asymptotically biased estimates for regression coefficients (Prentice, 1982 [56]). Unfortunately, the time-dependent Cox model (or extended Cox model) is only theoretically valid for exogenous time-varying covariates, meaning that it is not appropriate when it comes to studying endogenous covariates such as biomarkers or other patient parameters. This is because the extended Cox model assumes that the marker's level of a patient remains constant from one visit to the next and a sudden change occurs in the levels when the patients come for a visit. Such an assumption is only reasonable for covariates such as treatment dose, but leads to a crude approximation of the path of biomarkers, such as CD4 cell count, aortic gradient, and BNP.

Consequently, the accuracy of measurements taken on the endogenous processes is often contaminated by a random measurement error, which is inherent in the observed process, since the value is affected by unpredictable subject-specific features. In Chapter 4, we will introduce a modeling framework specially designed to account for the special features of endogenous or internal covariates.

3.3 The Frailty Model

In this section, we consider the inclusion of frailty to a survival model. Vaupel et al. (1979) [83] and Lancaster, (1979) [34] first introduced the notion of frailty, a random effect, to represent the unobserved population heterogeneity. Frailty is a random component designed to account for variability due to unobserved individual-level factors that are otherwise unaccounted for by the other predictors in the model. Frailty is assumed to act either multiplicatively or additively on the hazard rates of all subjects within a cluster. It is often assumed a distribution in the exponential family, such as a log-normal distribution or a Gamma distribution. The frailty model is useful for modeling the association between individual survival times within clusters. Two categories of frailty models can be pointed out. The first one is the univariate frailty model that takes into account univariate survival times, and the second one is the multivariate frailty model used for multivariate survival times. Frailty models can deal with multiple frailty terms (Duchateau and Janssen, 2007 [14]), but in this thesis, we discuss the case of univariate frailty.

When time-to-event data are observed in clusters, e.g., groups of patients treated in different hospitals, or members of the same family who share the same genetic background, the inclusion of a frailty term in the survival model may be essential (Liquet et al., 2012, [45]).

Clustered data also occurs when we are interested in studying repeated or recurrent events in the same subject (here the cluster is the subject). Examples, in clinical trials, include

the occurrence of multiple episodes of relapses from remission, recurrent heart attacks of coronary patients as well as repeated asthma attacks of asthma patients, and recurrence of bladder cancer tumours in a cohort of patients randomized to different treatment groups. In this case, the frailty captures the correlation between the multiple event times of the same subject which is unexplained by the model's covariates.

Clustered data could be also considered in more general multi-state data, such as individual transition times. Putter and van Houwelingen, (2011) [59] discussed that the inclusion of frailties in such models allows capturing both heterogeneity in the data, and residual correlation in the transition times. Thus they argued that the interpretation of the frailty effect is not always obvious and they also claimed that the identifiability problems may occur in such models due to the small number of transitions.

3.3.1 The Weibull Proportional Hazards Model with Shared Frailty

The univariate shared frailty model can be defined as

$$h_i(t|u_i) = h_0(t) \exp(\mathbf{w}_i' \boldsymbol{\gamma} + u_i), \quad (3.26)$$

where $h_i(t|u)$ is the conditional hazard function for the i^{th} subject, $\boldsymbol{\gamma}$ is a vector of unknown regression coefficients and $h_0(t)$ is the baseline hazard function; $\mathbf{w}_i$ is a vector of covariates and u_i is the random effect (the frailty) for the i^{th} subject with a density function $f(u_i)$. In the case of the Weibull PH models, the conditional hazard and survival functions for an individual are given by

$$h_i(t|u_i) = \tau t^{\tau-1} \exp(\mathbf{w}_i' \boldsymbol{\gamma} + u_i), \quad S_i(t|u_i) = \exp\{-t^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma} + u_i)\}.$$

Throughout this thesis, a Weibull baseline hazard is assumed, and the frailties u_i 's are assumed to be independent and follow a normal distribution. The likelihood function for the

frailty model is given by

$$L(\boldsymbol{\tau}, \boldsymbol{\gamma}, \sigma_u^2) = \prod_{i=1}^N \int_{u_i} \{h_0(t_i; \boldsymbol{\tau}) \exp(\mathbf{w}_i' \boldsymbol{\gamma} + u_i)\}^{\delta_i} \exp\{-H_0(t_i; \boldsymbol{\tau}) \exp(\mathbf{w}_i' \boldsymbol{\gamma} + u_i)\} f(u_i | \sigma_u^2) du_i, \quad (3.27)$$

with a normal density for the frailty, given by

$$f(u_i | \sigma_u^2) = \frac{1}{\sigma_u \sqrt{2\pi}} e^{-\frac{u_i^2}{2\sigma_u^2}}. \quad (3.28)$$

The regression parameters $\boldsymbol{\gamma}$, parameter vector $\boldsymbol{\tau}$ of the baseline hazard function $h_0(t)$, and variance of the frailty σ_u^2 can be estimated by numerically maximizing the likelihood function (3.27). We use the R function "optim" to obtain the maximum likelihood estimates.

For given data $(t_i, \delta_i, \mathbf{w}_i)$ ($i = 1, 2, \dots, N$), the likelihood function may be rewritten in the form

$$\begin{aligned} L(\boldsymbol{\gamma}, \boldsymbol{\tau}, \sigma_u^2) &= \prod_{i=1}^N \int_{u_i} \{f_i(t_i)\}^{\delta_i} \{S_i(t_i)\}^{1-\delta_i} f(u_i | \sigma_u^2) du_i \\ &= \prod_{i=1}^N \int_{u_i} \{h_i(t_i)\}^{\delta_i} S_i(t_i) f(u_i | \sigma_u^2) du_i, \end{aligned} \quad (3.29)$$

which can be evaluated using a numerical integration technique.

3.4 Simulation Study

In this study, we evaluate the performance of the maximum likelihood method for a Weibull PH model with frailty. The data were generated from the Weibull frailty model

$$h_i(t_i | u_i, \lambda, \tau, \gamma, \varphi) = \tau t_i^{\tau-1} \exp(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + u_i), \quad i = 1, \dots, N. \quad (3.30)$$

The parameters were fixed at $\tau = 0.8$ and $(\gamma_0, \gamma_1, \gamma_2) = (-6.9, 1.5, 0.5)$. The frailty term, u_i is generated from a normal distribution with mean 0 and variance $\sigma_u^2 = 0.25$. The survival

time t_i were considered as the 180-day mortality of the i^{th} subject, given by $\min(s_i, c_i)$, where $c_i = \min(c_i^*, 180)$ is a non-informative censoring time, with c_i^* being generated from the Weibull distribution with shape= 0.8 and scale= -6.5 . The data were generated for three combinations of sample sizes $N = \{100, 200, 400\}$. Each simulation run was based on 500 replicates of dataset.

The empirical properties of the ML estimators are investigated in terms of biases, mean squared errors (MSEs), coverage probabilities (CPs), and average lengths of the 95% confidence intervals. Table 3.2 presents the empirical results from the simulations.

3.4.1 Simulation Results and Discussion

Table 3.2 shows that the ML estimators of regression parameters $(\gamma_0, \gamma_1, \gamma_2)$ and τ are generally unbiased when the sample size is reasonably large (e.g., when $N = 400$). For small samples (e.g., when $N = 100$), these estimators are slightly biased. The coverage probabilities of the estimators are close to the nominal 95% confidence levels. Also, as expected, the average lengths of confidence intervals decrease as the sample size increases.

Table 3.2: Empirical biases, mean squared errors (MSEs), coverage probabilities (CPs) and average lengths (Ave.Length) of 95% confidence intervals for the maximum likelihood estimators of Weibull PH models with frailty.

Parameter	True value	Bias	MSE	CP	Ave.Length
$N = 100$					
γ_0	-6.9	-0.45902	1.19910	96	4.7189
γ_1	1.5	0.10488	0.23191	97	1.9350
γ_2	0.5	0.01314	0.02204	95	0.5907
τ	0.8	0.07486	0.02528	95	0.6299
σ_u^2	0.25	0.74950	0.81520	100	8.8363
$N = 200$					
γ_0	-6.9	-0.19818	0.48324	98	3.2867
γ_1	1.5	0.02399	0.10773	97	1.3347
γ_2	0.5	-0.00611	0.00949	96	0.4040
τ	0.8	0.04211	0.01084	96	0.4357
σ_u^2	0.25	0.50638	0.53818	100	6.6629
$N = 400$					
γ_0	-6.9	-0.19262	0.30516	98	2.7795
γ_1	1.5	0.05037	0.06584	97	1.1004
γ_2	0.5	-0.00025	0.00703	97	0.3368
τ	0.8	0.03430	0.00663	99	0.3618
σ_u^2	0.25	0.46465	0.48082	100	5.7702

3.5 Application on HIV/AIDS Study

Recall the AIDS study introduced in Chapter 2. Here we consider different survival models to explain how the risk, or hazard, of death occurring at a given time is affected by covariates of interest in the study area. Specifically, we study a variety of survival models to predict time to death of the HIV-infected patients.

To start with, we present the number of patients at risk for each treatment at each observed time point (0, 2, 6, 12, 18 months) in Table 3.3.

Table 3.3: Number of patients at risk at observed time (months)

Number at risk	Time					Total
	0	2	6	12	18	
Zalcitabine(ddC)	237	186	157	123	14	717
Didanosine(ddl)	230	182	153	103	20	688
Total	467	368	310	226	34	1405

Figure 3.1 displays the Kaplan-Merier estimates of the survivor functions under the two treatment groups (ddC vs. ddl). The plot shows that the survival curves for ddC and ddl are almost the same in the beginning months (0-6). During months 6-18, the survival probability for patients who were assigned to ddl group is lower than the survival rates for the patients assigned to ddC. Therefore, we suspect that ddC was as effective as ddl in delaying disease progression and death. Most importantly, the plot of KM curves for both drugs does not indicate sufficient evidence to suggest that there is a significant difference in treatment effects. To investigate further the treatment effects, we constructed in the following sections a Cox regression model and Weibull regression model to compare the ddl and ddC treatments while controlling other covariates in the model, such as gender, prevOI, and AZT.

Figure 3.1: Kaplan-Meier survival curves for ddC (zalcitabine) and ddl (didanosine) from 0-24 months.

3.5.1 Cox Proportional Hazard Model

In this Section, we fit a Cox regression model to investigate whether the treatment and other covariates Gender, Prev OI and AZT are associated with the risk for the event time (death). Here we assumed that the hazard depends only on covariates whose value is constant during follow-up, that is, the covariates in the model are not time-varying. For modeling the survival times (18-month mortality rate) t_i of AIDS patients, we consider the following Cox-proportional hazard model

$$h_i(t) = h_0(t) \exp(\beta_1 Drug_i + \beta_2 Gender_i + \beta_3 PrevOI_i + \beta_4 AZT_i), \quad (3.31)$$

for $i = 1, 2, \dots, n$, where $h_i(\cdot)$ is the hazard function for the i^{th} patient, $h_0(t)$ is the baseline hazard, and $(\beta_1, \dots, \beta_4)$ are the regression parameters. The only assumption of this model is that the hazards ratio $\psi = \frac{h_i(t)}{h_0(t)}$ does not change over time (i.e., proportional hazards). Table 3.4 presents estimates of regression coefficients, their standard errors, Z-values and p-value for the Cox proportionals hazards model (3.31). We observe that patients in the ddl group are 1.2423 times more likely to die, as compared to the ddC group after adjusting for covariates. Moreover, males have a lower risk of death compared to females. On the other hand, patients

who experienced a failure in AZT is 1.5575 times more likely to die, as compared to the patients who are intolerant to AZT. The output shows that previous opportunistic infection of AIDS diagnosis is a significant predictor of survival outcome ($p < 0.0001$). Specifically, patients with a previous opportunistic infection of AIDS diagnosis are 3.6402 times more likely to pass away, as compared to patients with no previous diagnosis of AIDS. In short, we conclude that the risk of death is greater in female patients, in patients who were given the drug ddl, who had a previous diagnosis of AIDS at the beginning of the study, and who experienced failure to the drug AZT.

Table 3.4: Summary of Fitted Cox Proportional Hazards Model

Variable	Estimate	Exp(Est)	Std(Est)	z-value	p-value
Drug	0.2170	1.2423	0.1464	1.482	0.138
Gender	-0.3419	0.7104	0.2455	-1.393	0.164
PrevOI	1.2920	3.6402	0.2270	5.692	<0.0001
AZT	0.1575	1.1705	0.1634	0.964	0.335

3.5.2 Weibull Proportional Hazard Model

In this Section, we consider analyzing the AIDS data assuming different parametric distributions in order to find an appropriate survival model for the HIV-infected patients. For this, several survival models were compared based on the Akaike Information Criterion (AIC). Table 3.5 shows the AIC values for fitting the null model (when the model is fitted without covariate) and the full model (when the model is fitted with all covariates). It is clear from the table that the Weibull model has the smallest AIC value. We can also check the adequacy of the Weibull model with the categorical variables (drug, gender, prevOI, AZT) by stratified Kaplan-Meier curves. That is, a plot of log survival time versus $\log[-\log(\text{KM})]$ will show linear and parallel lines if the model is adequate ([32]). Compare to other covariates, Figure 3.2 presents the Weibull regression diagnostic plot and shows that the lines for patients with ddI and ddC are generally parallel and linear in its scale.

Table 3.5: AIC values for different parametric survival models.

Survival models	Null model		Full model	
	Log likelihood	AIC	Log likelihood	AIC
Exponential	-835.900	1673.767	-806.300	1622.631
Weibull	-826.500	1656.912	-794.200	1600.351
Log normal	-827.800	1659.543	-797.00	1605.956
Log logistic	-826.400	1656.825	-794.600	1601.188

Figure 3.2: Weibull regression diagnostic for covariates (drug, gender, prevOI, AZT).

For the regression analysis, it is essential to include statistically important covariates in the model. Here we select important covariates based on Chi-square statistics and associated p-values. From Table 3.6, it appears that AZT is the least important variable and previous opportunistic infection prevOI is the most important one. We also include the covariate drug in the model as a confounding factor in the area study.

Table 3.6: Summary of Chi-square statistics and associated p-values for different covariates

Factor	Chi-sq	df	p-value
Drug	2.11	1	0.1465
Gender	2.10	1	0.1468
PrevOI	29.83	1	< 0.0001
AZT	0.93	1	0.3354
TOTAL	42.44	4	< 0.0001

So we fit the Weibull proportional hazards model

$$\begin{aligned}
 h_i(t_i|\lambda, \tau, \gamma) &= \lambda \tau t_i^{\tau-1} \exp(\gamma_1 Drug_i + \gamma_2 PrevOI_i) \\
 &= \tau t_i^{\tau-1} \exp(\gamma_0 + \gamma_1 Drug_i + \gamma_2 PrevOI_i).
 \end{aligned} \tag{3.32}$$

As can be observed from the fitted Weibull PH model in Table (3.7), the treatment (drug) has a positive effect on the hazard rate of the survival time. However, the hazard rate is higher for prevOI in comparison to other covariates. Further, the estimated shape parameter ($\hat{\tau} = 0.6945 < 1$) shows that the death rate from HIV infection was decreasing over time.

Table 3.7: The results of fitted Weibull proportional hazards (PH) model

Parameter	Survival Model				
	Estimate	Stderr	z-value	LL	UL
Intercept (γ_0)	-7.1668	0.9869	-7.2623	-9.1011	-5.2326
Drug (γ_1)	0.3516	0.2887	1.2181	-0.2142	0.9175
PrevOI (γ_2)	1.4939	0.4362	3.4252	0.6391	2.3488
Shape (τ)	0.6945	0.0948	7.3283	0.5088	0.8803

To assess the validity of the Weibull model used in the survival analysis, we performed a residual analysis based on the widely used Cox-Snell residuals, $r_{C_i} = -\log \hat{S}_i(t_i)$, where $\hat{S}_i(t_i)$ is an estimate of the survivor function, given by

$$\hat{S}_i(t_i) = \exp \left\{ -\hat{t}_i^{\hat{\tau}} \exp(\hat{\gamma}_0 + \hat{\gamma}_1 Drug_i + \hat{\gamma}_2 PrevOI_i) \right\}.$$

As we know, the Cox-Snell residuals follow a unit exponential distribution, if the fitted model is correct. Thus we assess the overall fit of the model by determining whether the Cox-Snell residuals are a plausible sample from a unit exponential distribution. Figure 3.3 displays a plot in which the cumulative hazards $H_i(r_{C_i})$ of the residuals are plotted against the residuals r_{C_i} . The plot exhibits a straight line pattern that would not cast doubt on the validity of the Weibull model, as considered here.

Figure 3.3: Residual analysis of the survival model with Weibull distribution for AIDS data.

3.5.3 Weibull Frailty Model

To study the heterogeneity in HIV patients, here we fit the Weibull frailty model

$$h_i(t_i|u_i, \lambda, \tau, \gamma) = \tau t_i^{\tau-1} \exp(\gamma_0 + \gamma_1 Drug_i + \gamma_2 PrevOI_i + u_i). \quad (3.33)$$

The frailty u_i is assumed to follow a normal distribution with mean 0 and variance σ_u^2 . Table 3.8 presents ML estimates of the model parameters along with their standard errors, z-values, and lower limits (LL) and upper limits (UL) of the 95% confidence intervals. It appears that the drug is not a significant predictor of survival times, and an increase in survival times is associated with an increase in previous opportunistic infection of HIV (prevOI).

Table 3.8: Weibull Model with Frailty (3.33) for AIDS Data					
Parameter	Frailty Survival Model				
	Estimate	Stderr	z-value	LL	UL
Intercept (γ_0)	-5.8002	0.7367	-7.8732	-7.2441	-4.3563
Drug (γ_1)	0.2163	0.1647	1.3133	-0.1065	0.5391
PrevOI (γ_2)	1.4096	0.2672	5.2755	0.8859	1.9333
Shape (τ)	1.4461	0.1711	8.4518	1.1107	1.7815
Random effects					
Frailty (σ_u^2)	0.1213	0.6012	0.2018	-1.0571	1.2997

The value of the shape parameter in the Weibull frailty model is $\hat{\tau} = 1.4461$, which is more than unity, indicating that the risk of death among patients was increasing over time. Furthermore, there is no significant heterogeneity among the HIV patients ($\hat{\sigma}_u^2 = 0.1213$; $95\%CI = [-1.0571, 1.2997]$). However, it appears that the Wald statistic does not work well for testing the parameter values on the boundary, e.g. $H_0 : \sigma_u^2 = 0$ as it gives a confidence interval containing a negative lower limit. We shall consider using the LRT for assessing the variance components in a future study.

Chapter 4

Joint Likelihood Analysis of Longitudinal and Survival Data

4.1 Introduction

As discussed in Section 1.3, longitudinal and survival data are often collected together. As an example, consider the CD4 count data from AIDs patients studied by Abrams (1994)[1], illustrated in Chapter 2, in which the goal is to determine whether CD4 count has a significant confounding effect with the active treatment on survival time. In this study, the relationship between longitudinal covariate processes and time-to-death is the primary research interest. Although the methodologies are well established individually for longitudinal data and survival data, the separate models discussed in Chapter 2 are insufficient for our research question. Specifically, the extended Cox model can be used to study the association between the longitudinal measurements and the survival process; the relevant literature is reviewed in Section 3.2.6. However, it is not appropriate, especially for internal or endogenous time-dependent covariates because it leads to potentially biased parameter estimation and misleading statistical inferences (Wulfsohn and Tsiatis, 1997 [90]; Tsiatis, 1995 [81]).

Consequently, new modeling frameworks are required to address two issues: (i) the longitudinal covariates collected for each individual are usually contaminated by measurement errors; (ii) the complete knowledge of a covariate history for all individuals are unavailable due to the intermittent measurements and the occurrence of events or censoring. In order to meet such practical needs, joint modeling techniques have been developed in the past two decades as a powerful statistical tool to fit the two submodels simultaneously via their shared information. In comparison to the separate estimation procedure, joint modeling greatly improves the model estimation and provides more reliable inferences by taking the longitudinal-survival relationship into account.

A typical approach in joint modeling literature is to consider a survival submodel with measurement errors in time-varying covariates (Ibrahim ,2010 [28]; Wu and Liu, 2012 [88]; Wulfsohn and Tsiatis, 1997 [90]). In this approach, a linear mixed-effects model is often used to model time-dependent covariates to address covariate measurement errors, and a Cox proportional hazards model is used for modelling the survival data (Wu and Liu, 2012 [88]). In this chapter, we consider a standard formulation of the joint model where a single longitudinal process is assessed by a linear mixed-effects model and a survival process by a Weibull frailty model and linking both processes using shared random effects.

This chapter is organized as follows. In Section 4.3, we present the joint model and discuss estimation method including optimization and numerical integration algorithms, as well as inference for the regression coefficients and random effects. In Section 4.5, we present results from a comprehensive simulation study, which was carried out to evaluate the performance of the joint model for analyzing longitudinal and survival data and to compare the results to those obtained from separate analyses. In Section 4.6, we re-analyze the AIDS data, illustrated in Chapter 2, using joint modelling.

4.2 Notation

In this chapter, we consider univariate longitudinal data of possibly different types. For simplicity, we focus on the continuous longitudinal variable. We first define some notation that will be used throughout the chapter.

For longitudinal data, let Y be a continuous response variable, with its observed values of y_{ij} for individual i at time t_{ij} , $i = 1, 2, \dots, N, j = 1, 2, \dots, n_i$. In addition, the matrix $\mathbf{X}_i$ may contain other longitudinal data, which may be used as covariates but are not of primary interest.

For survival data, let s_i be the true survival time of individual i , subject to right censoring. Let c_i be the censoring time for individual i . Here we assume that the censoring mechanism is non-informative or random. Let δ_i be the event indicator, denoted by $\delta_i = \mathbb{1}(s_i \leq c_i)$, with $\delta_i = 0$ when the event time is right censored, and $\delta_i = 1$, otherwise. Let $t_i = \min(s_i, c_i)$ be the observed time. We denote the observed survival data by $\{(t_i, \delta_i), i = 1, 2, \dots, N\}$.

For model parameters, we denote $\boldsymbol{\beta}$ as a vector of regression parameters in a longitudinal model and $\boldsymbol{\gamma}$ as a vector of regression parameters in a survival model. We also denote $\mathbf{u}_i$ as a vector of random effects for individual i and $\boldsymbol{\varepsilon}_i$ as a vector of random errors. As for variance-covariance parameters, we follow the same assumption and notation as introduced in Section 2.4.2.1. We denote $\boldsymbol{\theta}$ as a collection of all unknown parameters in the models.

4.3 The Basic Joint Model

The main purpose of a joint model is to link two processes: a longitudinal process (such as biomarker measurements) and an event process (such as time-to-event data), using a shared latent structure, as explained in Figure 4.1.

In the literature, two types of joint models have been developed: joint models with *shared latent classes* and joint models with *shared random effects* (Diggle et al., [25]). In the former, the population of subjects is assumed to be heterogeneous, divided into homogeneous latent

subgroups of subjects that share the same marker trajectory of the longitudinal process and the same risk of the event (Lin, 2002 [43]; Proust-Lima, 2014 [57]). In the latter, the population of subjects is assumed to be heterogeneous and then linking both the longitudinal and event processes by a function of the individual random effects that also illustrate the variability in the longitudinal trajectories over subjects (Tsiatis and Davilain, 2004 [79]; Wulfsohn and Tsiatis, 1997 [90]). In this thesis, we focus in joint models with shared random effects.

Figure 4.1: Graphical general representation of a joint model for a longitudinal process and an event process. Ovals represent latent variables while squares represent observed variables.

Also, in the literature, two different joint modeling approaches that used shared random effects to their parametrization were found. First, Rizopoulos (2010) [63] proposed a joint model, where the main interest is in the time-to-event process, which is influenced by a longitudinal time-dependent covariate measured with error. Later, Philipson et al (2012) [55] developed a shared random effects model, where the main focus is on both survival and longitudinal processes. In this chapter, our primary objective is to study the subject's survival (or survival process), so we consider the joint approach proposed by Rizopoulos (2010) [63] and Rizopoulos (2012) [62]. Here we would like to study a likelihood-based approach for the joint modelling of longitudinal and survival data, where the longitudinal process is assessed by a linear mixed-effects model and the survival process by a Weibull

parametric hazards model. Sousa (2011) [73]) and McCrink et al.(2013) [47]) give a good overview on selection and pattern-mixture models.

In the following sections, we briefly describe the motivation of the joint modelling approach. We introduce three factors that need to be considered in developing a joint model: the longitudinal submodel, the survival submodel, and the connection between the two submodels. The submodels are specified with regard to Rizopoulos's proposal (Rizopoulos, 2012 [62]). Particularly, the joint model consists of two linked submodels, known as the longitudinal submodel, and the survival submodel. In order to introduce this methodology, we will use the same notation as introduced in Section 4.2 with minor changes, as shown below:

- Let t_i be the true event time (survival time), c_i the censoring time, and $\delta_i = \mathbb{1}(s_i \leq c_i)$ the dichotomous event indicator.
- Let $y_i(t)$ be the observed value at time point t of the endogenous time-dependent covariate (such as CD4 cell count), in which measurements were taken only at the very specific occasions t_{ij} , equivalently, $y_{ij} = \{y_i(t_{ij}), j = 1, 2, \dots, n_i\}$.
- Let $m_i(t)$ be the *true* and *unobserved* value of the respective longitudinal outcome at time t , which is uncontaminated with any measurement error.

4.3.1 The Survival Submodel

Here the goal is to measure the association between the longitudinal marker $m_i(t)$ and the hazard of an event at time t . The natural approach is to postulate a proportional hazards model of the form:

$$\begin{aligned} h_i(\mathcal{M}_i(t), \mathbf{w}_i) &= \left\{ \lim_{\delta \rightarrow 0} \frac{P[t \leq T_i < t + \delta t | T_i \geq t, \mathcal{M}_i(t), \mathbf{w}_i]}{\delta t} \right\} \\ &= h_0(t) \exp\{\mathbf{w}_i' \boldsymbol{\gamma} + \varphi m_i(t)\}, \quad t > 0. \end{aligned} \tag{4.1}$$

Here $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$ indicates the history of the true unobserved values of longitudinal covariate up to time point t , $h_0(t)$ indicates the baseline risk function, and $\mathbf{w}_i$ is a vector of baseline covariates (such as a treatment indicator) with a corresponding vector of regression coefficients $\boldsymbol{\gamma}$. In model (4.1), the parameter φ measures the impact of the underlying longitudinal outcome to the risk for an event. Moreover, note that in the relative risk model (4.1) we assume that the risk for an event at time point t depends on the current value of the endogenous time-dependent covariate $m_i(t)$, whereas the extended Cox model (3.24) postulates that the risk for an event depends on the current value of the endogenous time-dependent covariates $y_i(t)$. As for the survivor function, however, this does not hold. The corresponding survivor function depends on the whole history of the true unobserved longitudinal process up to time point t , $\mathcal{M}_i(t)$. That is,

$$\begin{aligned} S_i(t|\mathcal{M}_i(t), \mathbf{w}_i) &= P(T_i > t|\mathcal{M}_i(t), \mathbf{w}_i) \\ &= \exp \left\{ - \int_0^t \lambda_0(s) \exp\{\mathbf{w}_i' \boldsymbol{\gamma} + \varphi m_i(s)\} ds \right\}. \end{aligned} \quad (4.2)$$

Note that both the hazard and survivor functions are written as a function of the baseline hazard $\lambda_0(t)$. In the PH literature, the function $\lambda_0(t)$ is usually recommended to remain completely unspecified in order to avoid the impact of misspecifying the distribution of survival times. In the joint modelling framework, however, it can lead to an underestimation of the standard errors of the parameter estimates (Hsieh, 2006 [27]). It is advisable to use parametric but flexible models for $\lambda_0(t)$ (Rizopoulos, 2012 [62]). We highlight a few useful models for the baseline hazard $\lambda_0(t)$ below.

(i) The *piecewise-constant* model, given by

$$\lambda_0(t) = \sum_{q=1}^Q \xi_q \mathbb{1}(\nu_{q-1} < t \leq \nu_q),$$

where $0 = \nu_0 < \nu_1 < \dots < \nu_Q$ denotes a partition of the time scale, with ν_Q being

larger than the largest observed time, and ν_q denotes the value of the hazard in the interval $(\nu_{q-1}, \nu_q]$.

(ii) The *regression splines* model, given by

$$\log \lambda_0(t) = \kappa_0 + \sum_{q=1}^Q \kappa_q B_q(t, \nu),$$

where $\boldsymbol{\kappa}^t$ is a vector spline coefficients, and $B_q(t, \nu)$ denotes the q^{th} basis function of B-splines basis functions with knots $\nu_0, \nu_1, \dots, \nu_Q$. However, to avoid the issue of underestimation of the standard errors of the parameter estimates, we may need to use a risk function corresponding to a known parametric distribution, such as the Weibull, log-normal, or the Gamma distribution. In this chapter we study the Weibull model for $\lambda_0(t)$.

(iii) The *Weibull* model, where the survival times follow a Weibull distribution, $W(\lambda, \tau)$, with the scale parameter λ and shape parameter τ . Then the hazard function is obtained as $h(t) = \lambda \tau t^{\tau-1}$ for $0 \leq t < \infty$. If $\tau = 1$, the survival times have an exponential distribution as a special case of the Weibull distribution. For other values of τ , the hazard function increases or decreases monotonously for $\tau > 1$ and $\tau < 1$, respectively. For this particular choice of hazard function with the distribution $W(\lambda \exp(\mathbf{w}'_i \boldsymbol{\gamma}), \tau)$, the survival submodel (4.1) can be written as

$$\begin{aligned} h_i(\mathcal{M}_i(t), \mathbf{w}_i) &= h_0(t) \exp\{\mathbf{w}'_i \boldsymbol{\gamma} + \varphi m_i(t)\} \\ &= \tau t^{\tau-1} \exp\{\lambda \mathbf{w}'_i \boldsymbol{\gamma} + \varphi m_i(t)\} \\ &= \tau t^{\tau-1} \exp\{\mathbf{w}'_i \boldsymbol{\gamma} + \varphi m_i(t)\}, \quad (\lambda = e^{\gamma_0}), \end{aligned} \tag{4.3}$$

where $m_i(t)$ involves random effects $\mathbf{u}_i$, which are assumed independent $N(0, \mathbf{D})$, with covariance $\mathbf{D} = \mathbf{D}(\boldsymbol{\xi})$.

Given data $\{(t_i, \delta_i); i = 1, 2, \dots, N\}$, the likelihood function for the frailty model parameters can be written as

$$\begin{aligned} L(\boldsymbol{\gamma}, \tau, \mathbf{D}, \varphi) &= \prod_{i=1}^N \int_{-\infty}^{\infty} \{f(t_i|\mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)\}^{\delta_i} \{S(t_i|\mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)\}^{\delta_i-1} f(\mathbf{u}_i|\mathbf{D}) d\mathbf{u}_i \\ &= \prod_{i=1}^N \int_{-\infty}^{\infty} \{h(t_i|\mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)\}^{\delta_i} S(t_i|\mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi) f(\mathbf{u}_i|\mathbf{D}) d\mathbf{u}_i, \end{aligned} \quad (4.4)$$

where $f(t_i|\mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)$ is the conditional density function of the survival time, given the frailty $\mathbf{u}_i$, $S(t_i|\mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)$ is the conditional survivor function for the i^{th} subject at time t_i , and $f(\mathbf{u}_i|\mathbf{D})$ is the conditional density of the random effects u_i . In this setting, the density function for the random effect u_i is given by $\mathbf{u}_i \sim N(0, \mathbf{D})$, that is,

$$f(\mathbf{u}_i|\mathbf{D}) = \frac{\exp\{-\frac{1}{2}\mathbf{u}_i'\mathbf{D}^{-1}\mathbf{u}_i\}}{(2\pi)^{q/2}|\mathbf{D}|^{1/2}}, \quad (4.5)$$

whereas the conditional density for survival times is given by the Weibull distribution

$$\begin{aligned} f(t_i|\mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi) &= \{h_i(t_i|\mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)\}^{\delta_i} S_i(t_i|\mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi) \\ &= \{\tau t_i^{\tau-1} \exp(\mathbf{w}_i'\boldsymbol{\gamma} + \varphi m_i(t))\}^{\delta_i} \exp\{-\tau t_i^{\tau-1} \exp(\mathbf{w}_i'\boldsymbol{\gamma} + \varphi m_i(t))\}. \end{aligned} \quad (4.6)$$

The corresponding log-likelihood function for the frailty submodel (4.4) is given by

$$\ell(\boldsymbol{\gamma}, \tau, \mathbf{D}, \varphi) = \sum_{i=1}^N \int_{-\infty}^{\infty} \left[\delta_i \log h_i(t_i|\mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi) + \log S_i(t_i|\mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi) \right] f(\mathbf{u}_i|\mathbf{D}) d\mathbf{u}_i.$$

Let $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \tau, \mathbf{D}, \varphi)$ denote the vector of all model parameters that need to be estimated. The ML estimators of $\boldsymbol{\theta}$ may be obtained by using a numerical maximization method. Given some initial estimates $\boldsymbol{\theta}^{(0)}$, we can obtain the ML estimates by solving the Newton-Raphson

iterative equations

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} + \{I(\boldsymbol{\theta}^{(m)})\}^{-1}\{U(\boldsymbol{\theta}^{(m)})\}, \quad (4.7)$$

for $m = 0, 1, 2, \dots$, where $U(\boldsymbol{\theta}^{(m)})$ is the likelihood score function $U(\boldsymbol{\theta})$, given by

$$U(\boldsymbol{\theta}) = \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^N \int_{-\infty}^{\infty} \left[\delta_i \frac{\partial \log h_i(t_i | \mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)}{\partial \boldsymbol{\theta}} + \frac{\partial \log S_i(t_i | \mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)}{\partial \boldsymbol{\theta}} \right] f(\mathbf{u}_i | t_i; \boldsymbol{\theta}) d\mathbf{u}_i,$$

evaluated at $\boldsymbol{\theta}^{(m)}$ and the Fisher information matrix $I(\boldsymbol{\theta})^{(m)}$ is obtained from the first derivative of the score function $U(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta}^{(m)}$. After some algebra ([66]), we can show that

$$\begin{aligned} I(\boldsymbol{\theta}) &= \sum_{i=1}^N \int_{-\infty}^{\infty} \left[\delta_i \frac{\partial^2 \log h_i(t_i | \mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} + \frac{\partial^2 \log S_i(t_i | \mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] f(\mathbf{u}_i | t_i; \boldsymbol{\theta}) d\mathbf{u}_i \\ &+ \sum_{i=1}^N \int_{-\infty}^{\infty} \left[\delta_i \frac{\partial \log h_i(t_i | \mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)}{\partial \boldsymbol{\theta}} + \frac{\partial \log S_i(t_i | \mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)}{\partial \boldsymbol{\theta}} \right] \\ &\times \left[\delta_i \frac{\partial \log h_i(t_i | \mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)}{\partial \boldsymbol{\theta}'} + \frac{\partial \log S_i(t_i | \mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)}{\partial \boldsymbol{\theta}'} \right] f(\mathbf{u}_i | t_i; \boldsymbol{\theta}) d\mathbf{u}_i \\ &- \sum_{i=1}^N \int_{-\infty}^{\infty} \left[\delta_i \frac{\partial \log h_i(t_i | \mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)}{\partial \boldsymbol{\theta}} + \frac{\partial \log S_i(t_i | \mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)}{\partial \boldsymbol{\theta}} \right] f(\mathbf{u}_i | t_i; \boldsymbol{\theta}) d\mathbf{u}_i \\ &\times \int_{-\infty}^{\infty} \left[\delta_i \frac{\partial \log h_i(t_i | \mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)}{\partial \boldsymbol{\theta}'} + \frac{\partial \log S_i(t_i | \mathbf{u}_i; \boldsymbol{\gamma}, \tau, \varphi)}{\partial \boldsymbol{\theta}'} \right] f(\mathbf{u}_i | t_i; \boldsymbol{\theta}) d\mathbf{u}_i. \end{aligned}$$

The likelihood function (4.4) does not have a closed form. Also, both the likelihood score function and Fisher information matrix involve the calculation of multi-dimensional integrations with respect to the conditional distribution of $\mathbf{u}_i | t_i$, which does not have a closed form. The integrations involving the conditional expectations may be computed numerically using an existing software. In this thesis, we use the R function "optim" to obtain the maximum likelihood (ML) estimates $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ in the Weibull frailty model.

4.3.2 The Longitudinal Submodel

We assume that the hazard function $h_i(t)$ in the above hazard model (4.1) depends on the *true unobserved* value of the longitudinal outcome $m_i(t)$ at time t (Wu and Liu, 2012 [88]). However, the longitudinal measurements y_{ij} are collected with error on each subject at times $t_{ij}; i = 1, 2, \dots, N, j = 1, 2, \dots, n_i$. Therefore, we need to estimate $m_i(t)$ and reconstruct the *true* longitudinal history $\mathcal{M}_i(t)$ for each subject in order to measure the impact of the longitudinal outcome to the hazard for an event (Rizopoulos, 2012 [62]). To accomplish this we can fit a suitable mixed-effects model with the observed repeated measurements $y_{ij} = \{y_i(t_{ij}), j = 1, 2, \dots, n_i\}$ of each subject i , in order to describe the subject-specific time evolutions. In this thesis, we assume that the longitudinal outcomes are normally distributed. The most commonly used sub-model for the longitudinal process is a linear mixed-effects (LME) model, given by

$$\begin{cases} y_i(t) = m_i(t) + \varepsilon_i(t), \\ m_i(t) = \mathbf{x}'_i(t)\boldsymbol{\beta} + \mathbf{z}'_i(t)\mathbf{u}_i, \\ \mathbf{u}_i \sim N(\mathbf{0}, \mathbf{D}), \quad \varepsilon_i(t) \sim N(0, \sigma_\varepsilon^2), \end{cases} \quad (4.8)$$

where $\boldsymbol{\beta}$ denotes the vector of the regression coefficients for the fixed effects covariates $\mathbf{x}_i$ and $\mathbf{z}_i$ denotes the covariate vector for the random effects $\mathbf{u}_i$. Particularly, the fixed and the random effects refer to the population-average and subject-specific effects, respectively. The error terms $\varepsilon_i(t)$ are mutually independent, normally distributed with variance σ_ε^2 , and independent of $\mathbf{u}_i$ (Rizopoulos, 2012 [62]). The random effects $\mathbf{u}_i$ follow a multivariate normal distribution with covariance matrix $\mathbf{D}$. The marginal mean vector and covariance matrix of the response vector $y_i(t)$ are $E(\mathbf{Y}_i) = \boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta}$ and $Var(\mathbf{Y}_i) = \mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}'_i + \sigma_\varepsilon^2\mathbf{I}_{n_i}$. For known variance components, $\boldsymbol{\alpha} = (\sigma_\varepsilon^2, \boldsymbol{\xi}')'$, the best linear unbiased estimator

(BLUE) of β and best linear unbiased predictor (BLUP) of $\mathbf{u}_i$ can be obtained as

$$\tilde{\beta}(\alpha) = \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i \right),$$

and

$$\tilde{\mathbf{u}}_i(\alpha) = \mathbf{D}_i \mathbf{Z}_i' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\beta}).$$

The empirical BLUE (EBLUE) of β and empirical BLUE (EBLUE) of $\mathbf{u}_i$ may be obtained by replacing α by a suitable estimator $\hat{\alpha}$, which leads to $\hat{\beta} = \tilde{\beta}(\hat{\alpha})$ and $\hat{\mathbf{u}}_i = \tilde{\mathbf{u}}_i(\hat{\alpha})$. In particular, the ML or restricted ML estimator of α can be used. Here the ML estimator of α may be obtained by solving the equations

$$\sum_{i=1}^N \left\{ (\mathbf{y}_i - \mathbf{X}_i \beta)' \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \alpha_k} \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta) - \text{tr} \left(\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \alpha_k} \right) \right\} = 0, \quad k = 1, 2, \dots, q,$$

where α_k is the k^{th} component of $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_q)'$. These equations are solved numerically using an iterative algorithm, known as the Newton-Raphson algorithm. Also, the variance-covariance matrix of the ML estimators of (β, α) is obtained from the Fisher information matrix.

4.3.3 Joint Model with Shared Random Effects

Recall the longitudinal submodel (4.8). For this, the survival model (4.1) may be written as

$$h_i(\mathcal{M}_i(t), \mathbf{w}_i) = h_0(t) \exp\{\mathbf{w}_i' \gamma + \varphi[\mathbf{x}_i'(t) \beta + \mathbf{z}_i'(t) \mathbf{u}_i]\}. \quad (4.9)$$

Furthermore, under the Weibull PH model, the above hazard function may be written as

$$h_i(\mathcal{M}_i(t), \mathbf{w}_i) = \tau t^{\tau-1} \exp\{\mathbf{w}_i' \gamma + \varphi[\mathbf{x}_i'(t) \gamma + \mathbf{z}_i'(t) \mathbf{u}_i]\}, \quad (\lambda = e^{\gamma_0}). \quad (4.10)$$

The classical joint model with shared random effects is decomposed into two submodels: a

linear mixed model for the longitudinal response (biomarker) $y_i(t)$ and a Weibull proportional hazards model for the survival time t , where the two processes are linked by a function of random effects $\mathbf{u}_i$. The joint model is given by

$$\begin{cases} y_i(t) = \mathbf{x}'_i(t)\boldsymbol{\beta} + \mathbf{z}'_i(t)\mathbf{u}_i + \varepsilon_i(t) \equiv m_i(t) + \varepsilon_i(t), \\ h_i(\mathcal{M}_i(t), \mathbf{w}_i) = \tau t^{\tau-1} \exp\{\mathbf{w}'_i\boldsymbol{\gamma} + \varphi[\mathbf{x}'_i(t)\boldsymbol{\beta} + \mathbf{z}'_i(t)\mathbf{u}_i]\}, \\ \mathbf{u}_i \sim N(0, \mathbf{D}), \quad \varepsilon_i(t) \sim N(0, \mathbf{R}_i), \end{cases} \quad (4.11)$$

where in the longitudinal submodel, $\mathbf{x}_i(t)$ and $\mathbf{z}_i(t)$ are vectors of possibly time-dependent covariates associated with the p-vector of fixed effects $\boldsymbol{\beta}$ and the q-vector of individual random effects $\mathbf{u}_i$, with $\mathbf{u}_i \sim N(0, \mathbf{D})$. The error terms $\varepsilon_i(t)$ and $\mathbf{u}_i$ are assumed independent.

In the survival submodel, $h_0(t) = \lambda\tau t^{\tau-1}$ is the baseline hazard when survival times follow the Weibull distribution. Usually, the baseline hazard is parametric (e.g., Weibull, piecewise constant, or a small number of B-splines). It is rare to keep $h_0(t)$ unspecified (like in the Cox model); the partial likelihood of Cox model cannot be employed and the full likelihood has to be defined. As a solution, one might consider a piecewise constant function with jumps at each event time but this would produce too many parameters and lead to computational problems (Van, 2000 [82]; Hsieh, 2006 [27]). The vector $\mathbf{w}_i$ denotes baseline covariates associated with the vector of coefficients $\boldsymbol{\gamma}$, while the multivariate function of marker $m_i(t) = \mathbf{x}'_i(t)\boldsymbol{\beta} + \mathbf{z}'_i(t)\mathbf{u}_i$ is associated with the parameter φ , which quantifies the degree of association between the longitudinal outcome evaluated at time t and the corresponding hazard for an event.

As for interpretation, this joint model framework has the advantage to keep the same type of interpretation of the linear mixed effects model for the longitudinal data and of the Weibull frailty model for the survival submodel. The quantity $\exp(\varphi)$ in the joint model (4.11) indicates the relative increase in the hazard at time t for a one-unit increase in $m_i(t)$ at time t , given covariates $\mathbf{w}_i$.

Another commonly used joint model framework, which is perhaps the most popular one, is to link the survival and longitudinal models via shared random effects, called shared-parameter models. In this case, the random effects may be interpreted as a summary of individual-specific longitudinal characteristics, or a latent process (shared variables) which governs both longitudinal and the event progressions. Such a shared parameter model may be written as

$$\begin{cases} \mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i, \\ h_i(t) = \tau t^{\tau-1} \exp\{\mathbf{w}_i'\boldsymbol{\gamma} + \boldsymbol{\varphi}'\mathbf{u}_i\}, \end{cases} \quad (4.12)$$

where the Weibull PH model and LME model share the same random effects $\mathbf{u}_i$. This joint model framework is frequently used when the survival risk is influenced by summaries of the longitudinal process (e.g., individual-specific intercepts and slopes). In this thesis, the joint model assumes a latent process based on random effects to describe the association between longitudinal and time-to-event data, formulated as (4.12). This joint model is not much different than the previous one when same set of covariates are used on both submodels. In this random-effects model, the survival and longitudinal outcomes are assumed to be conditionally independent, given a set of unobserved random effects.

Estimation of Joint Model Parameters

The maximum likelihood (ML) method is commonly used to estimate the parameters of joint models (Schluchter, 1992 [68]; Wulfsohn and Tsiatis, 1997 [90]; Henderson et al., 2000 [25]; Tsiatis and Davidian, 2004 [79]; Diggle et al., 2008 [11]; Rizopoulos, 2010 [63]). To specify the likelihood, we assume the observed longitudinal process $\{\mathbf{y}_i\}$ and the event processes $\{t_i, \delta_i\}$ are independent given the random effects $\mathbf{u}_i$ (Schluchter, 1992 [68]; Faucett and Thomas, 1996 [16]; Wulfsohn and Tsiatis, 1997 [90]; Henderson et al., 2000 [25]; Tsiatis and Davidian, 2004 [79]; Diggle et al., 2008 [11]; Rizopoulos, 2010 [63]). Given observed data $\{\mathbf{y}_i, t_i, \delta_i\} (i = 1, 2, \dots, N)$ from both survival and longitudinal outcomes, the joint likelihood

is given by

$$\begin{aligned}
L(\boldsymbol{\theta}) &= \prod_{i=1}^N f(t_i, \delta_i, \mathbf{y}_i | \boldsymbol{\theta}) \\
&= \prod_{i=1}^N \int f(t_i, \delta_i, \mathbf{y}_i, \mathbf{u}_i | \boldsymbol{\theta}) d\mathbf{u}_i \\
&= \prod_{i=1}^N \int f(t_i | \mathbf{u}_i; \gamma, \tau, \varphi) f(\mathbf{y}_i | \mathbf{u}_i; \boldsymbol{\beta}, \sigma_\epsilon^2) f(\mathbf{u}_i | \mathbf{D}) d\mathbf{u}_i \\
&= \prod_{i=1}^N \int f(t_i | \mathbf{u}_i; \gamma, \tau, \varphi) \prod_{j=1}^{n_i} f(y_{ij} | \mathbf{u}_i; \boldsymbol{\beta}, \sigma_\epsilon^2) f(\mathbf{u}_i | \mathbf{D}) d\mathbf{u}_i, \tag{4.13}
\end{aligned}$$

where $\boldsymbol{\theta}$ denotes the vector of all model parameters that need to be estimated. In this setting:

- The conditional density for the Weibull survival time t_i takes the form

$$\begin{aligned}
f(t_i | \mathbf{u}_i; \gamma, \tau, \varphi) &= \{h_i(t_i | \mathbf{u}_i; \gamma, \tau, \varphi)\}^{\delta_i} S_i(t_i | \mathbf{u}_i; \gamma, \tau, \varphi) \\
&= \{\tau t_i^{\tau-1} \exp(\mathbf{w}'_i \gamma + \boldsymbol{\varphi}' \mathbf{u}_i)\}^{\delta_i} \exp\{-t_i^{\tau-1} \exp(\mathbf{w}'_i \gamma + \boldsymbol{\varphi}' \mathbf{u}_i)\}. \tag{4.14}
\end{aligned}$$

- The conditional density for the longitudinal outcome y_{ij} is assumed to be normal, with mean $\mu_{ij} = \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{u}_i$ and variance σ_ϵ^2 . The conditional density for of $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})'$ is given by

$$\begin{aligned}
f(\mathbf{y}_i | \mathbf{u}_i; \boldsymbol{\beta}, \sigma_\epsilon^2) &= \prod_{j=1}^{n_i} f(y_{ij} | \mathbf{u}_i; \boldsymbol{\beta}, \sigma_\epsilon^2) \\
&= (2\pi\sigma_\epsilon^2)^{-\frac{n_i}{2}} \exp\{-\|\mathbf{y}_i - \mathbf{x}'_{ij} \boldsymbol{\beta} - \mathbf{z}'_{ij} \mathbf{u}_i\|^2 / 2\sigma_\epsilon^2\}, \tag{4.15}
\end{aligned}$$

where $\|x\| = \{\sum_i x_i^2\}^{1/2}$ is the norm of the Euclidian vector.

Note that the two outcome processes are linked via the random effects $\mathbf{u}_i$, which are assumed to follow a normal distribution, $\mathbf{u}_i \sim N(0, \mathbf{D})$. When the association parameter $\varphi = 0$, the

joint analysis is equivalent to the separate analysis based on models (4.8) and (4.3). The observed data log-likelihood for all individuals in the study can be formulated as

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= \sum_{i=1}^N \log \int f(t_i, \delta_i, \mathbf{y}_i, \mathbf{u}_i | \boldsymbol{\theta}) d\mathbf{u}_i \\ &= \sum_{i=1}^N \log \int f(t_i | \mathbf{u}_i; \boldsymbol{\gamma}, \tau, \boldsymbol{\varphi}) \prod_{j=1}^{n_i} f(y_{ij} | \mathbf{u}_i; \boldsymbol{\beta}, \sigma_\varepsilon^2) f(\mathbf{u}_i | \mathbf{D}) d\mathbf{u}_i.\end{aligned}\quad (4.16)$$

Note that the log-likelihood involves the calculation of integrations with respect to the conditional distribution of $\mathbf{u}_i | (\mathbf{y}_i, t_i, \delta_i)$, which does not have a closed form. Thus, evaluation of the integral in $\ell(\boldsymbol{\theta})$ can be very challenging. However, the likelihood based method is the recommended approach for the joint analysis of longitudinal and survival data (Sweeting and Thompson, 2011 [75]). The ML of estimator of $\boldsymbol{\theta}$ is obtained by maximizing the log-likelihood function (4.16) with respect to $\boldsymbol{\theta}$ using an iterative algorithm, such as the Newton-Raphson iterative algorithm. Many approaches have been proposed to approximate the integral with high dimensional random effects, including adaptive Gaussian quadrature methods (Liu, 2009 [46]), some Monte Carlo resampling algorithm, such as the Metropolis-Hastings algorithm (Li, 2020 [39]), and fully Bayesian estimation methods (Béguin and Glas, 2001 [3]; Bolt and Lall, 2003 [4]; Edwards, 2010 [15]). It can be shown that the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is unbiased and asymptotically normally distributed, so if $N \rightarrow \infty$ then $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, [E\{I(\boldsymbol{\theta})\}]^{-1})$, where $I(\boldsymbol{\theta})$ is the observed Fisher information matrix.

For the simulation study, we consider a single frailty term for which the likelihood involves only one-dimensional integration. In this case, the computation of the likelihood is not tedious, so we consider finding the exact ML estimators of the model parameters. Specifically, we use the R function "optim" to obtain the maximum likelihood (ML) estimates. The R code is presented in the Appendix.

4.4 Review of Estimation Methods for Joint Models

Several estimation procedures have been developed to deal with the challenge of estimating the parameters of joint model. There are three general types of estimation methods in joint models of longitudinal and survival outcomes: the two-stage approach, Bayesian Markov Chain Monte Carlo (MCMC) method, and maximum-likelihood approach. Below we summarize two major approaches in the literature: two-stage approach and likelihood based approach.

Two-stage Approach

The numerical computation of the likelihood for the joint model is complicated as explained in the previous section. To estimate the model parameters, several two-stage methods were used in the joint modeling framework (Tsiatis, 1995 [81]; Self and Pwitan, 1992 [70]; Wu and Liu, 2012 [88]). Rizopoulos (2012) [62] and Self and Pawitan (1992) [70] used a least-squares method to estimate the random effects at step one and then used these estimates at step two to assign suitable values of $m_i(t)$, which are replaced in the partial likelihood of the Cox model. In addition, Tsiatis et al. (1995) [81] proposed a two-stage procedure with the attempt of reducing the bias of the parameter estimates introduced by the use of the time-dependent Cox model that incorporates a longitudinal covariate measured with error. In the first stage, a linear mixed effects model is fitted to estimate the true longitudinal process without measurement error. In the second stage, the estimates obtained by the LME model are plugged into a Cox proportional hazards model as covariates. In short, a two-stage method works as follows (Wu and Liu, 2012 [88]):

- Stage 1: The longitudinal covariate is modelled using a linear mixed effects (LME) model in order to estimate the subject-specific values of the covariate.
- Stage 2: The survival model is fitted using the modelled values in Stage 1 as time-varying covariate values.

This two-stage modeling is preferable to the extended Cox model with time-dependent covariates because this approach is not based on the observed longitudinal data in the survival model (Ibrahim, 2010 [28]), and it also reduces the bias of the parameter estimates in the Cox model (Wulfsohn and Tsiatis, 1997 [90]; Yu et al., 2004 [91]). However, this is not an unbiased method because no survival information is utilized when estimating the true longitudinal response and possible selection bias due to informative dropout is not taken into account. In other words, this approach does not use information from the longitudinal process and the survival process simultaneously in each model fitting stage (Yu et al., 2004 [91]; Wu and Liu, 2012 [88]). When the longitudinal process is only used, the linear mixed effects (LME) model can provide biased estimates in the first stage. As a consequence, estimates of the parameters in the survival model are potentially biased and inefficient in the second stage (Ibrahim, 2010 [28]).

Likelihood Based Approach

An increasingly popular alternative is the use of a joint likelihood approach, where the estimates of the parameters are obtained by maximizing the joint likelihood of the longitudinal process and survival time. As mentioned earlier, the joint model is comprised of two linked sub-models, one for the "true" longitudinal process and the other for the survival time along with additional assumptions that allow a full representation of the joint distribution of the observed data. As a result, this approach uses the entire available information optimally because both the longitudinal and survival data are utilized simultaneously. We focus on shared parameter models in this dissertation because this type of model has been widely used in recent studies. Let Y denote the longitudinal response, T the failure time, and b the random effects shared by models for both Y and T . The joint distribution of Y and T takes the form

$$f(Y, T) = \int_u f(Y|u)f(T|u)f(u)du. \quad (4.17)$$

Here the two outcome progresses are linked through their shared random effects, which account for the association between both outcomes. The most widely used method to estimate the parameters in joint models is the *joint likelihood approach* proposed by Wulfsohn and Tsiatis (1997) [90], which simultaneously fits models (4.8) and (4.3) linked by the random effects u_i via the joint likelihood.

4.5 Simulation Study

In this section, we conduct a comprehensive simulation study to evaluate the maximum likelihood method for the joint analysis of longitudinal and survival data. The models considered in the simulations are similar to those as used in Sections 2.6.2 and 3.4. We evaluate the performance of the ML estimators with respect to the bias, mean square error (MSE), coverage probabilities of 95% confidence intervals, and average length of 95% confidence intervals. These empirical properties were defined earlier in Section 2.6.1.

4.5.1 Generating longitudinal and survival data

In order to study the empirical properties of the estimators in the joint model, we ran a series of simulations using a linear mixed model for longitudinal outcomes y_{ij} and a frailty model for survival outcomes t_i . Specifically, the longitudinal outcomes were generated from the mixed model

$$y_{ij} = \beta_0 + \beta_1 time_j + \beta_2 time_j \times x_{1i} + u_i + \varepsilon_{ij}, \quad i = 1, \dots, N, j = 1, \dots, n_i, \quad (4.18)$$

where the random effects u_i were generated from $N(0, \sigma_u^2)$ and the random errors ε_{ij} were generated from $N(0, \sigma_\varepsilon^2)$. The covariates x_{1i} were generated as binary indicators ($x_{1i} = 0$ for control group and $x_{1i} = 1$ for treatment group). The follow-up times were set in months ($time_j = 0, 1, 2, \dots$) at which the longitudinal measurements were taken between the baseline time (month 0) and survival time (t_i). The true values of regression parameters and

variance components were fixed at $(\beta_0, \beta_1, \beta_2) = (10.0, -0.2, 0.6)$ and $(\sigma_\varepsilon^2, \sigma_u^2) = (0.5, 1.0)$, respectively. The true survival times s_i were generated from the Weibull frailty model

$$h_i(s_i|u_i, \lambda, \tau, \gamma, \varphi) = \tau s_i^{\tau-1} \exp(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \varphi u_i), \quad i = 1, 2, \dots, N, \quad (4.19)$$

where $\gamma_0 = \log(\lambda)$. The values of the baseline covariate x_{2i} were generated from $N(0, \sigma_x^2)$ with $\mu = 2$ and $\sigma^2 = 0.25$. The parameters of the frailty model were fixed at $\tau = 0.8$ and $(\gamma_0, \gamma_1, \gamma_2) = (-6, 1.5, 0.50)$, along with the association parameter $\varphi = 0.5$. The observed survival times t_i were defined by $t_i = \min(s_i, c_i)$, $(1, 2, \dots, n)$, where the censoring times c_i were generated from the Weibull distribution with shape= 0.8 and scale= -6.5. Finally, the censoring status was defined as $\delta_i = 1$ if $s_i \leq c_i$ and $\delta_i = 0$ if $s_i > c_i$. The data were generated for two combinations of sample sizes, $N = 300, 500$. For comparison, the parameters in models (4.18) and (4.19) were estimated using both joint modeling and separate modeling approaches described earlier. The results are discussed in the next section.

4.5.2 Simulation Results and Discussion

Table 4.1 presents empirical biases, mean squared errors (MSEs), coverage probabilities (CPs) and average lengths of 95% confidence intervals of the ML estimators under two scenarios: (i) joint analysis of the longitudinal and survival submodels, and (ii) separate analysis of the two submodels. The frailty variance was fixed at $\sigma_u^2 = 0.5$. Table 4.2 repeats them for a different value of frailty variance component $\sigma_u^2 = 1.0$. It is apparent from Table 4.1 that the estimators of the regression parameters $(\beta_0, \beta_1, \beta_2)$ and variance components $(\sigma_\varepsilon^2, \sigma_u^2)$ of the longitudinal submodel are unbiased under both approaches. However, the joint model provides much less biases compared to the separate model for longitudinal outcomes, especially for the parameter β_0 . The corresponding MSEs of the estimators are also very close by the two methods for both small and large sample sizes considered, except for the variance component σ_u^2 . As expected, the MSEs tend to decrease as the sample size N

increases. Moreover, both separate and joint approaches give CPs that are generally close to the nominal 95% confidence level. However, in the case of the survival submodel, there is a large difference between the estimates from the two approaches. As can be seen from Table 4.1, the separate analysis provides biased estimates, larger MSEs, and CPs for the regression parameters, as compared to the joint analysis. For instance, when estimating the regression coefficient γ_0 at sample size $N = 500$, the separate ML method provides a large bias of -0.2719, whereas the joint method gives a smaller bias of -0.0521. We also see that both separate and joint models generally provide close CPs for the parameters in the survival submodel. For example, when finding a 95% confidence interval for γ_0 at $N = 500$, the separate method gives a CP value of 0.96, whereas the proposed JM method gives a CP value of 0.96, which is close to the nominal 95% confidence level. However, under the separate analysis of the survival submodel, we notice that the variance estimates of frailty term are roughly biased, with large MSEs, poor CPs, and large average lengths of 95% confidence intervals for both sample sizes $N = \{300, 500\}$. This indicates that the ML estimators for the survival submodel under the joint model perform reasonably well and produces reasonably accurate estimates. Tables 4.2 presents empirical biases, MSEs, and CPs of the estimators for the frailty variance component $\sigma_u^2 = 1.0$. The ML method of the joint model appears to provide roughly unbiased estimates, smaller MSEs, good CPs, and better average lengths of 95 % CIs for most values of the parameters considered in the simulations. For example, the biases of the estimates of the longitudinal regression parameters $(\beta_0, \beta_1, \beta_2)$ at $N = 300$ are (-0.0067, -0.0016, -0.0003), whereas the corresponding MSEs are (0.0037, 0.0005, 0.0001). Also for the survival submodel under the joint modeling, when estimating the regression parameters $(\gamma_0, \gamma_1, \gamma_2)$ and shape parameter τ at $N = 300$, the biases are (-0.0373, 0.0079, 0.0064) and (0.0024), respectively, whereas the corresponding MSEs are (0.02858, 0.036, 0.0037) and (0.0005), respectively. It is important to note that the estimates σ_u^{2*} from the separate analysis of the survival submodel tend to provide large biases. For example, the bias of $\hat{\sigma}_u^{2*}$ at $N = 500$ is 0.265, with a CP value of 80%. In

summary, the joint modelling approach generally provides less bias and smaller MSEs of the estimates, and also gives CPs closer to the nominal 95% confidence level and smaller average lengths of confidence intervals, as compared to the separate modeling approach.

Table 4.1: Empirical biases, mean squared errors, coverage probabilities (CPs), and average lengths of 95% confident intervals. Simulations are based on 500 replicates of data set. Frailty variance $\sigma_u^2 = 0.5$.

Par	True Value	Separate models				Joint model			
		Bias	MSE	CP%	Length	Bias	MSE	CP%	Length
$N = 300$									
Longitudinal Submodel									
β_0	10	-0.0197	0.0028	92	0.1916	-0.0083	0.0023	95	0.1936
β_1	-0.2	-0.0025	0.0003	95	0.0746	-0.0006	0.0004	95	0.0788
β_2	0.6	-0.0008	0.0001	96	0.0378	-0.0006	0.0001	95	0.0399
σ_ϵ^2	0.5	0.0006	0.0003	94	0.0688	-0.0010	0.0440	94	0.0712
σ_u^2	0.5	-0.0099	0.0028	94	0.2015	-0.0070	0.2811	93	0.2075
Survival Submodel									
γ_0	-6.0	-0.3343	0.7136	97	3.3519	-0.0767	0.2870	96	2.1485
γ_1	1.5	0.0963	0.0852	97	1.1211	0.0167	0.0442	94	0.7856
γ_2	0.5	0.0260	0.0376	97	0.8208	0.0076	0.0023	95	0.7392
τ	0.8	0.0409	0.0113	95	0.4024	0.0091	0.0005	94	0.2410
σ_u^{2*}	0.125	0.2726	0.3313	69	1.8177				
φ	0.5					0.0336	0.0012	96	0.6334
$N = 500$									
Longitudinal Submodel									
β_0	10	-0.0198	0.0017	93	0.1484	-0.0095	0.0016	93	0.1504
β_1	-0.2	-0.0003	0.0002	96	0.0576	-0.0002	0.0003	95	0.0601
β_2	0.6	-0.0021	0.0006	93	0.0291	-0.0008	0.0001	94	0.0305
σ_ϵ^2	0.5	0.0008	0.0002	95	0.0532	-0.0004	0.0251	95	0.0551
σ_u^2	0.5	-0.0102	0.0017	93	0.1559	-0.0035	0.1734	93	0.1615
Survival Submodel									
γ_0	-6.0	-0.2719	0.5579	96	2.6307	-0.0521	0.1761	95	1.6566
γ_1	1.5	0.0685	0.0568	96	0.8730	0.0230	0.0256	95	0.6090
γ_2	0.5	0.0235	0.0266	97	0.6335	0.0008	0.0015	94	0.5700
τ	0.8	0.0307	0.0085	94	0.3146	0.0039	0.0003	96	0.1862
σ_u^{2*}	0.125	0.2410	0.2534	74	1.4664				
φ	0.5					0.0299	0.0010	95	0.4844

Table 4.2: Empirical biases, mean squared errors, coverage probabilities (CPs), and average lengths of 95% confident intervals. Simulations are based on 500 replicates of data set. Frailty variance $\sigma_u^2 = 0.1$.

Par	True Value	Separate models				Joint model			
		Bias	MSE	CP	Length	Bias	MSE	CP	Length
$N = 300$									
Longitudinal Submodel									
β_0	10	-0.0184	0.0042	95	0.2491	-0.0067	0.0037	96	0.2521
β_1	-0.2	-0.0029	0.0004	94	0.0773	-0.0016	0.0005	94	0.0810
β_2	0.6	-0.0011	0.0001	95	0.0392	-0.0003	0.0001	94	0.0412
σ_ϵ^2	0.5	0.0010	0.0003	95	0.0691	-0.0008	0.0359	94	0.0713
σ_u^2	1.0	-0.0240	0.0092	93	0.3637	-0.0066	0.2845	95	0.3763
Survival Submodel									
γ_0	-6.0	-0.3205	0.9068	96	3.5703	-0.0373	0.2858	96	2.1438
γ_1	1.5	0.0747	0.0995	94	1.1828	0.0079	0.0360	96	0.7887
γ_2	0.5	0.0216	0.0455	96	0.8631	0.0064	0.0037	94	0.7386
τ	0.8	0.0347	0.0127	95	0.4263	0.0024	0.0005	95	0.2391
σ_u^{2*}	0.25	0.3183	0.4895	70	2.2301				
φ	0.5					0.0117	0.0003	94	0.4287
$N = 500$									
Longitudinal Submodel									
β_0	10	-0.0231	0.0032	91	0.1931	-0.0135	0.0025	94	0.1947
β_1	-0.2	-0.0015	0.0002	94	0.0598	-0.0015	0.0002	96	0.0625
β_2	0.6	-0.00161	0.0001	94	0.0302	-0.0006	0.0001	96	0.0317
σ_ϵ^2	0.5	0.00001	0.0001	95	0.0534	0.0004	0.0230	96	0.0553
σ_u^2	1.0	-0.0226	0.0055	92	0.2821	-0.0148	0.1751	92	0.2893
Survival Submodel									
γ_0	-6.0	-0.2251	0.6135	93	2.8601	-0.0269	0.1757	96	1.6559
γ_1	1.5	0.0467	0.0548	96	0.9384	0.0164	0.0232	95	0.6108
γ_2	0.5	0.0112	0.0320	95	0.6637	-0.0068	0.0023	95	0.5698
τ	0.8	0.0229	0.0082	93	0.3427	0.0035	0.0003	96	0.1858
σ_u^{2*}	0.25	0.2650	0.3428	80	1.8700				
φ	0.5					0.0141	0.0003	95	0.3318

4.6 Application on HIV/AIDS Data

In this section, we re-analyze the HIV/AIDS data set introduced earlier in Section 2.7.2, based on the proposed joint modelling approach introduced earlier in Section 4.3. In the AIDS dataset, 467 patients with HIV infection were given antiretroviral treatment who had failed or were intolerant to zidovudine therapy, in which patients were randomly assigned to receive either didanosine (ddI) and zalcitabine (ddC) drugs. The CD4 cell counts were recorded at study entry, where randomization took place, as well as at 2, 6, 12, and 18 months thereafter. By the end of the study, a total of 188 deaths were recorded, resulting in about 59.7% censored survival times, and out of the 2335 planned measurements, 1405 were actually recorded, leading to 39.8% missing responses.

Our objective is to check if individual-specific longitudinal characteristics of CD4 cell counts are associated with the risk of death. As for comparisons, we consider analyzing the data based on the joint modeling of longitudinal CD4 response and time-to-death in this chapter. Instead of a separate analysis for each outcome illustrated in Chapter 2 and 3, an integrated (joint) analysis of correlated outcome processes is the desirable approach.

4.6.1 Longitudinal submodel

Figure 4.2 presents CD4 counts of randomly selected 50 AIDS patients. This figure indicates that, for many individuals, there are substantial variations in CD4 cell counts with different intercept levels over time. This suggests that a random intercept model would be necessary to explain within and between subjects variation. Ideally, we can use both random intercept and slope in our models. However, for the sake of computational simplicity, we use a random intercept model for the ASID data analysis.

To normalize CD4 counts, the square root of the CD4 counts is considered in the models. For modeling the longitudinal CD4 measurements y_{ij} , we consider a linear mixed effects model with a random intercept, with covariates observed time "obstime" (in months), binary

indicator of treatment use *Drug* and the interaction between observed time and drug use $Obstime \times Drug$, given by

$$y_{ij} = \beta_0 + \beta_1 Obstime_{ij} + \beta_2 Drug_i \times Obstime_{ij} + u_i + \varepsilon_{ij}, \quad (4.20)$$

where the random intercept u_i and random error ε_{ij} are assumed to be independently distributed as $N(0, \sigma_u^2)$ and $N(0, \sigma_\varepsilon^2)$, respectively. Moreover, u_i and ε_{ij} are assumed to be mutually independent. The linear mixed effects model (4.20) is similar to the model that Rizopoulos (2012) [62] considered when comparing the joint model to the extended Cox model. This model is adequate for studying the effects of various confounding factors in the AIDS study.

Figure 4.2: Spaghetti plot of CD4 cell counts of randomly selected 50 AISD patients.

The parameter estimates of the linear mixed model shown in Table 4.3 indicates that at

baseline the CD4 measurements of patients assigned to the ddI group were 0.5065 time lower in comparison with the ddC group. Furthermore, with a unit increase in the observation time, the CD4 measurements decrease by 0.1535. The result shows that the observed time was a significant predictor of CD4 cell counts ($\hat{\beta}_1 = -0.1561$; $95\%CI = [-0.1892, -0.1221]$), while the interaction between the observation time and treatment type did not have a significant impact on the CD4 cell count ($\hat{\beta}_3 = 0.0158$; $95\%CI = [-0.0310, 0.0626]$).

Table 4.3: Fitted Linear Mixed Effects Model with Random Intercept

Parameter	Estimate	Stderr	z-value	95 % CIs
Fixed effects				
Intercept (β_0)	7.1899	0.2192	32.8006	[6.7603, 7.6195]
Obstime (β_1)	-0.1561	0.0169	-9.2367	[-0.1892, -0.1230]*
Obstime:DrugddI (β_3)	0.0158	0.0239	0.6611	[-0.0310, 0.0626]
Random effects				
σ_u^2	20.0111	1.3952	14.3428	[17.2765, 22.7457]
Between subject error				
σ_ϵ^2	3.8450	0.1775	21.6620	[3.4971, 4.1929]
AIC= 7170.669				

* Indicates significance covariate at 5% level of significance

4.6.2 Survival submodel

We chose the Weibull distribution after some preliminary analysis of the data with several other distributions, including log-normal and log-logistic distributions, as the Weibull distribution possesses the desirable properties of both proportional hazards and accelerated failure time models. The covariate previous opportunistic infection PrevOI is included in the analysis because of the importance of exposure risk categories at diagnosis. Although Drug at diagnosis was not significant, it is included in the model as a potential confounder.

For modeling the survival times (18-months mortality rate) t_i of the HIV infected patients, we consider a frailty model in the framework of the Weibull proportional hazards model with

the covariates Drug and PrevOI measured at baseline time, given by

$$h_i(t_i|u_i, \lambda, \tau, \gamma, \varphi) = \tau t_i^{\tau-1} \exp(\gamma_0 + \gamma_1 Drug_i + \gamma_2 PrevOI_i + \varphi u_i), \quad (4.21)$$

for $i = 1, \dots, N$, where $h_i(\cdot)$ is the hazard function for the i^{th} patient, $\lambda = \exp(\gamma_0)$ and τ are the scale and shape parameters of the Weibull distribution, (γ_1, γ_2) are the regression parameters, and φ is an association parameter measuring the strength of association between the two outcome processes. Table 4.4 presents results from separate analyses of the outcome processes (survival) under the assumption that the association parameter φ is zero, that is

$$h_i(T_i|u_i, \lambda, \tau, \gamma, \varphi) = \tau t_i^{\tau-1} \exp(\gamma_0 + \gamma_1 Drug_i + \gamma_2 PrevOI_i). \quad (4.22)$$

The table displays ML estimates of the Weibull PH with no frailty along with corresponding standard errors, z-values, and lower limits (LL) and upper limits (UL) limits of the parameters at 95% confidence level. As can be observed from the fitted Weibull PH model, the relative hazard rate for patients with ddI treatment to ddC group is $\exp(0.3516) = 1.4213$. At a given instant in time, a patient who has taken the ddI drug is 1.42 times as likely to die as a patient who has taken the ddC drug adjusting for other covariates, though not statistically significant. We also observe that one month increase in the baseline previous opportunistic infection is associated with a 49.39% increase in the risk of death. Furthermore, the estimated shape parameter for the Weibull ($\hat{\tau} = 0.6945$) shows that the death rate from the HIV infection was decreasing over time.

Table 4.4: Fitted Weibull proportional hazards (PH) model (4.22).

Parameter	Survival submodel ($\varphi = 0$)				
	Estimate	Stderr	z-value	LL	UL
Intercept (γ_0)	-7.1668	0.9869	-7.2623	-9.1011	-5.2326
Drug (γ_1)	0.3516	0.2887	1.2181	-0.2142	0.9175
PrevOI (γ_2)	1.4939	0.4362	3.4252	0.6391	2.3488
Shape (τ)	0.6945	0.0948	7.3283	0.5088	0.8803

4.6.3 Joint modeling

After fitting separate models that predict the change of CD4 measurements over time and the time to death of HIV infected patients, the next step is to explore the joint model that associates the longitudinally measured CD4 biomarker with time to death of patients with HIV infection. In the joint modeling, we combine the linear mixed-effects submodel (4.20) and Weibull proportional hazards submodel (4.21). Note that the survival submodel includes the association parameter φ in the Weibull PH model, in which φu_i defines the nature of association structure between the two processes, φ measures the strength of association between the two processes. As explained earlier, the joint model was fitted by the ML approach using the R function "optim" that implements the "L-BFGS-B" method presented by Byrd et. al. (1995), which allows box constraints, that is, each variable can be given a lower and/or upper bound. Table 4.5 presents ML estimates of the joint model parameters along with corresponding standard errors, z-values, and lower limits (LL) and upper limits (UL) of the parameters at 95% confidence level. It is clear from the longitudinal submodel that the observation time (p.value<0.0001, with 95% CI [-0.1989,-0.1327]) is a significant predictor of the CD4 cell count at the 5% level of significance, whereas the interaction between the drug and observation time is not significant at the 5% level of significance (p.value=0.6200, with 95% CI [0.0348,0.0584]). This indicates that for every one month increase in the observation time, the CD4 cell count decreases by 0.1658, and there is no significant interaction between the observation time and the type of drug the patient has taken. From the survival

submodel, the hazard rate for patients assigned to ddI is $\exp(0.2842) = 1.33$, which indicates a greater risk of death from HIV infection. The results do not give a significant estimate $\hat{\gamma}_1 = 0.2842$ (95% CI [-0.0116, 0.5800]). Also, patients with a previous diagnosis of AIDS have a higher risk, as the hazard rate is $\exp(0.7340) = 2.0834$. That clearly indicates that the previous diagnosis of AIDS for patients is a significant predictor of death from disease (p.value<0.0001, with 95% CI [0.3101, 1.1579]). An important focus of the joint analysis is on the association parameter φ , which is used to assess the strength of association between the two submodels. In this study, the purpose of the joint modelling is to link the longitudinally measured CD4 measurements with the time to death of HIV infected patients to understand their association. We can see that the joint likelihood provides a significant estimate of $\hat{\varphi} = -0.2212$ (95% CI [-0.2857, -0.1567]) with a unit decrease in the CD4 marker corresponding to a $\exp(\varphi) = 1.25$ increase in the risk for death, which justifies the use of the joint modeling approach for the data analysis. The association term (p-value<0.0001) is significant; this implies that the CD4 cell count is associated with the risk of death or disease progression. Table 4.5 also presents results from separate analyses of the two outcome processes (longitudinal and survival) under the assumption that the association parameter φ is zero. We clearly observe differences in the two sets of estimates obtained under the two methods (joint vs. separate). For instance, with regards to the covariate "PrevOI", the JM and the separate survival models provide estimates of (0.7340, 1.4939) and standard errors of (0.2163, 0.4362), respectively. As another example, with regards to the binary covariate "Drug", the JM provides an estimate of 0.2842 with a standard error of 0.1509, whereas the separate survival model provides an estimate of 0.3516 with a standard error of 0.2887.

Table 4.5: Joint and separate analyses of time-to-death and longitudinal CD4 outcomes from HIV/AIDS study.

Parameter	Separate model					Joint model				
	Estimate	SE	z-vlaue	LL	UL	Estimate	SE	z-vlaue	LL	UL
Longitudinal Submodel										
Intercept (β_0)	7.1899	0.2192	32.8006	6.7603	7.6195	7.1708	0.2219	32.3155	6.7359	7.6057
Obstime (β_1)	-0.1561	0.0169	-9.2367	-0.1892	-0.1221	-0.1658	0.0169	-9.8107	-0.1989	-0.1327
Obstime $\times$ Drug (β_2)	0.0158	0.0239	0.6611	-0.0310	0.0626	0.0118	0.0238	0.4958	-0.0348	0.0584
σ_u^2	20.0111	1.3952	14.3428	17.2765	22.7457	20.6003	1.4481	14.2257	17.7620	23.4386
σ_ϵ^2	3.8450	0.1775	21.6620	3.4971	4.1929	3.8283	0.1760	21.7517	3.4833	4.1733
Survival Submodel										
Intercept (γ_0)	-7.1668	0.9869	-7.2623	-9.1011	-5.2326	-5.7317	0.3480	-16.4704	-6.4138	-5.0496
Drug (γ_1)	0.3516	0.2887	1.2181	-0.2142	0.9175	0.2842	0.1509	1.8834	-0.0116	0.5800
PrevOI (γ_2)	1.4939	0.4362	3.4252	0.6391	2.3488	0.7340	0.2163	3.3934	0.3101	1.1579
Shape (τ)	0.6945	0.0948	7.3283	0.5088	0.8803	1.5062	0.0986	15.2759	1.3129	1.6995
φ						-0.2212	0.0329	-6.7234	-0.2857	-0.1567

Chapter 5

Joint Analysis Based On H-likelihood

5.1 Introduction

Joint modelling of longitudinal and survival data encounters many challenges that cannot be addressed by standard likelihood methods. Therefore, new statistical models and methods are needed in order to conduct a comprehensive statistical analysis. In the literature, numerous methods have been proposed for the parameter estimation in joint models, including the Monte Carlo EM algorithm, Adaptive Gauss-Hermite (AGH) quadrature method, and approximate methods based on Laplace approximation. In this chapter, we address the computational challenges of the likelihood inference by proposing an approximate method based on the Laplace approximation, referred to as the *hierarchical likelihood (h-likelihood) method*.

Lee and Nelder (1996) [36] introduced the hierarchical-likelihood (h-likelihood) approach, which provides a simple integrated framework and a numerically efficient fitting algorithm for several random effect models, such as frailty models (Ha, Lee and Song, 2001 [22]; Ha and Lee, 2003 [21]), mixed linear survival models (Ha, Lee and Song, 2002 [12]), spatial correlation models for non-normal data (Lee and Nelder, 2001a [37]) and structured dispersion models, which provide statistically reliable and efficient estimators (Lee and Nelder, 2001b [36]).

Using a numerical study, Ha and Lee (2003a) [21] explained that the h-likelihood methods provides less bias in the parameter estimation, as compared to the maximum likelihood method proposed by Nielsen et al. (1992) [52], as well as the penalized likelihood method introduced by Ripatti and Palmgren (2000) [61].

5.2 Approximate Inference Based on the H-likelihood

The h-likelihood method is basically a first-order Laplace approximation method that can produce approximate MLEs for mean parameters and restricted MLEs (REMLs) for variance-covariance parameters in a model (Lee and Nelder, 1996 [36]; Ha, Noh and Lee, 2017 [23]).

The h-likelihood function for the joint model (4.12) can be written as

$$\ell_h = \sum_{i=1}^N \{ \log f(\varsigma_i | \mathbf{u}_i, \boldsymbol{\theta}) + \log f(\mathbf{u}_i | \mathbf{D}) \}, \quad (5.1)$$

where $f(\varsigma_i | \mathbf{u}_i, \boldsymbol{\theta})$ is the density function of the observed data ς_i . Generally, the basic h-likelihood method consists of three steps. Let $\boldsymbol{\theta}$ be the vector of mean parameters and $\boldsymbol{\xi}$ be the vector of dispersion parameters in the collection of all model parameters $\boldsymbol{\eta}$, i.e. $\boldsymbol{\eta} = (\boldsymbol{\theta}', \boldsymbol{\xi}')$. Beginning with some initial parameter estimates $\boldsymbol{\theta}^{(0)}$ and $\boldsymbol{\xi}^{(0)}$, we can obtain the h-likelihood-based estimates by iterating the following steps until convergence at iteration k , ($k = 0, 1, 2, \dots$):

- Step 1: Given $\hat{\boldsymbol{\theta}}^{(k)}$ and $\hat{\boldsymbol{\xi}}^{(k)}$, obtain $\hat{\mathbf{u}}_i^{(k+1)}$ by maximizing ℓ_h with respect to $\mathbf{u}_i$ for $i = 1, 2, \dots, N$.
- Step 2: Given $\hat{\mathbf{u}}_i^{(k+1)}$ and $\hat{\boldsymbol{\xi}}^{(k)}$, obtain $\hat{\boldsymbol{\theta}}^{(k+1)}$ by maximizing the following adjusted profile h-likelihood with respect to $\boldsymbol{\theta}$:

$$p_{\boldsymbol{\theta}} = \left[\ell_h - 0.5 \log \left| \frac{H(\ell_h, \mathbf{u})}{2\pi} \right| \right] \Big|_{\mathbf{u}=\hat{\mathbf{u}}^{(k+1)}, \boldsymbol{\xi}=\hat{\boldsymbol{\xi}}^{(k)}},$$

where $H(\ell_h, \mathbf{u}) = -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \mathbf{u}'}$.

- Step 3: Given $\hat{\mathbf{u}}_i^{(k+1)}$ and $\hat{\boldsymbol{\theta}}^{(k+1)}$, obtain $\hat{\boldsymbol{\xi}}^{(k+1)}$ by maximizing the following adjusted profile h-likelihood with respect to $\boldsymbol{\xi}$:

$$p_{\boldsymbol{\xi}} = \left[\ell_h - 0.5 \log \left| \frac{H[\ell_h, (\mathbf{u}, \boldsymbol{\theta})]}{2\pi} \right| \right] \Big|_{\mathbf{u}=\hat{\mathbf{u}}^{(k+1)}, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(k+1)}},$$

$$\text{where } H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})] = - \begin{bmatrix} \frac{\partial^2 \ell_h}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} & \frac{\partial^2 \ell_h}{\partial \boldsymbol{\theta} \partial \mathbf{u}'} \\ \frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \boldsymbol{\theta}'} & \frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \mathbf{u}'} \end{bmatrix}.$$

At convergence, we obtain approximate MLEs $\hat{\boldsymbol{\theta}}$ for the mean parameters $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\xi}}$ for the variance-covariance parameters $\boldsymbol{\xi}$.

5.3 Joint Analysis Based on H-Likelihood

The longitudinal and survival models described in the previous chapter are associated through shared random effects. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \varphi)$ be the collection of the mean parameters in the two models, and let $\boldsymbol{\xi} = (\sigma_\varepsilon^2, \mathbf{D})$ be the collection of the variance (dispersion) parameters. Given observed data (t_i, δ_i, y_i) from both survival and longitudinal outcomes, the joint likelihood is given by

$$L(\boldsymbol{\theta}, \boldsymbol{\xi}) = \prod_{i=1}^N \left\{ \int f(t_i, \delta_i | \mathbf{u}_i; \tau, \boldsymbol{\gamma}, \varphi) f(\mathbf{y}_i | \mathbf{u}_i; \boldsymbol{\beta}, \sigma_\varepsilon^2) f(\mathbf{u}_i | \mathbf{D}) d\mathbf{u}_i \right\}. \quad (5.2)$$

Following Lee and Nelder (1996) [36] and Ha et al. (2001) [12], the h-likelihood for the joint model, is defined by

$$h = \prod_{i=1}^N h_{1i} \times h_{2i} \times h_{3i}, \quad (5.3)$$

where

$$\begin{aligned}
h_{1i} = f(t_i, \delta_i | \mathbf{u}_i; \gamma, \tau, \varphi) &= \{\tau t_i^{\tau-1} \exp(\mathbf{w}_i' \gamma + \varphi' \mathbf{u}_i)\}^{\delta_i} \times \exp \left\{ - \int_0^{s_i} t_i^\tau \exp(\mathbf{w}_i' \gamma + \varphi' \mathbf{u}_i) dt \right\}, \\
h_{2i} = f(\mathbf{y}_i | \mathbf{u}_i; \beta, \sigma_\varepsilon^2) &= (2\pi\sigma_\varepsilon^2)^{-n_i/2} \exp \left\{ - \frac{[\mathbf{y}_i - \mathbf{X}_i \beta - \mathbf{Z}_i \mathbf{u}_i]' [\mathbf{y}_i - \mathbf{X}_i \beta - \mathbf{Z}_i \mathbf{u}_i]}{2\sigma_\varepsilon^2} \right\}, \\
\text{and } h_{3i} = f(\mathbf{u}_i | \mathbf{D}) &= \frac{\exp\{-\frac{1}{2} \mathbf{u}_i' \mathbf{D}^{-1} \mathbf{u}_i\}}{(2\pi)^{q/2} |\mathbf{D}|^{1/2}}
\end{aligned}$$

are the conditional density function for t_i given $\mathbf{u}_i$, conditional density function for $\mathbf{y}_i$ given $\mathbf{u}_i$, and marginal density function for the frailty term $\mathbf{u}_i$, respectively. The ML estimators of $\boldsymbol{\eta} = (\boldsymbol{\theta}', \boldsymbol{\xi}')'$ are obtained by maximizing the log-likelihood function. The corresponding logarithm of the h-likelihood is defined by

$$\begin{aligned}
\ell_h = \sum_i \ell_{h_i} &= \sum_i \ell_{h_{1i}} + \sum_i \ell_{h_{2i}} + \sum_i \ell_{h_{3i}} \\
&= \left\{ \sum_i \log f(t_i, \delta_i | \mathbf{u}_i; \tau, \gamma, \varphi) \right\} + \left\{ \sum_i \log f(\mathbf{y}_i | \mathbf{u}_i; \beta, \sigma_\varepsilon^2) \right\} + \left\{ \sum_i \log f(\mathbf{u}_i; \mathbf{D}) \right\} \\
&= \sum_i \delta_i \log(\tau) + (\tau - 1) \sum_i \delta_i \log(t_i) + \sum_i \delta_i (\mathbf{w}_i' \gamma + \varphi' \mathbf{u}_i) - \sum_i t_i^\tau \exp(\mathbf{w}_i' \gamma + \varphi' \mathbf{u}_i) \\
&\quad + \sum_i \left\{ -\frac{n_i}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}_i| - \frac{1}{2} (\mathbf{y}_i - \mathbf{X}_i \beta - \mathbf{Z}_i \mathbf{u}_i)' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta - \mathbf{Z}_i \mathbf{u}_i) \right\} \\
&\quad + \sum_i \left\{ -\frac{q}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{D}| - \frac{1}{2} \mathbf{u}_i' \mathbf{D}^{-1} \mathbf{u}_i \right\}. \tag{5.4}
\end{aligned}$$

Estimation of random effects $\mathbf{u}_i$

Let $\mathbf{u}_i = (u_{1i}, u_{2i}, \dots, u_{qi})$ be the $q \times 1$ vector of random effects. Given the parameters $(\boldsymbol{\theta}, \boldsymbol{\xi})$, the maximum h-likelihood estimators (MHLEs) for $\mathbf{u}$ are obtained by solving

$$\partial \ell_h / \partial \mathbf{u}_i = \mathbf{0}. \tag{5.5}$$

We use the Newton-Raphson method to solve the estimating equations (5.5). The Newton-Raphson method obtains the MHLEs of random effects by solving the iterative equation

$$\hat{\mathbf{u}}_i^{(m+1)} = \hat{\mathbf{u}}_i^{(m)} + \left(-\frac{\partial^2 \ell_h}{\partial \mathbf{u}_i \partial \mathbf{u}_i} \right)^{-1} \frac{\partial \ell_h}{\partial \mathbf{u}_i}, \quad (5.6)$$

where

$$\begin{aligned} \frac{\partial \ell_h}{\partial \mathbf{u}_i} = & \mathbf{Z}_i' \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{u}_i) + \varphi \delta_i - \varphi t_i^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \varphi' \mathbf{u}_i) \\ & - \mathbf{u}_i \mathbf{D}^{-1}, \end{aligned} \quad (5.7)$$

and

$$\frac{\partial^2 \ell_h}{\partial \mathbf{u}_i \partial \mathbf{u}_i'} = -\mathbf{Z}_i' \mathbf{V}_i^{-1} \mathbf{Z}_i - \varphi \varphi' t_i^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \varphi' \mathbf{u}_i) - \mathbf{D}^{-1}. \quad (5.8)$$

Estimation of fixed effects $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \varphi)$

Generally, Noh and Lee (2007) [53] suggest a Laplace approximation to the marginal likelihood to estimate fixed effects in the mean structure for GLMMs. Similarly, we obtain the maximum h-likelihood estimators (MHLEs) of fixed effects parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \varphi)$ by maximizing the adjusted profile h-likelihood. This involves the evaluation of the adjusted profile (log)-likelihood

$$\begin{aligned} p_{\boldsymbol{\theta}} &= \left[\ell_h - 0.5 \log \left| \frac{H[(\ell_h, \mathbf{u})]}{2\pi} \right| \right] \Big|_{\mathbf{u}=\hat{\mathbf{u}}} \\ &= \left[\ell_h - 0.5 \log \left(\frac{1}{2\pi} \left| H[(\ell_h, \mathbf{u})] \right| \right) \right] \Big|_{\mathbf{u}=\hat{\mathbf{u}}} \\ &= \left[\ell_h - 0.5 \left(\log \left(\frac{1}{2\pi} \right) + \log \left| H[(\ell_h, \mathbf{u})] \right| \right) \right] \Big|_{\mathbf{u}=\hat{\mathbf{u}}}, \end{aligned} \quad (5.9)$$

where $H(\ell_h, \mathbf{u}) = -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \mathbf{u}'}$. The maximum adjusted profile h-likelihood estimators are obtained by solving the score equations

$$\begin{aligned} S(\boldsymbol{\theta}) = \frac{\partial p_\theta}{\partial \boldsymbol{\theta}} &= \frac{\partial \ell_h}{\partial \boldsymbol{\theta}} - 0.5 \frac{\partial \log |H(\ell_h, \mathbf{u})|}{\partial \boldsymbol{\theta}} \\ &= \frac{\partial \ell_h}{\partial \boldsymbol{\theta}} - 0.5 \frac{1}{|H(\ell_h, \mathbf{u})|} \frac{\partial |H(\ell_h, \mathbf{u})|}{\partial \boldsymbol{\theta}} = \mathbf{0}, \end{aligned}$$

where

$$\frac{\partial \ell_h}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \ell_h}{\partial \boldsymbol{\beta}} \\ \frac{\partial \ell_h}{\partial \boldsymbol{\gamma}} \\ \frac{\partial \ell_h}{\partial \tau} \\ \frac{\partial \ell_h}{\partial \varphi} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{Y}_i - \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \boldsymbol{\beta} - \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{u}_i \\ \sum_{i=1}^N \delta_i \mathbf{w}_i - \sum_{i=1}^N t_i^\tau \mathbf{w}_i \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \varphi' \mathbf{u}_i) \\ \sum_{i=1}^N \frac{\delta_i}{\tau} + \sum_{i=1}^N \delta_i \log(t_i) - \sum_{i=1}^N t_i^\tau \log(t_i) \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \varphi' \mathbf{u}_i) \\ \sum_{i=1}^N \delta_i \mathbf{u}_i - \sum_{i=1}^N \mathbf{u}_i t_i^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \varphi' \mathbf{u}_i) \end{bmatrix},$$

and

$$\begin{aligned} |H(\ell_h, \mathbf{u})| &= \left| H[(\ell_h, \mathbf{u})] \right| \\ &= \left| -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \mathbf{u}'} \right| \end{aligned}$$

We use the Newton-Raphson method to obtain the MHLEs of fixed effects parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \varphi)$, which leads to the iterative equation

$$\hat{\boldsymbol{\theta}}^{(m+1)} = \hat{\boldsymbol{\theta}}^{(m)} + \left(-\frac{\partial^2 p_\theta}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)^{-1} \frac{\partial p_\theta}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}^{(m)}}, \quad (5.10)$$

for $m = 0, 1, 2, \dots$. More details of the derivations can be found in the Appendix.

Estimation of variance components $\boldsymbol{\xi}$

To estimate $\boldsymbol{\xi}$, a Newton-Raphson algorithm can be applied to maximize the appropriate adjusted profile likelihood. Following Lee and Nelder (1996) [36], we use the adjusted profile

h-likelihood defined by

$$\begin{aligned}
p_{\xi} &= \left[\ell_h - 0.5 \log \left| \frac{H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})]}{2\pi} \right| \right] \Big|_{\xi=\hat{\xi}, u=\hat{u}} \\
&= \left[\ell_h - 0.5 \log \left(\frac{1}{2\pi} \left| H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})] \right| \right) \right] \Big|_{\xi=\hat{\xi}, u=\hat{u}} \\
&= \left[\ell_h - 0.5 \left(\log \left(\frac{1}{2\pi} \right) + \log \left| H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})] \right| \right) \right] \Big|_{\xi=\hat{\xi}, u=\hat{u}}, \tag{5.11}
\end{aligned}$$

where $H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})] = - \begin{bmatrix} \frac{\partial^2 \ell_h}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} & \frac{\partial^2 \ell_h}{\partial \boldsymbol{\theta} \partial \mathbf{u}'} \\ \frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \boldsymbol{\theta}'} & \frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \mathbf{u}'} \end{bmatrix}$.

The maximum adjusted profile h-likelihood estimators of p_{ξ} are obtained by solving the score equations

$$\begin{aligned}
S(\boldsymbol{\xi}) = \frac{\partial p_{\xi}}{\partial \boldsymbol{\xi}} &= \frac{\partial \ell_h}{\partial \boldsymbol{\xi}} - 0.5 \frac{\partial \log |H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})]|}{\partial \boldsymbol{\theta}} \\
&= \frac{\partial \ell_h}{\partial \boldsymbol{\xi}} - 0.5 \frac{1}{|H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})]|} \frac{\partial |H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})]|}{\partial \boldsymbol{\theta}} = 0.
\end{aligned}$$

The Newton-Raphson method leads to the iterative equation

$$\hat{\boldsymbol{\xi}}^{(m+1)} = \hat{\boldsymbol{\xi}}^{(m)} + \left(- \frac{\partial^2 p_{\xi}}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}'} \right)^{-1} \frac{\partial p_{\xi}}{\partial \boldsymbol{\xi}} \Big|_{\boldsymbol{\xi}^{(m)}}, \tag{5.12}$$

for $m = 0, 1, 2, \dots$.

5.4 Simulation Studies

In this section, we conduct a simulation study to assess the performance of the h-likelihood method. In the simulation study, we consider a joint model with one- dimensional of random effects and with an exponential survival model (as a special case of the Weibull distribution). We compare the performance of the h-likelihood and likelihood methods based on the bias, mean square error (MSE), coverage Probability (CP) and average length of 95% confidence

intervals (Length) of the parameter estimates.

In this study, the data are generated as follows. The continuous longitudinal response y_{ij} are generated from a liner mixed model with random intercept, specifically from the mixed model

$$y_{ij} = \beta x_i + u_i + \epsilon_{ij}, \quad i = 1, 2, \dots, N, j = 1, 2, \dots, n_i, \quad (5.13)$$

where x_i is a baseline covariate generated from the standard normal distribution. The random effect, u_i , is generated from a normal distribution with mean 0 and variance σ_u^2 , while the random error, ϵ_{ij} , is generated from $N(0, \sigma_\epsilon^2)$. The true values of the parameters are set to be: $\beta = 0.5, \sigma_u^2 = \{0.25, 0.5\}, \sigma_\epsilon^2 = 0.5$.

The survival times t_i are generated from an exponential distribution given by

$$h_i(t) = \exp(\gamma + \varphi u_i), \quad i = 1, 2, \dots, N, \quad (5.14)$$

where $\exp(\gamma) = \lambda$. The noninformative censoring survival times c_i are generated from an exponential distribution with $\lambda = -5.0$. The true value of the regression parameter is $\gamma = -4.5$ and different values for the association parameter φ are $\varphi = \{0.25, 0.5, 0.8\}$, where the number of replications in each simulation is $S = 500$.

5.4.1 Simulation Results

We assess the performance of both maximum likelihood (ML) and h-likelihood (HL) methods for analyzing joint models with different values of the association parameter that links the longitudinal and survival outcomes. Tables 5.1-5.3 summarize the results of the simulation study for three combinations of sample sizes $N = \{200, 300, 500\}$ and for three combinations of $(\sigma_u^2, \varphi) = \{(0.5, 0.25), (0.5, 0.5), (0.5, 0.8)\}$.

From the tables, the estimates obtained by the H-likelihood method appear to perform well. Both ML and H-likelihood estimates are roughly unbiased under all scenarios considered here. The H-likelihood methods, however, loses slight efficiency in terms of higher mean

squared errors of the parameter estimates, as compared to the ML method. For example, with $N = 500$ and $(\varphi, \sigma_u^2) = (0.8, 0.8)$, Table 5.3 shows that the ML estimates of $(\beta, \sigma_\varepsilon^2, \sigma_u^2)$ gives MSEs of $(0.00059, 0.0030, 0.00456)$, whereas the corresponding H-likelihood estimates give slightly higher MSEs of $(0.00063, 0.00036, 0.00470)$, respectively. We expect to lose such small efficiencies from the computationally simpler H-likelihood method for approximating the exact likelihood by the Laplace approximation.

The coverage probabilities of the ML estimates of 95% confidence intervals are generally better than those obtained from the H-likelihood estimates. For example, with $N = 300$ and $(\varphi, \sigma_u^2) = (0.8, 0.8)$, it is observed in Table 5.3 that the coverage probabilities of the ML estimates of $(\beta, \sigma_\varepsilon^2, \sigma_u^2)$ are $(0.94, 0.95, 0.93)$, whereas the corresponding coverage probabilities of H-likelihood estimates are $(0.86, 0.88, 0.92)$, respectively. For smaller values of (φ, σ_u^2) , however, both ML and H-likelihood methods provides coverage probabilities that are generally close to the nominal confidence level, as can be seen from Table 5.1 for the parameters $(\varphi, \sigma_u^2) = (0.25, 0.5)$.

Table 5.1: Empirical biases, mean squared errors, coverage probabilities (CPs), and average length of 95% confidence intervals obtained by both ML and H-likelihood methods. Association parameter $\varphi = 0.25$, variance component $\sigma_u^2 = 0.5$. Number of simulations $S = 500$.

Par	True Value	ML Method				H-likelihood Method			
		Bias	MSE	CP%	Length	Bias	MSE	CP%	Length
$N = 200$									
Longitudinal Submodel									
β	0.5	-0.00213	0.00084	95	0.1121	-0.00431	0.00074	96	0.1126
σ_{ϵ}^2	0.5	0.00379	0.00091	94	0.1110	-0.00023	0.00091	94	0.1195
σ_u^2	0.5	-0.00513	0.00502	94	0.2664	0.00073	0.00345	99	0.2986
Survival Submodel									
γ	-4.5	0.00395	0.00919	95	0.3655	0.00188	0.00852	95	0.3659
φ	0.25	0.00121	0.02563	94	0.6015	0.00893	0.02384	98	0.6905
$N = 300$									
Longitudinal Submodel									
β	0.5	-0.00317	0.00059	96	0.0918	-0.00375	0.00050	96	0.0920
σ_{ϵ}^2	0.5	-0.00147	0.00053	95	0.0895	0.00030	0.00062	95	0.0974
σ_u^2	0.5	0.00222	0.00298	96	0.2193	0.00252	0.00209	99	0.2418
Survival Submodel									
γ	-4.5	-0.00058	0.00547	97	0.2985	0.00094	0.00571	96	0.2981
φ	0.25	0.01787	0.01532	94	0.4819	0.01307	0.01563	98	0.5522
$N = 500$									
Longitudinal Submodel									
β	0.5	-0.00317	0.00034	95	0.0709	-0.00371	0.00033	94	0.0712
σ_{ϵ}^2	0.5	-0.00081	0.00029	96	0.0693	0.00009	0.00035	96	0.0755
σ_u^2	0.5	-0.00225	0.00172	95	0.1685	-0.00035	0.00133	99	0.1861
Survival Submodel									
γ	-4.5	-0.00308	0.00319	96	0.2305	-0.00196	0.00357	95	0.2310
φ	0.25	0.00854	0.00906	96	0.3732	0.00854	0.00906	98	0.4251

Table 5.2: Empirical biases, mean squared errors, coverage probabilities (CPs), and average length of 95% confidence intervals obtained by both ML and H-likelihood methods. Association parameter $\varphi = 0.5$, variance component $\sigma_u^2 = 0.5$. Number of simulations $S = 500$.

Par	True Value	ML Method				H-likelihood Method			
		Bias	MSE	CP%	Length	Bias	MSE	CP%	Length
<i>N</i> = 200									
Longitudinal Submodel									
β	0.5	-0.00570	0.00086	94	0.1120	-0.00772	0.00086	95	0.1121
σ_{ϵ}^2	0.5	0.00017	0.00067	97	0.1087	-0.00083	0.00079	95	0.1175
σ_u^2	0.5	-0.00314	0.00505	93	0.2671	0.00160	0.00318	99	0.2968
Survival Submodel									
γ	-4.5	-0.00525	0.00788	97	0.3794	-0.01109	0.00955	94	0.3833
φ	0.5	0.01986	0.02836	94	0.6150	0.03325	0.02662	98	0.7320
<i>N</i> = 300									
Longitudinal Submodel									
β	0.5	-0.00813	0.00066	92	0.0918	-0.00822	0.00066	93	0.0920
σ_{ϵ}^2	0.5	0.00023	0.00046	96	0.0887	0.00363	0.00056	97	0.0967
σ_u^2	0.5	0.00016	0.00321	95	0.2192	0.00032	0.00187	99	0.2416
Survival Submodel									
γ	-4.5	-0.01020	0.00741	93	0.3101	-0.00525	0.00692	94	0.3113
φ	0.5	0.02305	0.01776	93	0.4997	0.03106	0.01766	98	0.5862
<i>N</i> = 500									
Longitudinal Submodel									
β	0.5	-0.00452	0.00032	95	0.0711	-0.00687	0.00036	95	0.0711
σ_{ϵ}^2	0.5	-0.00033	0.00030	95	0.0686	0.00037	0.00036	95	0.0743
σ_u^2	0.5	-0.00004	0.00193	96	0.1698	-0.00252	0.00135	99	0.1851
Survival Submodel									
γ	-4.5	-0.00648	0.00363	95	0.2398	-0.00777	0.00387	95	0.2408
φ	0.5	0.02718	0.01144	94	0.3852	0.03017	0.01100	98	0.4532

Table 5.3: Empirical biases, mean squared errors, coverage probabilities (CPs), and average length of 95% confidence intervals obtained by both ML and H-likelihood methods. Association parameter $\varphi = 0.8$, variance component $\sigma_u^2 = 0.8$. Number of simulations $S = 500$.

Par	True	ML Method				H-likelihood Method			
	Value	Bias	MSE	CP%	Length	Bias	MSE	CP%	Length
<i>N</i> = 200									
Longitudinal Submodel									
β	0.5	-0.00892	0.00134	93	0.1328	-0.01590	0.00118	90	0.1105
σ_{ϵ}^2	0.5	0.00040	0.00064	96	0.1057	-0.00716	0.00091	93	0.1108
σ_u^2	0.8	-0.02175	0.00998	93	0.3853	-0.03056	0.00915	93	0.3754
Survival Submodel									
γ	-4.5	-0.02291	0.01138	95	0.4302	-0.03504	0.01199	96	0.4157
φ	0.8	0.03660	0.02018	94	0.5198	0.03218	0.01817	97	0.5843
<i>N</i> = 300									
Longitudinal Submodel									
β	0.5	-0.01129	0.00090	94	0.1092	-0.01723	0.00096	86	0.0901
σ_{ϵ}^2	0.5	0.00184	0.00046	95	0.0863	-0.00957	0.00069	88	0.0898
σ_u^2	0.8	-0.00849	0.00666	93	0.3188	-0.03294	0.00660	92	0.3047
Survival Submodel									
γ	-4.5	-0.02699	0.00859	95	0.3502	-0.03315	0.00850	94	0.3371
φ	0.8	0.02513	0.01211	96	0.4183	0.02192	0.01327	95	0.4715
<i>N</i> = 500									
Longitudinal Submodel									
β	0.5	-0.01022	0.00059	91	0.0845	-0.01590	0.00063	82	0.0700
σ_{ϵ}^2	0.5	0.00276	0.00030	94	0.0669	-0.00722	0.00036	92	0.0697
σ_u^2	0.8	-0.01134	0.00456	93	0.2462	-0.03742	0.00470	89	0.2349
Survival Submodel									
γ	-4.5	-0.02432	0.00525	93	0.2704	-0.04093	0.00661	90	0.2621
φ	0.8	0.02541	0.00803	93	0.3237	0.03171	0.00718	97	0.3672

Chapter 6

Joint Analysis under Misspecified Random Effects

6.1 Introduction

Note that random effects in joint models are usually assumed to follow a normal distribution, even though this choice is not made for computational simplicity. Many authors have questioned the Gaussian assumption, in the sense that the resulting inferences can be sensitive to assumptions not easily verifiable from available data; for instance, see discussion of Scharfstein et al. (1999) [67]. Other authors have proposed approaches that either make no parametric assumption about the random effects distribution (Tsiatis & Davidian, 2001 [78]) or relax the distributional assumptions (Song et al., 2002 [72]). In this chapter, we investigate the effects of misspecified the random effects distribution on inference with joint models. In particular, we study the empirical properties of the ML estimators in joint models with misspecified random effects distributions.

6.2 The Joint Model and Estimation

6.2.1 Two Submodels

Model for longitudinal data

Let y_{ij} be the continuous longitudinal measurement for subject i taken at time t_{ij} , $i = 1, 2, \dots, N; j = 1, 2, \dots, n_i$. Let $\mathbf{y}_i = \{y_i(t_{ij}), j = 1, 2, \dots, n_i\}$ be the observed longitudinal process for the i^{th} subject. We consider a linear mixed model for the process, given by

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{1}u_i + \boldsymbol{\varepsilon}_i, & i &= 1, 2, \dots, N, \\ u_i &\sim f(u_i|\sigma_u^2), & \boldsymbol{\varepsilon}_i &\sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_{n_i}), \end{aligned} \tag{6.1}$$

where $\mathbf{1}$ is the $n_i \times 1$ vector of 1's, $\boldsymbol{\beta}$ is the vector of fixed effects parameters, $\mathbf{X}_i$ is an $n_i \times p$ design matrix of fixed effects. The error terms $\boldsymbol{\varepsilon}_i$ are assumed to be normally distributed with zero mean and covariance matrix $\sigma_\varepsilon^2 \mathbf{I}_{n_i}$, where $\mathbf{I}_{n_i}$ denotes the n_i -dimensional identity matrix and σ_ε^2 is the error variance. Further, the errors $\boldsymbol{\varepsilon}_i$ are assumed to be independent of the random effects u_i .

Model for time-to-event data

Let s_i be the true survival time and c_i the censoring time. The observed survival times are defined as $t_i = \min(s_i, c_i)$, $i = 1, 2, \dots, N$. Let $\delta_i = I(s_i \leq c_i)$ be the censoring status, defined by $\delta_i = 1$ if $t_i = s_i$ and $\delta_i = 0$ if $t_i = c_i$. The survival times are assumed to follow the Weibull proportional hazards model

$$h_i(t) = \lambda \tau t_i^{\tau-1} \exp\{\mathbf{w}_i' \boldsymbol{\gamma} + \varphi u_i\}, \tag{6.2}$$

where $h_0(t) = \lambda \tau t^{\tau-1}$ is the baseline hazard, $\mathbf{w}_i$ denotes the vector of baseline covariates (or risk factors) with an associated coefficient vector $\boldsymbol{\gamma}$, and φ denotes a vector of association parameters linking the survival process with the longitudinal process.

6.2.2 The ML Estimation

Recall the joint model for two outcome processes

$$\begin{cases} \mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{1}u_i + \boldsymbol{\varepsilon}_i, \\ h_i(t_i) = \tau t_i^{\tau-1} \exp\{\mathbf{w}_i' \boldsymbol{\gamma} + \varphi u_i\}. \end{cases} \quad (6.3)$$

The likelihood function for all observed longitudinal and survival data is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^N \int f(t_i|u_i; \boldsymbol{\gamma}, \varphi) f(\mathbf{y}_i|u_i; \boldsymbol{\beta}, \sigma_\epsilon^2) f(u_i|\sigma_u^2) du_i, \quad (6.4)$$

where

$$f(t_i|u_i; \boldsymbol{\gamma}, \varphi) = \{\tau t_i^{\tau-1} \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \varphi u_i)\}^{\delta_i} \times \exp\left\{-\int_0^{t_i} t_i^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \varphi u_i) dt\right\},$$

$$f(\mathbf{y}_i|u_i; \boldsymbol{\beta}, \sigma_\epsilon^2) = (2\pi\sigma_\epsilon^2)^{-n_i/2} \exp\left\{-\frac{(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{1}u_i)'(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{1}u_i)}{2\sigma_\epsilon^2}\right\}.$$

The MLEs of the model parameters can be obtained by maximizing the corresponding log-likelihood, given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^N \log \left\{ \int f(t_i|u_i; \boldsymbol{\gamma}, \varphi) f(\mathbf{y}_i|u_i; \boldsymbol{\beta}, \sigma_\epsilon^2) f(u_i|\sigma_u^2) du_i \right\}. \quad (6.5)$$

The computation of the likelihood is not tedious as we consider a single frailty term for which the likelihood involves only one-dimensional integration. Thus, we find the exact ML estimators of the parameters in the joint model. Specifically, we use the R function "optim" to obtain the maximum likelihood (ML) estimates $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\gamma}}', \hat{\tau}, \hat{\varphi}, \hat{\sigma}_\epsilon^2, \hat{\sigma}_u^2)'$.

6.3 Simulation Study

In this section, we carry out a simulation study to evaluate the impact of a misspecified the random effects distribution on the parameter estimators in joint models. The models and parameter values used in the simulations are similar to those used in previous sections, except we assume that the random effects follow either a t-distribution or a normal distribution, where the true distribution of random effects is Chi-square with 3 degrees of freedom.

6.3.1 Generating longitudinal and survival data

The longitudinal outcomes were generated from the following mixed effects model with a random intercept

$$y_{ij} = \beta_0 + \beta_1 time_j + \beta_2 time_j \times x_{1i} + u_i + \epsilon_{ij}, \quad i = 1, 2, \dots, N, j = 1, 2, \dots, n_i, \quad (6.6)$$

where the random effects u_i were generated under two scenarios: (i) from a Chi-square distribution with 3 **d.f.**, so that $u_i \sim \sigma_u(\chi_3^2 - 3)/\sqrt{6}$, and (ii) from a Gamma distribution with parameters $a = 2$ and $s = 0.5$, that is $u_i \sim Gamma(a, s)$. The random errors ϵ_{ij} were generated from $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. The values of the covariate x_{1i} were generated as binary indicators ($x_{1i} = 0$ for control group and $x_{1i} = 1$ for treatment group). The follow-up times at which the longitudinal measurements were taken between the baseline time and survival time (t_i) were set in months ($time_j = 0, 1, 2, \dots$). The true survival times s_i were generated from the Weibull PH model

$$h_i(s_i|u_i, \lambda, \tau, \gamma, \varphi) = \tau s_i^{\tau-1} \exp(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \varphi u_i), \quad i = 1, 2, \dots, N, \quad (6.7)$$

where $\gamma_0 = \log(\lambda)$ and values of the baseline covariate x_{2i} were generated from $N(\mu_{x_2} = 2, \sigma_{x_2}^2 = 0.25)$. The observed survival times t_i were defined by $t_i = \min(s_i, c_i); (1, 2, \dots, N)$, where the non-informative censoring times c_i were generated from the same survival distribu-

tion as of s_i in (6.7), but with a different parameter value of $\gamma_0 = -6.5$. The parameter values are given as $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)' = (10.0, -0.2, 0.6)'$, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2, \tau)' = (-6.0, 1.5, 0.5, 0.8)'$, $\varphi = 0.5$, $(\sigma_\varepsilon^2, \sigma_u^2) = (0.5, 0.5)$. Each simulation run was based on $S = 500$ replications of data sets, with each data set containing $N = 100$ or 200 subjects.

Evaluation Criteria

We compare the performance of the methods based on empirical biases, mean squared errors (MSE), coverage probabilities (CP), and average lengths of 95% confidence intervals obtained by the ML method under the assumption of a normal and t random effects, whereas the true random effects distribution is a χ_3^2 distribution.

6.3.2 Simulation Results

Tables 6.1 and 6.2 present empirical biases, mean squared errors (MSEs), coverage probability (CPs) and average lengths of 95% confidence intervals (Length) obtained by the ML method. We study the performance of the ML method under two misspecified random effects distributions:

- (i) $u_i \sim N(0, \sigma_u^2)$,
- (ii) $u_i \simeq \sigma_u \sqrt{\frac{3}{5}} t_5$, t_5 is the t -distribution with 5 d.f.,

where the actual values of u_i are obtained under two scenarios: (i) from a Chi-square distribution with 3 d.f., (i.e., $u_i \sim \sigma_u(\chi_3^2 - 3)/\sqrt{6}$), and (ii) from a Gamma distribution (i.e., $u_i = \sigma_u(u_i^* - E(u_i^*)/\sqrt{V(u_i^*)})$, where u_i^* follows a Gamma distribution with the shape parameter $a = 2$ and scale parameter $s = 0.5$, so that $E(u_i^*) = a(s) = 1$ and $V(u_i^*) = a(s^2) = 0.5$. Here we consider both normal and t -distributions for random effects, as the normal distribution is commonly used for random effects, whereas the t -distribution is often used as a robust alternative to the normal to accommodate potential outliers in the tail areas. In both

cases, the random effects distribution is misspecified, and we are interested in finding out how the ML method is influenced by the misspecified random effects.

Table 6.1 presents the simulation results for the case when the true random effects distribution is Chi-square. Table 6.2 repeats the results for the case when the true random effects distribution is Gamma. It appears from the empirical results that under the misspecified random effects as considered here, the ML method provides biased estimates for some parameters. For example, when estimating the variance component σ_u^2 with $N = 100$, Table 6.1 under the true Chi-square distribution shows that the misspecified normal random effect provides a percentage relative bias of $(-0.0846/0.5) \times 100\% = -16.92\%$, whereas the t random effect provides a smaller percentage relative bias of $(-0.0126/0.5) \times 100\% = -2.52\%$ in absolute value. Also, for $N = 100$, Table 6.2 under the true Gamma distribution shows that the misspecified normal random effect provides a percentage relative bias of $(-0.0851/0.5) \times 100\% = -17.02\%$, whereas the t random effect provides a very small percentage relative bias of $(-0.0086/0.5) \times 100\% = -1.72\%$.

The difference between the estimates of σ_u^2 under the two settings is more evident in coverage probabilities (CPs). When $N = 100$ and the true distribution is Chi-square, Table 6.1 shows that the normal random effect gives a CP value of just 68.5%, whereas the t random effect gives a much better CP value of 86.5%. Also, when $N = 100$ and the true distribution is Gamma, Table 6.2 shows that the normal random effect gives a poor CP value of just 72.9%, whereas the t random effect gives a much better CP value of 89.5%. The gain in efficiency from the t -distribution, however, is not so obvious for estimating the regression parameters, as both methods give reasonably good coverage probabilities for the regression parameters.

The simulation results for larger samples with $N = 200$ follows a similar trend, as shown above. For example, when $N = 200$ and the true distribution is Gamma, Table 6.2 shows that the normal random effect gives a poor coverage probability of 63%, whereas the t random effect gives a much better coverage probability of 89%. We also performed a simulation

study for $N = 300$ (not shown here). As before, the misspecified normal random effects appeared to give systematic biases in the estimates of the variance component, and hence poor coverage probabilities of the confidence intervals.

From this empirical study, we can conclude that when the random effects distribution is misspecified (i.e., when the true distribution is asymmetric or non-normal), the t random effects can provide a better alternative to the commonly used normal random effects for analyzing longitudinal and survival data based on joint models.

Table 6.1: Empirical biases, mean squared errors, coverage probabilities (CPs), and average lengths of 95% confidence intervals obtained by the ML method under both normal and t distributions of random effects. True random effects follow a Chi-square distribution.

Par	True Value	Normal distribution				t distribution			
		Bias	MSE	CP%	Length	Bias	MSE	CP%	Length
$N = 100$									
Longitudinal Submodel									
β_0	10.0	-0.0110	0.0075	91.9	0.3167	-0.0724	0.0118	83.5	0.3100
β_1	-0.2	-0.0006	0.0014	94.9	0.1399	-0.0015	0.0013	94.5	0.1383
β_2	0.6	-0.0001	0.0004	94.4	0.0710	0.00003	0.0004	94.0	0.0702
σ_ϵ^2	0.5	0.0065	0.0009	96	0.1266	0.0043	0.0009	96.5	0.1258
σ_u^2	0.5	-0.0846	0.0197	69	0.3308	-0.0126	0.0172	86.5	0.4155
Survival Submodel									
γ_0	-6.0	-0.1515	1.1535	94.4	3.8440	-0.1647	1.1793	94.0	3.8435
γ_1	1.5	0.0390	0.1573	94.9	1.4051	0.0357	0.1555	95.0	1.4023
γ_2	0.5	0.0106	0.1462	92.9	1.3264	0.0113	0.1468	92.0	1.3232
τ	0.8	0.0200	0.0152	92.9	0.4270	0.0171	0.0150	93.0	0.4250
φ	0.5	0.0811	0.1279	93.9	1.2746	0.0300	0.1159	94.0	1.2373
$N = 200$									
Longitudinal Submodel									
β_0	10.0	-0.0218	0.0041	92.0	0.2222	-0.0719	0.0086	72.0	0.2179
β_1	-0.2	0.0001	0.0004	99.5	0.0950	-0.0008	0.0004	99.5	0.0940
β_2	0.6	-0.0008	0.0001	99.0	0.0483	-0.0007	0.0001	99.5	0.0478
σ_ϵ^2	0.5	0.0040	0.0004	97.5	0.0882	0.0020	0.0004	98.0	0.0877
σ_u^2	0.5	-0.0933	0.0143	53.8	0.2285	-0.0183	0.0082	87.5	0.2893
Survival Submodel									
γ_0	-6.0	-0.1678	0.4671	96.5	2.6833	-0.1831	0.4684	96.0	2.6843
γ_1	1.5	0.0459	0.0616	97.5	0.9786	0.0420	0.0600	97.5	0.9770
γ_2	0.5	0.0180	0.0573	97.0	0.9226	0.0192	0.0572	98.0	0.9207
τ	0.8	0.0166	0.0060	96.0	0.3004	0.0139	0.0059	95.0	0.2993
φ	0.5	0.1017	0.0576	93.5	0.8585	0.0641	0.0471	96.5	0.8304

Generated data: $u_i \sim \sigma_u(\chi_3^2 - 3)/\sqrt{6}$.

Fitted distribution: $u_i \sim N(0, \sigma_u^2)$ or $u_i \sim \sigma_u \sqrt{\frac{3}{5}} t_5$.

Table 6.2: Empirical biases, mean squared errors, coverage probabilities (CPs), and average lengths of 95% confidence intervals obtained by the ML method under both normal and t distributions of random effects. True random effects follow a Gamma distribution.

Par	True Value	Normal distribution				t distribution			
		Bias	MSE	CP%	Length	Bias	MSE	CP%	Length
<i>N</i> = 100									
Longitudinal Submodel									
β_0	10.0	-0.0121	0.0076	93.0	0.3169	-0.0591	0.0102	87.5	0.3121
β_1	-0.2	-0.0015	0.0012	93.5	0.1366	-0.0023	0.0013	93.5	0.1355
β_2	0.6	-0.0007	0.0003	94.0	0.0696	-0.0005	0.0003	94.5	0.0690
σ_ϵ^2	0.5	0.0056	0.0012	92.5	0.1259	0.0037	0.0011	92.5	0.1251
σ_u^2	0.5	-0.0851	0.0172	72.9	0.3295	-0.0086	0.0146	89.5	0.4261
Survival Submodel									
γ_0	-6.0	-0.3156	1.1809	95.5	3.8933	-0.3330	1.1922	95.0	3.8975
γ_1	1.5	0.0812	0.1661	93.0	1.4119	0.0811	0.1633	93.0	1.4103
γ_2	0.5	0.0513	0.1549	91.0	1.3300	0.0508	0.1538	91.0	1.3278
τ	0.8	0.0337	0.0142	95.5	0.4339	0.0322	0.0140	95.5	0.4329
φ	0.5	0.0743	0.1315	93.5	1.2961	0.0434	0.1230	94.5	1.2570
<i>N</i> = 200									
Longitudinal Submodel									
β_0	10.0	-0.0172	0.0040	91.0	0.2247	-0.0641	0.0075	78.0	0.2215
β_1	-0.2	-0.0031	0.0006	96.0	0.0938	-0.0038	0.0006	95.0	0.0930
β_2	0.6	0.0005	0.0002	95.0	0.0477	0.0006	0.0002	95.0	0.0473
σ_ϵ^2	0.5	0.0059	0.0006	92.9	0.0884	0.0042	0.0006	93.0	0.0880
σ_u^2	0.5	-0.0796	0.0125	63.0	0.2334	0.0001	0.0088	89.0	0.3005
Survival Submodel									
γ_0	-6.0	-0.2197	0.6069	92.0	2.6671	-0.2326	0.6165	91.5	2.6690
γ_1	1.5	0.0560	0.0634	94.0	0.9764	0.0536	0.0630	94.5	0.9752
γ_2	0.5	0.0287	0.0639	93.5	0.9057	0.0283	0.0644	94.5	0.9046
τ	0.8	0.0220	0.0075	91.5	0.3012	0.0199	0.0074	92.5	0.3003
φ	0.5	0.0864	0.0586	93.0	0.8486	0.0549	0.0502	95.0	0.8231

Generated data: $u_i \sim \text{Gamma}$.

Fitted distribution: $u_i \sim N(0, \sigma_u^2)$ or $u_i \sim \sigma_u \sqrt{\frac{3}{5}} t_5$.

Chapter 7

Application: GenIMS Study

7.1 Introduction

Severe sepsis is widespread and frequently fatal if not recognized early and managed promptly. Sepsis a syndromic response to a serious complication of infection caused by acute organ dysfunction (Singer, 2016 [71]). Recent scientific research estimated that in 2017 there were 48.9 million cases and 11 million sepsis-related deaths worldwide, leading to nearly 20% of all global deaths (Rudd, 2020 [65]). However, implementing preventive measures against infections in both the community and health care settings, such as ensuring access to vaccination programs, adequate hygiene practices, and improved water quality and sanitation, would be essential steps in reducing the occurrence of sepsis. Our research was motivated by a clinical study of subjects who were admitted to the emergency department (ED) with community-acquired pneumonia (CAP), the leading cause of severe sepsis, referred to as the Genetic and Inflammatory Markers of Sepsis (GenIMS) study. The goals of this study were to describe the systemic cytokine response to pneumonia and determine if there were specific patterns, including the balance of pro-inflammatory and anti-inflammatory markers, which were associated with severe sepsis and death (Kellum et al., 2007 [31]).

In this chapter, we present an analysis of the **GenIMS** data. The GenIMS study is

a large longitudinal cohort study of patients with community-acquired pneumonia (CAP) conducted in four states (Pennsylvania, Connecticut, Tennessee, and Michigan) of the USA from December 2001 to November 2003. The main objective of the GenIMS study was to investigate the association between the longitudinal response of a group of inflammatory and coagulation biomarkers and their effects on the survival outcomes of "pneumonia-to-death" of sepsis patients. In this study, antithrombin is one of the substantial coagulation biomarkers for understanding the mechanisms of pneumonia progression to sepsis. The patients were enrolled in the study through the emergency departments (ED) of US 28 academic and community hospitals.

The GenIMS study enrolled 2320 patients of age 18 years or older with a clinical and radiologic diagnosis of pneumonia. However, after discharge from emergency departments and changes in primary diagnosis, the inpatient CAP cohort comprised 1886 patients. During the study, deaths of inpatients were recorded, while patients' survivals who were discharged from the hospitals were determined by telephone follow-ups and from a National Death Index search. Among 1886 patients in the CAP cohort, 583 subjects (31%) developed severe sepsis and 212 subjects (11.24%) died (see Figure 7.1).

Among several goals of the study, a primary study goal was to examine the association between cytokines and the 90-day mortality rate of the patients. Blood samples were drawn for cytokine and coagulation assays starting at the enrollment of a patient, daily for the first 7 days, and weekly thereafter while the patient remained in the hospital. Only day 1 samples from subjects admitted after 11 PM or on weekends and holidays were not drawn for logistic reasons. Several cytokines were measured in this study: circulating levels of tumor necrosis factor (TNF) and IL-6 (interleukin-6) as markers of the pro-inflammatory, IL-10 (interleukin-10) as a marker of the anti-inflammatory cytokine response. The cohort study also collected patients' characteristics, demographics, comorbidities, etc. The 90-day mortality was used as a primary measure of survival. More details about the subject disposition (enrollment, exclusion, death) in the study can be found in Kellum et al.(2007) [31].

Figure 7.1: Subject disposition for the entire Genetic and Inflammatory Markers of Sepsis (GenIMS) cohort. CAP indicates community-acquired pneumonia [31].

For our study, we consider the joint modeling of longitudinal D-dimer (a coagulation biomarker) and 90 day mortality (time-to-event) of the patients, along with some covariates as potential confounders.

Before we describe the joint model for the aforementioned two outcomes, here we present some summary statistics of the outcomes. Figure (7.2) shows plots of individual D-dimer levels of fifteen patients over time. The individual D-dimer levels indicate distinct longitudinal trends with random intercepts and slopes for patients, which justify the use of a mixed model with random intercept and slope for explaining the D-dimer.

Figure 7.2: Logarithm of D-dimer levels of fifteen randomly selected sepsis patients over a week from GenIMS study.

Figure 7.3 presents boxplots of overall $\log(\text{D-dimer})$ levels drawn for four baseline covariates "Age", and group indicators of "Steroid", "Diabetic", and "Cardiac". Specifically, we dichotomize the patients into two age-groups: below 70 ($\text{age} \leq 70$) and over 70 ($\text{age} > 70$). Boxplots of D-dimer levels are shown for these two age groups. It is clear from the plots that

the D-dimer levels are generally higher for older patients. Also, patients who were treated with steroid appear to have lower D-dimer levels. Patients diagnosed with cardiac disease and diabetes appear to have slightly higher D-dimer levels.

Figure 7.3: The boxplot of overall D-dimer levels versus four baseline covariates age, steroid use, diabetes, and cardiac disease from GenIMS study.

Figure 7.4 shows Kaplan-Meier estimates of survival probabilities for two groups of D-dimer levels, $\log(\text{D-dimer}) \leq 6.5$ and $\log(\text{D-dimer}) > 6.5$, at the baseline time point. It is clear from the Kaplan-Meier estimates that the survival probabilities are different for the two D-dimer levels, where patients with higher D-dimer levels are at higher risk of death.

Figure 7.4: Kaplan-Meier estimates of the survival probability for two groups of D-dimer levels: $\log(\text{D-dimer}) \leq 6.5$ and $\log(\text{D-dimer}) > 6.5$.

We also study the nonparametric Kaplan-Meier estimates of the survivor function for different demographic groups of the sepsis patients. Figure 7.5 exhibits estimated survival probabilities for two age groups ($\text{age} \leq 70$ vs. $\text{age} > 70$), as well as for the group indicators of steroid use (Steroid vs. No steroid), diabetes (Diabetic vs. No diabetic), and heart disease (Cardiac vs. No cardiac). The survival probabilities appears to be lower for older patients and also lower among patients who use steroid, who are diabetic, and who have cardiac disease. In the next section, we perform a joint analysis to assess the significance of these covariates and D-dimer biomarker for predicting the survival probabilities.

Figure 7.5: Kaplan-Meier estimates of the survivor function for different groups of sepsis patients from GenIMS study. The estimates appears to be different among the groups.

7.2 The Joint Models

Let y_{ij} be the longitudinal measurement on the D-dimer biomarker of the i^{th} patient observed at time, $Time_{ij}$, $i = 1, \dots, N$ and $j = 1, \dots, n_i$. Let t_i be the survival time (90-day mortality) of the i^{th} pneumonia patient. The censoring status was defined as $\delta_i = 1$ if a patient died and $\delta_i = 0$ if a patient is still alive. Among the predictors considered in the study, "Steroid" represents an indicator variable, which is defined to be 1 if a patient used steroid, and 0 if not. Similarly, the binary covariates "Respdis", "Carddis", and "Diab" are used to indicate the disease status of a patient at baseline for respiratory disease, cardiac disease, and diabetes, respectively. "Age" denotes the baseline age of a patient in years (18 years or older), "Sex" is labeled as 1 for male and 2 for female, and "Time" represents the follow-up times (in days) at which the longitudinal measurements were taken between the baseline time and survival time t_i . Our goal is to investigate how these covariates affect the

longitudinal response of D-dimer measurements as well as the risk to death.

Model for Longitudinal D-dimer Data

For modeling the longitudinal D-dimer levels y_{ij} , we consider a linear mixed effects model with random intercept and slope terms, given by

$$y_{ij} = \beta_0 + \beta_1 Age_i + \beta_2 Time_{ij} + \beta_3 Steroid_i \times Time_{ij} + \beta_4 Carddise_i + \beta_5 Diab_i + u_{0i} + u_{1i} Time_{ij} + \varepsilon_{ij}, \quad (7.1)$$

for $i = 1, 2, \dots, N, j = 1, 2, \dots, n_i$, where the random intercept u_{0i} and random slope u_{1i} are assumed to follow a bivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{D}$. Here, the elements in matrix $\mathbf{D}$ are $d_{11} = Var(u_{0i}) = \sigma_{u_{11}}^2$, $d_{22} = Var(u_{1i}) = \sigma_{u_{22}}^2$, and $d_{12} = d_{21} = Cov(u_{0i}, u_{1i}) = \sigma_{u_{12}}$. Moreover, the random error term ε_{ij} is assumed to follow a normal distribution with mean 0 and variance σ_ε^2 , and is assumed to be independent of the random effects terms $\mathbf{u}_i = (u_{0i}, u_{1i})'$.

Model for Survival Data

For modeling the survival outcome t_i of the 90-day mortality, we consider a frailty model in the framework of the Weibull accelerated failure time model with the baseline covariates Age, Sex, Steroid use, Cardiac disease, and Diabetes. The model is given by

$$h_i(t_i) = \exp(-\eta_i) h_0\{t_i / \exp(-\eta_i)\}, \quad (7.2)$$

with the baseline hazard $h_0(t_i) = \lambda \tau t_i^{\tau-1}$ and the linear predictor $\eta_i = \gamma_1 Age_i + \gamma_2 Sex_i + \gamma_3 Steroid_i + \gamma_4 Cardiac_i + \gamma_5 Diab_i + \varphi_1 u_{0i} + \varphi_2 u_{1i}$. In this case, Eq.(7.2) can be rewritten

in the form

$$\begin{aligned}
h_i(t_i) &= \exp(-\eta_i) \lambda \tau t_i^{\tau-1} \exp\{-(\tau-1)\eta_i\} \\
&= \tau t_i^{\tau-1} \exp(\log \lambda - \tau \eta_i) \\
&= \tau t_i^{\tau-1} \exp\{-\tau(\gamma_0 + \eta_i)\} \\
&= \tau t_i^{\tau-1} \exp\{-\eta_i^*\}^\tau,
\end{aligned} \tag{7.3}$$

where $\gamma_0 = -\log(\lambda^{1/\tau})$, $\eta_i^* = \gamma_0 + \eta_i$ and the association parameters $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)'$ connect the frailty model (7.3) to the mixed-effects model (7.1) through the random effects terms $\mathbf{u}_i = (u_{0i}, u_{1i})'$.

7.3 ML Estimation

We now define our joint model for the longitudinal outcome y_{ij} and time-to-event outcome t_i . The observed data are given by $\{(y_{ij}, t_i, \delta_i), i = 1, 2, \dots, N, j = 1, 2, \dots, n_i\}$. The joint likelihood for the observed data is obtained as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^N \left\{ \int f(t_i, \delta_i | \mathbf{u}_i; \tau, \gamma, \boldsymbol{\varphi}) f(\mathbf{y}_i | \mathbf{u}_i; \boldsymbol{\beta}, \sigma_\varepsilon^2) f(\mathbf{u}_i | \mathbf{D}) d\mathbf{u}_i \right\}, \tag{7.4}$$

where:

- The conditional density function of the i^{th} response vector $\mathbf{y}_i = (y_{i1}, \dots, y_{n_i})'$ given the random effects $\mathbf{u}_i$ is given by

$$f(\mathbf{y}_i | \mathbf{u}_i; \boldsymbol{\beta}, \sigma_\varepsilon^2) = (2\pi)^{-n_i/2} |\sigma_\varepsilon^2 \mathbf{I}_{n_i}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{u}_i)' (\sigma_\varepsilon^2 \mathbf{I}_{n_i})^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{u}_i) \right\},$$

with $\mathbf{x}_{ij}' \boldsymbol{\beta} = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Sex}_i + \beta_3 \text{Steroid}_i + \beta_4 \text{Cardiac}_i + \beta_5 \text{Diab}_i$ and $\mathbf{z}_{ij}' \mathbf{u}_i = u_{0i} + u_{1i} \text{Time}_{ij}$, and $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$ being the j^{th} rows of the design matrix $\mathbf{X}_i$ for fixed effects and $\mathbf{Z}_i$ for random effects, respectively.

- The conditional density function of the observed survival time (t_i, δ_i) for the i^{th} individual may be written as

$$f(t_i|\mathbf{u}_i; \gamma, \tau, \boldsymbol{\varphi}) = \{h(t_i|\mathbf{u}_i; \gamma, \tau, \boldsymbol{\varphi})\}^{\delta_i} S(t_i|\mathbf{u}_i; \gamma, \tau, \boldsymbol{\varphi}),$$

with $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)'$ and $h(t_i|\mathbf{u}_i; \gamma, \tau, \boldsymbol{\varphi})$ and $S(t_i|\mathbf{u}_i; \gamma, \tau, \boldsymbol{\varphi})$ being the conditional hazard and survivor functions for the i^{th} individual at time t_i , respectively. More specifically, the hazard function for the i^{th} individual at time t_i is given by the accelerated failure time (AFT) model (7.2), where the linear predictor $\eta_i = \gamma_1 Age_i + \gamma_2 Sex_i + \gamma_3 Steroid_i + \gamma_4 Cardiac_i + \gamma_5 Diab_i + \varphi_1 u_{0i} + \varphi_2 u_{1i}$.

- The density function of the random effects is given by

$$f(\mathbf{u}_i|\mathbf{D}) = (2\pi)^{-q/2} |\mathbf{D}|^{-1/2} \exp\{\mathbf{u}_i' \mathbf{D}^{-1} \mathbf{u}_i\}, \quad (q = 2),$$

with the variance-covariance matrix

$$\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} = \begin{bmatrix} Var(u_{0i}) & Cov(u_{0i}, u_{1i}) \\ Cov(u_{0i}, u_{1i})' & Var(u_{1i}) \end{bmatrix} = \begin{bmatrix} \sigma_{u_{11}}^2 & \sigma_{u_{12}} \\ \sigma_{u_{21}} & \sigma_{u_{22}}^2 \end{bmatrix}.$$

The association parameters $\boldsymbol{\varphi}$ and shared random effects $\mathbf{u}_i$ connect the two different outcome processes and their respective models. When all association parameters are zero, i.e., $\boldsymbol{\varphi} = 0$, the joint analysis is equivalent to the separate analyses of the two outcome processes. However, for nonzero $\boldsymbol{\varphi}$, we expect to achieve gain in efficiency from the joint analysis. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \boldsymbol{\varphi}, \sigma_\varepsilon^2, \mathbf{D})$ be the vector of regression parameters and variance components in the joint model. We can find the maximum likelihood (ML) estimators of the model parameters by numerically maximizing the joint likelihood (7.4). Specifically, the integration involved in evaluating the likelihood (7.4) was performed numerically using an adaptive Gaussian quadrature method. The maximization of the likelihood was performed

using a dual quasi-Newton algorithm, where we set the relative gradient convergence criterion to zero that leads to gradient values of all parameter estimates close to zero. We use the SAS NLMIXED procedure for estimating the model parameters. The iterative procedure took about 20 minutes CPU time in a PC that has Windows 10 with a 64-bit operating system, Intel Core i7-6650U CPU @ 2.21GHz and 16.0 GB RAM.

7.4 Results

Table 7.1 presents the maximum likelihood estimates, their standard errors, t-values and individual 95% confidence intervals of the model parameters based on both joint and separate modeling approaches. It is obvious from the table that the shared random effects variance components are significant, with their estimates $(\hat{\sigma}_{u_{11}}^2, \hat{\sigma}_{u_{22}}^2) = (0.9142, 0.0217)$ and corresponding standard errors $(0.0451, 0.0015)$. Also, the covariance term $(\sigma_{u_{21}})$ that links the two random effects is significant with the estimate -0.0556 and its corresponding standard error 0.0067 .

An important focus of the joint analysis is the association parameter φ , which is used to assess the strength of association between the two submodels. The association parameters $\varphi = (\varphi_1, \varphi_2)'$ are estimated as $\hat{\varphi} = (-1.0435, -5.5962)'$ with small standard errors $SE(\hat{\varphi}) = (0.2081, 1.3054)'$, justifying the use of the joint modeling approach for the data analysis. The separate model gives a corresponding frailty variance of 6.7435 (95%CI : $4.7361, 8.7510$).

From the linear mixed model for D-dimer, it appears that the levels are higher among older patients. Also, the D-dimer level increases over time, with an estimated time effect of 0.0357 and a corresponding standard error of 0.0075 , as obtained by the joint analysis. The interaction effect of $(Steroid \times Time)$ also appears to be significant by the joint analysis. It is also obvious from the joint analysis that patients with no cardiac disease have lower D-dimer levels, while the D-dimer level increases for diabetic patients, although not statistically significant.

From the frailty model for 90-day mortality of the pneumonia patients, Table 7.1 shows that the estimated value of the shape parameter τ is 0.6600, with a small standard error of 0.0681. In addition, older patients, females, patients treated with steroid, and patients with diabetes and cardiac disease are all at the higher risk of 90-day mortality in the cohort of community acquired pneumonia patients.

Finally, with regards to comparing the joint estimates with their separate counterparts, the two sets of parameter estimates appear to be somewhat different. For example, for the group indicator of cardiac disease, the joint model provides an estimate of -0.0224 with a standard error of 0.3873, whereas the separate model provides an estimate of 0.0789 with a standard error of 0.2706.

To assess the validity of the Weibull accelerated failure time model used in our survival analysis, we performed a residual analysis based on the widely used Cox-Snell residuals, $r_{Ci} = -\log \hat{S}_i(t_i)$, where $\hat{S}_i(t_i)$ is an estimate of the survivor function $S_i(t_i)$, given by

$$\hat{S}_i(t_i) = \exp[-\{\exp(-\hat{\eta}_i^*)t_i\}^\tau],$$

with $\hat{\eta}_i^* = \hat{\gamma}_0 + \hat{\gamma}_1 \text{Age}_i + \hat{\gamma}_2 \text{Sex}_i + \hat{\gamma}_3 \text{Steroid}_i + \hat{\gamma}_4 \text{Cardiac}_i + \hat{\gamma}_5 \text{Diab}_i + \hat{\varphi}_1 \hat{u}_{0i} + \hat{\varphi}_2 \hat{u}_{1i}$. Here $\hat{\mathbf{u}}_i = (\hat{u}_{0i}, \hat{u}_{1i})'$ are predictors of the random effects $\mathbf{u}_i = (u_{0i}, u_{1i})'$, given by

$$\hat{\mathbf{u}}_i = E(\mathbf{u}_i | t_i, \mathbf{y}_i) = \int_{\mathbf{u}_i} \mathbf{u}_i f(\mathbf{u}_i | t_i, \mathbf{y}_i) d\mathbf{u}_i,$$

where $f(\mathbf{u}_i | t_i, \mathbf{y}_i)$ is the conditional density of $\mathbf{u}_i$ given the outcomes $(t_i, \mathbf{y}_i)$, which may be calculated from

$$f(\mathbf{u}_i | t_i, \mathbf{y}_i) = \frac{f(t_i | \mathbf{u}_i) f(\mathbf{y}_i | \mathbf{u}_i) f(\mathbf{u}_i)}{\int_{\mathbf{u}_i} f(t_i | \mathbf{u}_i) f(\mathbf{y}_i | \mathbf{u}_i) f(\mathbf{u}_i) d\mathbf{u}_i}.$$

The above expectation involves computation of two-dimensional integrations with respect to the random effects $\mathbf{u}_i = (u_{0i}, u_{1i})'$. We use a numerical integration technique to evaluate the above conditional expectation with respect to $\mathbf{u}_i | (t_i, \mathbf{y}_i)$.

Figure 7.6: Residual analysis of the survival submodel for GenIMS data. Top-left panel exhibits Q-Q plot of Cox-Snell residuals. Other three panels show scatterplots of Cox-Snell residuals against three covariates Age, Steroid use, and Diabetes.

The Cox-Snell residuals $r_{C_i} = -\log \hat{S}_i(t_i)$ obtained from the predictors of fixed and random effects are presented in Figure 7.6. Under a correctly specified model, the Cox-Snell residuals generally follow a unit exponential distribution. So to check the validity of our survival submodel, here we investigate if the Cox-Snell residuals represent a plausible sample from the unit exponential distribution. The top-left panel of Figure 7.6 presents a Q-Q plot of the Cox-Snell residuals r_{C_i} in which the cumulative hazards $H_i(r_{C_i})$ of the residuals are plotted against the residuals r_{C_i} . It is obvious that the plot displays a straight line pattern except for two points in the top-right corner, which may indicate potential outliers in the data. These two points are also visible as potential outliers in the residual plots against the covariates Age, Steroid, and Diabetic as can be seen in the remaining three panels of Figure 7.6. The plots, however, do not indicate any systematic pattern that would cast doubt on

Table 7.1: Joint analyses of pneumonia-to-death and longitudinal D-dimer outcomes from GenIMS study based on ML method.

Parameter	Separate Models				Joint Model			
	Estimate	SE	t-value	LL	UL	Estimate	SE	t-value
Longitudinal Submodel								
Intercept (β_0)	5.8372	0.1101	53.01	5.6211	6.0532	5.8365	0.1102	52.99
Age (β_1)	0.0837	0.0163	5.13	0.0517	0.1157	0.0839	0.0163	5.14
Time (β_2)	0.0364	0.0075	4.87	0.0217	0.0510	0.0357	0.0075	4.78
Time $\times$ Steroid (β_3)	-0.0485	0.0108	-4.50	-0.0696	-0.0273	-0.0480	0.0108	-4.45
Cardiac (β_4)	-0.1125	0.0691	-1.61	-0.2498	0.0247	-0.1136	0.0700	-1.62
Diabetic (β_5)	0.0387	0.0775	0.50	-0.1133	0.1906	0.0403	0.0775	0.52
σ_ϵ^2	0.0958	0.0026	37.32	0.0908	0.1008	0.0958	0.0026	37.32
$\sigma_{u_{11}}^2$	0.9148	0.0452	20.26	0.8262	1.0034	0.9142	0.0451	20.26
$\sigma_{u_{21}}$	-0.0551	0.0067	-8.36	-0.0691	-0.0428	-0.0556	0.0067	-8.32
$\sigma_{u_{22}}^2$	0.0217	0.0015	14.35	0.0188	0.0247	0.0217	0.0015	14.35
Survival Submodel								
Intercept (γ_0)	16.3714	1.1243	14.56	14.1666	18.5762	17.1927	1.9191	8.96
Age (γ_1)	-0.9334	0.1079	-8.65	-1.1450	-0.7218	-1.0206	0.1867	-5.47
Sex (γ_2)	-0.5529	0.2441	-2.26	1.0316	-0.0742	-0.6384	0.3479	-1.84
Steroid (γ_3)	-0.5854	0.2471	-2.37	-1.0700	-0.1008	-0.5874	0.3447	-1.70
Cardiac (γ_4)	0.0789	0.2706	0.29	-0.4518	0.6095	-0.0224	0.3873	-0.06
Diabetic (γ_5)	-0.7828	0.2914	-2.69	-1.3543	-0.2114	-1.1705	0.4098	-2.86
Shape (τ)	1.2114	0.1811	6.69	0.8563	1.5665	0.6600	0.0681	9.68
σ_u^2	6.7435	1.0237	6.59	4.7361	8.7510			
φ_1						-1.0435	0.2081	-5.02
φ_2						-5.5962	1.3054	-4.29

the validity of the Weibull model as used here. Also to assess the assumptions of the linear mixed-effects model, we analyze the conditional residuals $e_{ij} = y_{ij} - \hat{y}_{ij}$, where $\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 Age_i + \hat{\beta}_2 Time_{ij} + \hat{\beta}_3 Steroid_i \times Time_{ij} + \hat{\beta}_4 Carddise_i + \hat{\beta}_5 Diab_i + \hat{u}_{0i} + \hat{u}_{1i} Time_{ij}$. The top-left panel of Figure 7.7 shows the scatterplot of standardized residuals $r_{ij} = e_{ij}/\hat{\sigma}_\varepsilon = (y_{ij} - \hat{y}_{ij})/\hat{\sigma}_\varepsilon$ over seven follow-up days (Day 1-7). The residual plot does not indicate any systematic pattern that would cast doubt on the validity of the linear mixed model as used here. We also present normal quantile-quantile (Q-Q) plots of the standardized residuals $r_{ij} = e_{ij}/\hat{\sigma}_\varepsilon$ over five observation days (Day 1-5) in Figure 7.7. The Q-Q plots do not indicate any clear violation of the normality assumption, as used for the random error term.

Figure 7.7: Analysis of standardized residuals $r_{ij} = e_{ij}/\hat{\sigma}_\varepsilon = (y_{ij} - \hat{y}_{ij})/\hat{\sigma}_\varepsilon$ obtained from the linear mixed model for D-dimer level. Top-left panel shows scatterplot of r_{ij} against time (Day 1-7). Remaining five panels show Q-Q plots of the standardized residuals for five follow-up times (Day 1-5).

Chapter 8

Conclusions and Future Work

Joint modelling of longitudinal and survival data has received much attention in the literature in recent years, especially in biomedical studies.

Our research may be extended in several directions:

- The association among different types of longitudinal processes can be modelled by correlated or shared random effects. This association may also be modeled in other ways. For example, we may assume that different types of longitudinal processes may be governed by the same latent disease processes, which introduce the association between the longitudinal processes.
- The MLEs of the parameters in the joint model can also be obtained by Monte Carlo EM algorithms, treating the random effects as "missing data". Although such EM algorithms may be computationally intensive, it may be desirable to compare them with the h-likelihood method in terms of efficiency and computational simplicity.
- In the joint models, the Cox model for survival data may be replaced by an accelerated failure time (AFT) model. The extension is conceptually straightforward. In this case, the baseline hazard may involve unknown parameters, but we may consider parametric models such as AFT models based on Gumbel distributions. The Cox model may also be replaced by survival models for interval censored data or competing risks data.

- In practice, there may be dropouts or missing data in the clinical trials, and these dropouts or missingness may be associated with the longitudinal patterns of responses. Thus, we may consider incorporating missing data mechanisms into the joint models in future research.
- The longitudinal data may also contain outliers, so we may use robust joint model by applying the M-estimator for analyzing longitudinal data.
- Lastly, our proposed approximate h-likelihood and likelihood methods may be extended to joint models with left censored covariates often observed due to the limit of detection.

In conclusion, many issues still remain for joint modeling of longitudinal and survival data. We plan to investigate these issues in our future research.

Appendix A

Estimation Procedure Based On H-likelihood Method

A.1 The (Joint) h-likelihood function

The likelihood function

$$\begin{aligned} L(\theta) &= f(t_i, \delta_i, \mathbf{y}_i; \theta) \\ &= \int f(t_i, \delta_i | \mathbf{u}_i) f(\mathbf{y}_i | \mathbf{u}_i) f(\mathbf{u}_i) d\mathbf{u}_i \\ &= \int \left[\{f(t_i, \delta_i | \mathbf{u}_i)\}^{\delta_i} \{S(t_i | \mathbf{u}_i)\}^{1-\delta_i} \right] \left[\prod_j f(y_{ij} | \mathbf{u}_i) \right] f(\mathbf{u}_i) d\mathbf{u}_i \\ &= \int \left[\{h(t_i, \delta_i | \mathbf{u}_i)\}^{\delta_i} S(t_i | \mathbf{u}_i) \right] \left[\prod_j f(y_{ij} | \mathbf{u}_i) \right] f(\mathbf{u}_i) d\mathbf{u}_i. \end{aligned} \tag{A.1}$$

The log-likelihood:

$$\ell(\theta) = \log \left[\int f(t_i, \delta_i | \mathbf{u}_i) f(\mathbf{y}_i | \mathbf{u}_i) f(\mathbf{u}_i) d\mathbf{u}_i \right]. \tag{A.2}$$

Following Lee and Nelder (1996, [36]) and Ha et al.(2001, [12]), h-likelihood for the joint

model, denoted by h , is given by

$$h = h(t_i, \delta_i, y_i; \boldsymbol{\theta}, \boldsymbol{\xi}) = \prod_i h_{1i} \times h_{2i} \times \ell_{3i}, \quad (\text{A.3})$$

where

$$\begin{aligned} h_{1i} = f(t_i \delta_i | \mathbf{u}_i; \boldsymbol{\gamma}, \tau, \boldsymbol{\varphi}) &= \{\tau t_i^{\tau-1} \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i)\}^{\delta_i} \times \exp \left\{ - \int_0^{s_i} t_i^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) dt \right\}, \\ h_{2i} = f(\mathbf{y}_i | \mathbf{u}_i; \boldsymbol{\beta}, \sigma_\epsilon^2) &= (2\pi\sigma_\epsilon^2)^{-n_i/2} \exp \left\{ - \frac{[\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{u}_i]' [\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{u}_i]}{2\sigma_\epsilon^2} \right\}, \\ h_{3i} = f(\mathbf{u}_i | \mathbf{D}) &= \frac{\exp\{-\frac{1}{2} \mathbf{u}_i' \mathbf{D}^{-1} \mathbf{u}_i\}}{(2\pi)^{q/2} |\mathbf{D}|^{1/2}}. \end{aligned}$$

The corresponding logarithm of h-likelihood for the joint model (A.3), denoted by ℓ_h , is given by

$$\begin{aligned} \ell_h = \sum_i \ell_{h_i} &= \sum_i \ell_{h_{1i}} + \sum_i \ell_{h_{2i}} + \sum_i \ell_{h_{3i}} \\ &= \left\{ \sum_i \log f(t_i, \delta_i | \mathbf{u}_i; \tau, \boldsymbol{\gamma}, \boldsymbol{\varphi}) \right\} + \left\{ \sum_i \log f(\mathbf{y}_i | \mathbf{u}_i; \boldsymbol{\beta}, \sigma_2) \right\} + \left\{ \sum_i \log f(\mathbf{u}_i; \mathbf{D}) \right\} \\ &= \sum_i \delta_i \log(\tau) + (\tau - 1) \sum_i \delta_i \log(t_i) + \sum_i \delta_i (\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) - \sum_i t_i^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \\ &\quad + \sum_i \left\{ -\frac{n_i}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}_i| - \frac{1}{2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{u}_i)' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{u}_i) \right\} \\ &\quad + \sum_i \left\{ -\frac{q}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{D}| - \frac{1}{2} \mathbf{u}_i' \mathbf{D}^{-1} \mathbf{u}_i \right\}. \end{aligned} \quad (\text{A.4})$$

A.1.1 Estimation for $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\varphi}')'$

We now develop an estimation procedure via the h-likelihood (A.4). Given the dispersion parameters $\boldsymbol{\xi} = (\sigma_\epsilon^2, \mathbf{D})$, the maximum h-likelihood estimators (MHLEs) for $\boldsymbol{\theta}$ are obtained

by maximizing the following adjusted profile h-likelihood with respect to $\boldsymbol{\theta}$:

$$\begin{aligned}
p_{\boldsymbol{\theta}} &= \left[\ell_h - 0.5 \log \left| \frac{H[(\ell_h, \mathbf{u})]}{2\pi} \right| \right] \Big|_{\xi=\hat{\xi}, \mathbf{u}=\hat{\mathbf{u}}} \\
&= \left[\ell_h - 0.5 \log \left(\frac{1}{2\pi} \left| H[(\ell_h, \mathbf{u})] \right| \right) \right] \Big|_{\xi=\hat{\xi}, \mathbf{u}=\hat{\mathbf{u}}} \\
&= \left[\ell_h - 0.5 \left(\log \left(\frac{1}{2\pi} \right) + \log \left| H[(\ell_h, \mathbf{u})] \right| \right) \right] \Big|_{\xi=\hat{\xi}, \mathbf{u}=\hat{\mathbf{u}}}.
\end{aligned} \tag{A.5}$$

The maximum adjusted profile h-likelihood, $p_{\boldsymbol{\theta}}$, for estimators for are obtained by solving iteratively

$$\begin{aligned}
S(\boldsymbol{\theta}) = \frac{\partial p_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} &= \frac{\partial \ell_h}{\partial \boldsymbol{\theta}} - 0.5 \frac{\partial \log |H^*(\boldsymbol{\theta})|}{\partial \boldsymbol{\theta}} \\
&= \frac{\partial \ell_h}{\partial \boldsymbol{\theta}} - 0.5 \frac{1}{|H^*(\boldsymbol{\theta})|} \frac{\partial |H^*(\boldsymbol{\theta})|}{\partial \boldsymbol{\theta}} = 0,
\end{aligned} \tag{A.6}$$

where

$$\frac{\partial \ell_h}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \ell_h}{\partial \boldsymbol{\beta}} \\ \frac{\partial \ell_h}{\partial \boldsymbol{\gamma}} \\ \frac{\partial \ell_h}{\partial \boldsymbol{\tau}} \\ \frac{\partial \ell_h}{\partial \boldsymbol{\varphi}} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{Y}_i - \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \boldsymbol{\beta} - \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{u}_i \\ \sum_{i=1}^N \delta_i \mathbf{w}_i - \sum_{i=1}^N t_i^{\tau} \mathbf{w}_i \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \\ \sum_{i=1}^N \frac{\delta_i}{\tau} + \sum_{i=1}^N \delta_i \log(t_i) - \sum_{i=1}^N t_i^{\tau} \log(t_i) \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \\ \sum_{i=1}^N \delta_i \mathbf{u}_i - \sum_{i=1}^N \mathbf{u}_i t_i^{\tau} \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \end{bmatrix},$$

$$\begin{aligned}
\frac{\partial \ell_h}{\partial \mathbf{u}_i} &= \mathbf{Z}_i' \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{u}_i) + \boldsymbol{\varphi}' \delta_i - \boldsymbol{\varphi}' t_i^{\tau} \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \\
&\quad - \mathbf{u}_i \mathbf{D}^{-1},
\end{aligned} \tag{A.7}$$

$$\frac{\partial^2 \ell_h}{\partial \mathbf{u}_i \partial \mathbf{u}_i'} = -\mathbf{Z}_i' \mathbf{V}_i^{-1} \mathbf{Z}_i - \boldsymbol{\varphi} \boldsymbol{\varphi}' t_i^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) - \mathbf{D}^{-1}, \quad (\text{A.8})$$

$$\begin{aligned} |H^*(\boldsymbol{\theta})| &= \left| H[(\ell_h, \mathbf{u})] \right| \\ &= \left| -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \mathbf{u}'} \right| \\ &= \left| -\left(-\sum_{i=1}^N \mathbf{Z}_i' \mathbf{V}_i^{-1} \mathbf{Z}_i - \boldsymbol{\varphi} \boldsymbol{\varphi}' \sum_i t_i^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) - \mathbf{D}^{-1} \right) \right| \\ &= \left| \sum_{i=1}^N \mathbf{Z}_i' \mathbf{V}_i^{-1} \mathbf{Z}_i + \boldsymbol{\varphi} \boldsymbol{\varphi}' \sum_i t_i^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) + \mathbf{D}^{-1} \right| = H_i, \end{aligned} \quad (\text{A.9})$$

$$\begin{aligned} \frac{\partial |H^*(\boldsymbol{\theta})|}{\partial \boldsymbol{\theta}} &= \frac{\partial H_i}{\partial \boldsymbol{\theta}} \\ &= \begin{bmatrix} \frac{\partial H_i}{\partial \boldsymbol{\beta}} \\ \frac{\partial H_i}{\partial \boldsymbol{\gamma}} \\ \frac{\partial H_i}{\partial \boldsymbol{\tau}} \\ \frac{\partial H_i}{\partial \boldsymbol{\varphi}} \end{bmatrix} = \begin{bmatrix} 0 \\ \boldsymbol{\varphi} \boldsymbol{\varphi}' t_i^\tau \mathbf{w}_i \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \\ \boldsymbol{\varphi} \boldsymbol{\varphi}' t_i^\tau \log(t_i) \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \\ \boldsymbol{\varphi} \boldsymbol{\varphi}' t_i^\tau \mathbf{u}_i \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) + 2\boldsymbol{\varphi} t_i^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \end{bmatrix} \end{aligned} \quad (\text{A.10})$$

Then, use Newton-Raphson equation to solve MHL estimation of equation (A.5) as follows

$$\hat{\boldsymbol{\theta}}^{(m+1)} = \hat{\boldsymbol{\theta}}^{(m)} + \left(-\frac{\partial^2 p_\theta}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)^{-1} \frac{\partial p_\theta}{\partial \boldsymbol{\theta}}, \quad (\text{A.11})$$

Replaced by

$$\hat{\boldsymbol{\theta}}^{(m+1)} = \hat{\boldsymbol{\theta}}^{(m)} + \left(-\frac{\partial^2 \ell_h}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)^{-1} \frac{\partial p_\theta}{\partial \boldsymbol{\theta}}. \quad (\text{A.12})$$

where the score function is given by $\frac{\partial \ell_h}{\partial \boldsymbol{\theta}}$.

$$\frac{\partial \ell_h}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \ell_h}{\partial \boldsymbol{\beta}} \\ \frac{\partial \ell_h}{\partial \boldsymbol{\gamma}} \\ \frac{\partial \ell_h}{\partial \tau} \\ \frac{\partial \ell_h}{\partial \boldsymbol{\varphi}} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{Y}_i - \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} X_i \boldsymbol{\beta} - \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{u}_i \\ \sum_{i=1}^N \delta_i \mathbf{w}_i - \sum_{i=1}^N t_i^\tau \mathbf{w}_i \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \\ \sum_{i=1}^N \frac{\delta_i}{\tau} + \sum_{i=1}^N \delta_i \log(t_i) - \sum_{i=1}^N t_i^\tau \log(t_i) \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \\ \sum_{i=1}^N \delta_i \mathbf{u}_i - \sum_{i=1}^N \mathbf{u}_i t_i^\tau \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \end{bmatrix},$$

and the fisher information is given by

$$I(\boldsymbol{\theta}) = -E \left(\frac{\partial^2 \ell_h}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) = \begin{bmatrix} -E \left(\frac{\partial^2 \ell_h}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right) & -E \left(\frac{\partial^2 \ell_h}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}'} \right) & -E \left(\frac{\partial^2 \ell_h}{\partial \boldsymbol{\beta} \partial \tau'} \right) & -E \left(\frac{\partial^2 \ell_h}{\partial \boldsymbol{\beta} \partial \boldsymbol{\varphi}'} \right) \\ -E \left(\frac{\partial^2 \ell_h}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}'} \right) & -E \left(\frac{\partial^2 \ell_h}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} \right) & -E \left(\frac{\partial^2 \ell_h}{\partial \boldsymbol{\gamma} \partial \tau'} \right) & -E \left(\frac{\partial^2 \ell_h}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\varphi}'} \right) \\ -E \left(\frac{\partial^2 \ell_h}{\partial \tau \partial \boldsymbol{\beta}'} \right) & -E \left(\frac{\partial^2 \ell_h}{\partial \tau \partial \boldsymbol{\gamma}'} \right) & -E \left(\frac{\partial^2 \ell_h}{\partial \tau \partial \tau'} \right) & -E \left(\frac{\partial^2 \ell_h}{\partial \tau \partial \boldsymbol{\varphi}'} \right) \\ -E \left(\frac{\partial^2 \ell_h}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\beta}'} \right) & -E \left(\frac{\partial^2 \ell_h}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\gamma}'} \right) & -E \left(\frac{\partial^2 \ell_h}{\partial \boldsymbol{\varphi} \partial \tau'} \right) & -E \left(\frac{\partial^2 \ell_h}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'} \right) \end{bmatrix}$$

$$\begin{aligned}
& \left[\sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} X_i \right. \\
& \quad 0 \quad -\sum_{i=1}^N t_i^T \mathbf{w}_i \mathbf{w}_i' \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \quad \sum_{i=1}^N t_i^T \log(t_i) \mathbf{w}_i \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \quad \sum_{i=1}^N t_i^T \mathbf{w}_i \mathbf{u}_i \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \quad 0 \\
& \quad 0 \quad \sum_{i=1}^N t_i \log(t_i) \mathbf{w}_{ii} \exp(\mathbf{w}_{ii}' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \quad \sum_{i=1}^N \frac{\delta_{ii}}{\tau^2} + \sum_{i=1}^N t_i^T \{\log(t_i)\}^2 \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \quad \sum_{i=1}^N t_i^T u_i \log(t_i) \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \quad \sum_{i=1}^N t_i^T u_i \log(t_i) \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \\
& \quad 0 \quad \sum_{i=1}^N u_i \mathbf{w}_i t_i^T \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \quad \sum_{i=1}^N u_i t_i^T \log(t_i) \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \quad \sum_{i=1}^N \mathbf{u}_i^2 t_i^T \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \quad \sum_{i=1}^N \mathbf{u}_i^2 t_i^T \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \boldsymbol{\varphi}' \mathbf{u}_i) \left. \right] \\
& =
\end{aligned}$$

Then iterative estimating equation of adjusted profile (A.6) becomes:

$$S(\boldsymbol{\theta}) = \left[\frac{\partial \ell_h}{\partial \boldsymbol{\theta}} \right]_{q \times 1} - 0.5 \frac{1}{H_i} \left[\frac{\partial H_i}{\partial \boldsymbol{\theta}} \right]_{q \times 1}, \quad (\text{A.13})$$

$$S(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial \ell_h}{\partial \boldsymbol{\beta}} \\ \frac{\partial \ell_h}{\partial \gamma} \\ \frac{\partial \ell_h}{\partial \tau} \\ \frac{\partial \ell_h}{\partial \boldsymbol{\varphi}} \end{bmatrix} - 0.5 \frac{1}{H_i} \begin{bmatrix} \frac{\partial H_i}{\partial \boldsymbol{\beta}} \\ \frac{\partial H_i}{\partial \gamma} \\ \frac{\partial H_i}{\partial \tau} \\ \frac{\partial H_i}{\partial \boldsymbol{\varphi}} \end{bmatrix}.$$

Then, use Newton-Raphson equation to solve MHL, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\gamma}', \hat{\tau}, \hat{\boldsymbol{\varphi}})'$, estimation of equation (A.12) as follows

$$\hat{\boldsymbol{\theta}}^{(m+1)} = \hat{\boldsymbol{\theta}}^{(m)} + \left(-\frac{\partial^2 \ell_h}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)^{-1} \frac{\partial p_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}. \quad (\text{A.14})$$

A.1.2 Estimation for $\boldsymbol{\xi}$

We now develop an estimation procedure via the h-likelihood (A.4). Given the fixed effect parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}', \gamma', \tau, \boldsymbol{\varphi}')$, the maximum h-likelihood estimators (MHLEs) for dispersion parameters $\boldsymbol{\xi} = (\sigma_{\varepsilon}^2, \mathbf{D})$ are obtained by maximizing the following adjusted profile h-likelihood

$$\begin{aligned} p_{\boldsymbol{\xi}} &= \left[\ell_h - 0.5 \log \left| \frac{H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})]}{2\pi} \right| \right] \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \mathbf{u}=\hat{\mathbf{u}}} \\ &= \left[\ell_h - 0.5 \log \left(\frac{1}{2\pi} \left| H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})] \right| \right) \right] \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \mathbf{u}=\hat{\mathbf{u}}} \\ &= \left[\ell_h - 0.5 \left(\log \left(\frac{1}{2\pi} \right) + \log \left| H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})] \right| \right) \right] \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \mathbf{u}=\hat{\mathbf{u}}}, \end{aligned} \quad (\text{A.15})$$

where $H[\ell_h, (\boldsymbol{\theta}, \mathbf{u}')] = - \begin{bmatrix} \frac{\partial^2 \ell_h}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} & \frac{\partial^2 \ell_h}{\partial \boldsymbol{\theta} \partial \mathbf{u}'} \\ \frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \boldsymbol{\theta}'} & \frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \mathbf{u}'} \end{bmatrix}$

The maximum adjusted profile h-likelihood, p_ξ , for estimators for are obtained by solving iteratively

$$\begin{aligned} S(\boldsymbol{\xi}) = \frac{\partial p_\xi}{\partial \boldsymbol{\xi}} &= \frac{\partial \ell_h}{\partial \boldsymbol{\xi}} - 0.5 \frac{\partial \log |H^*(\boldsymbol{\theta})|}{\partial \boldsymbol{\theta}} \\ &= \frac{\partial \ell_h}{\partial \boldsymbol{\xi}} - 0.5 \frac{1}{|H^*(\boldsymbol{\theta}, \mathbf{u})|} \frac{\partial |H^*(\boldsymbol{\theta}, \mathbf{u})|}{\partial \boldsymbol{\theta}} = 0, \end{aligned} \quad (\text{A.16})$$

where $|H^*(\boldsymbol{\theta}, \mathbf{u})| = \left| H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})] \right| = \left| - \begin{bmatrix} \frac{\partial^2 \ell_h}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} & \frac{\partial^2 \ell_h}{\partial \boldsymbol{\theta} \partial \mathbf{u}'} \\ \frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \boldsymbol{\theta}'} & \frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \mathbf{u}'} \end{bmatrix} \right|$

$$= \left| \begin{bmatrix} -\frac{\partial^2 \ell_h}{\partial \beta \partial \beta'} & -\frac{\partial^2 \ell_h}{\partial \beta \partial \gamma'} & -\frac{\partial^2 \ell_h}{\partial \beta \partial \tau'} & -\frac{\partial^2 \ell_h}{\partial \beta \partial \varphi'} & -\frac{\partial^2 \ell_h}{\partial \beta \partial \mathbf{u}'} \\ -\frac{\partial^2 \ell_h}{\partial \gamma \partial \beta'} & -\frac{\partial^2 \ell_h}{\partial \gamma \partial \gamma'} & -\frac{\partial^2 \ell_h}{\partial \gamma \partial \tau'} & -\frac{\partial^2 \ell_h}{\partial \gamma \partial \varphi'} & -\frac{\partial^2 \ell_h}{\partial \gamma \partial \mathbf{u}'} \\ -\frac{\partial^2 \ell_h}{\partial \tau \partial \beta'} & -\frac{\partial^2 \ell_h}{\partial \tau \partial \gamma'} & -\frac{\partial^2 \ell_h}{\partial \tau \partial \tau'} & -\frac{\partial^2 \ell_h}{\partial \tau \partial \varphi'} & -\frac{\partial^2 \ell_h}{\partial \tau \partial \mathbf{u}'} \\ -\frac{\partial^2 \ell_h}{\partial \varphi \partial \beta'} & -\frac{\partial^2 \ell_h}{\partial \varphi \partial \gamma'} & -\frac{\partial^2 \ell_h}{\partial \varphi \partial \tau'} & -\frac{\partial^2 \ell_h}{\partial \varphi \partial \varphi'} & -\frac{\partial^2 \ell_h}{\partial \varphi \partial \mathbf{u}'} \\ -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \beta'} & -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \gamma'} & -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \tau'} & -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \varphi'} & -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \mathbf{u}'} \end{bmatrix} \right|.$$

Then, use Newton-Raphson equation to solve MHL estimation of equation (??) as follows

$$\hat{\boldsymbol{\xi}}^{(m+1)} = \hat{\boldsymbol{\xi}}^{(m)} + \left(-\frac{\partial^2 p_\xi}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}'} \right)^{-1} \frac{\partial p_\xi}{\partial \boldsymbol{\xi}}, \quad (\text{A.17})$$

Replaced by

$$\hat{\boldsymbol{\xi}}^{(m+1)} = \hat{\boldsymbol{\xi}}^{(m)} + \left(-\frac{\partial^2 \ell_h}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}'} \right)^{-1} \frac{\partial p_\xi}{\partial \boldsymbol{\xi}}. \quad (\text{A.18})$$

where (In case $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ and $u_i \sim N(0, \sigma_u^2)$)

$$\begin{aligned} \frac{\partial \ell_h}{\partial \boldsymbol{\xi}} &= \begin{bmatrix} \frac{\partial \ell_h}{\partial \sigma_\varepsilon^2} \\ \frac{\partial \ell_h}{\partial \sigma_u^2} \end{bmatrix} = \begin{bmatrix} -\frac{N}{2\sigma_\varepsilon^2} + \frac{1}{2(\sigma_\varepsilon^2)^2} \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i u_i) (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i u_i)' \\ -\frac{1}{2\sigma_u^2} + \frac{1}{2(\sigma_u^2)^2} \sum_{i=1}^N u_i^2 \end{bmatrix} \\ &= \begin{bmatrix} -N \sigma_\varepsilon^2 + \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i u_i) (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i u_i)' \\ -\sigma_u^2 + \sum_{i=1}^N u_i^2 \end{bmatrix}, \\ -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}'} &= \begin{bmatrix} -\frac{\partial^2 \ell_h}{\partial \sigma_\varepsilon^2 \partial \sigma_\varepsilon^2} \\ -\frac{\partial^2 \ell_h}{\partial \sigma_u^2 \partial \sigma_u^2} \end{bmatrix} = \begin{bmatrix} N \\ 1 \end{bmatrix}, \end{aligned}$$

and $|H^*(\boldsymbol{\theta}, \mathbf{u})| = \left| H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})] \right| = |W|$ is given by

$$\begin{aligned} &\left| \begin{bmatrix} \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i & 0 & 0 \\ 0 & + \sum_{i=1}^N t_i^\tau \mathbf{w}_i \mathbf{w}_i' \exp(\vartheta) & \sum_{i=1}^N t_i^\tau \log(t_i) \mathbf{w}_i \exp(\vartheta) \\ 0 & \sum_{i=1}^N t_i^\tau \log(t_i) \mathbf{w}_i \exp(\vartheta) & \sum_{i=1}^N \frac{\delta_i}{\tau^2} + \sum_{i=1}^N t_i^\tau \{\log(t_i)\}^2 \exp(\vartheta) \\ 0 & \sum_{i=1}^N \mathbf{u}_i \mathbf{w}_i t_i^\tau \exp(\vartheta) & \sum_{i=1}^N \mathbf{u}_i t_i^\tau \log(t_i) \exp(\vartheta) \\ \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^N \mathbf{X}_i' \mathbf{Z}_i & \varphi \sum_{i=1}^N t_i^\tau \mathbf{w}_i \exp(\vartheta) & \varphi \sum_{i=1}^N t_i^\tau \log(t_i) \exp(\vartheta) \end{bmatrix} \right| \\ &\left| \begin{bmatrix} 0 & \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^N \mathbf{X}_i' \mathbf{Z}_i \\ \sum_{i=1}^N t_i^\tau \mathbf{w}_i \mathbf{u}_i \exp(\vartheta) & \varphi \sum_{i=1}^N t_i^\tau \mathbf{w}_i \exp(\vartheta) \\ \sum_{i=1}^N u_i t_i^\tau \log(t_i) \exp(\vartheta) & \varphi \sum_{i=1}^N t_i^\tau \log(t_i) \exp(\vartheta) \\ \sum_{i=1}^N \mathbf{u}_i^2 t_i^\tau \exp(\vartheta) & -\sum_{i=1}^N \delta_i + \sum_{i=1}^N t_i^\tau \exp(\vartheta) + \varphi \sum_{i=1}^N t_i^\tau \mathbf{u}_i \exp(\vartheta) \\ -\sum_{i=1}^N \delta_i + \sum_{i=1}^N t_i^\tau \exp(\vartheta) + \varphi \sum_{i=1}^N t_i^\tau \mathbf{u}_i \exp(\vartheta) & \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^N \mathbf{Z}_i' \mathbf{Z}_i + \sum_i \frac{1}{\sigma_u^2} + \varphi \varphi' \sum_{i=1}^N t_i \exp(\vartheta) \end{bmatrix} \right|, \end{aligned}$$

where $\exp(\vartheta) = \exp(\mathbf{w}'_i \boldsymbol{\gamma} + \varphi u_i)$

$$\begin{aligned} \text{Let find } H^*(\boldsymbol{\theta}, \mathbf{u}) &= H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})] = - \begin{bmatrix} \frac{\partial^2 \ell_h}{\boldsymbol{\theta} \partial \boldsymbol{\theta}'} & \frac{\partial^2 \ell_h}{\partial \boldsymbol{\theta} \partial \mathbf{u}'} \\ \frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \boldsymbol{\theta}'} & \frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \mathbf{u}'} \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} & -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}'} & -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\beta} \partial \boldsymbol{\tau}'} & -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\beta} \partial \boldsymbol{\varphi}'} \\ -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}'} & -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}'} & -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\varphi}'} & -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\varphi}'} \\ -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\tau} \partial \boldsymbol{\beta}'} & -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\tau} \partial \boldsymbol{\gamma}'} & -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'} & -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\tau} \partial \boldsymbol{\varphi}'} \\ -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\beta}'} & -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\gamma}'} & -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\tau}'} & -\frac{\partial^2 \ell_h}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'} \end{bmatrix} & \begin{bmatrix} -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \boldsymbol{\beta}'} \\ -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \boldsymbol{\gamma}'} \\ -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \boldsymbol{\tau}'} \\ -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \boldsymbol{\varphi}'} \end{bmatrix} \\ \begin{bmatrix} -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \boldsymbol{\beta}'} & -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \boldsymbol{\gamma}'} & -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \boldsymbol{\tau}'} & -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \boldsymbol{\varphi}'} \end{bmatrix} & \begin{bmatrix} -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \mathbf{u}'} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} A & B \\ B^t & C \end{bmatrix}, \end{aligned}$$

where

$$\begin{aligned} B &= \begin{bmatrix} \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^N \mathbf{X}'_i \mathbf{Z}_i \\ \varphi \sum_{i=1}^N t_i^\tau \mathbf{w}_i \exp(\mathbf{w}'_i \boldsymbol{\gamma} + \varphi u_i) \\ \varphi \sum_{i=1}^N t_i^\tau \log(t_i) \exp(\mathbf{w}'_i \boldsymbol{\gamma} + \varphi u_i) \\ -\sum_{i=1}^N \delta_i + \sum_{i=1}^N t_i^\tau \mathbf{u}_i \exp(\mathbf{w}'_i \boldsymbol{\gamma} + \varphi u_i) \end{bmatrix}, \\ C &= \left[\frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^N \mathbf{Z}'_i \mathbf{Z}_i + \sum_{i=1}^N \frac{1}{\sigma_u^2} + \varphi^2 \sum_{i=1}^N t_i^\tau \exp(\mathbf{w}'_i \boldsymbol{\gamma} + \varphi u_i) \right], \end{aligned}$$

$$A = \begin{bmatrix} \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i & 0 & 0 & 0 \\ 0 & \sum_{i=1}^N t_i^\tau \exp(\vartheta) & \sum_{i=1}^N t_i^\tau \log(t_i) \mathbf{w}_i \exp(\vartheta) & \sum_{i=1}^N t_i^\tau \mathbf{w}_i u_i \exp(\vartheta) \\ 0 & \sum_{i=1}^N t_i^\tau \log(t_i) \mathbf{w}_i \exp(\vartheta) & \sum_{i=1}^N \frac{\delta_i}{\tau^2} + \sum_{i=1}^N t_i^\tau \{\log(t_i)\}^2 \exp(\vartheta) & \sum_{i=1}^N u_i t_i^\tau \log(t_i) \exp(\vartheta) \\ 0 & \sum_{i=1}^N u_i \mathbf{w}_i t_i^\tau \exp(\vartheta) & \sum_{i=1}^N u_i t_i^\tau \log(t_i) \exp(\vartheta) & \sum_{i=1}^N \mathbf{u}_i u_i t_i^\tau \exp(\vartheta) \\ \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^N \mathbf{X}_i' \mathbf{Z}_i & \varphi \sum_{i=1}^N t_i^\tau w_i \exp(\vartheta) & \psi \sum_{i=1}^N t_i^\tau \log(t_i) \exp(\vartheta) & - \sum_{i=1}^N \delta_i + \varphi \sum_{i=1}^N t_i^\tau \exp(\vartheta) + \sum_{i=1}^N t_i^\tau \exp(\vartheta) \end{bmatrix},$$

where $\exp(\vartheta) = \exp(\mathbf{w}_i' \boldsymbol{\gamma} + \varphi u_i)$

Let find $\frac{\partial H^*(\boldsymbol{\theta}, \mathbf{u})}{\partial \boldsymbol{\xi}}$

$$\begin{aligned} \frac{\partial W}{\partial \boldsymbol{\xi}} &= \frac{\partial}{\partial \boldsymbol{\xi}} \left(H[\ell_h, (\boldsymbol{\theta}, \mathbf{u})] \right) = \frac{\partial}{\partial \boldsymbol{\xi}} \left(- \begin{bmatrix} \frac{\partial^2 \ell_h}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} & \frac{\partial^2 \ell_h}{\partial \boldsymbol{\theta} \partial \mathbf{u}'} \\ \frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \boldsymbol{\theta}'} & \frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \mathbf{u}'} \end{bmatrix} \right) \\ &= \frac{\partial}{\partial \boldsymbol{\xi}} \left(- \begin{bmatrix} \left[\begin{array}{cccc} -\frac{\partial^2 \ell_h}{\partial \beta \partial \beta'} & -\frac{\partial^2 \ell_h}{\partial \beta \partial \gamma'} & -\frac{\partial^2 \ell_h}{\partial \beta \partial \tau'} & -\frac{\partial^2 \ell_h}{\partial \beta \partial \varphi'} \\ -\frac{\partial^2 \ell_h}{\partial \gamma \partial \beta'} & -\frac{\partial^2 \ell_h}{\partial \gamma \partial \tau'} & -\frac{\partial^2 \ell_h}{\partial \gamma \partial \tau^t} & -\frac{\partial^2 \ell_h}{\partial \gamma \partial \varphi'} \\ -\frac{\partial^2 \ell_h}{\partial \tau \partial \beta'} & -\frac{\partial^2 \ell_h}{\partial \tau \partial \gamma'} & -\frac{\partial^2 \ell_h}{\partial \tau \partial \tau'} & -\frac{\partial^2 \ell_h}{\partial \tau \partial \varphi'} \\ -\frac{\partial^2 \ell_h}{\partial \varphi \partial \beta'} & -\frac{\partial^2 \ell_h}{\partial \varphi \partial \gamma'} & -\frac{\partial^2 \ell_h}{\partial \varphi \partial \tau^t} & -\frac{\partial^2 \ell_h}{\partial \varphi \partial \varphi'} \end{array} \right] \left[\begin{array}{c} -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \beta'} \\ -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \gamma'} \\ -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \tau'} \\ -\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \varphi'} \end{array} \right] \end{bmatrix} \right) = \begin{bmatrix} A & B \\ B^t & C \end{bmatrix}. \end{aligned}$$

That is,

$$\begin{aligned} \frac{\partial W}{\partial \sigma_\epsilon^2} &= - \begin{bmatrix} \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \beta \partial \beta'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \beta \partial \gamma'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \beta \partial \tau'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \beta \partial \varphi'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \beta'} \right) \\ \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \gamma \partial \beta'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \gamma \partial \tau'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \gamma \partial \tau^t} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \gamma \partial \varphi'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \gamma'} \right) \\ \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \tau \partial \beta'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \tau \partial \gamma'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \tau \partial \tau'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \tau \partial \varphi'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \tau'} \right) \\ \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \varphi \partial \beta'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \varphi \partial \gamma'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \varphi \partial \tau^t} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \varphi \partial \varphi'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \varphi'} \right) \\ \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \beta'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \gamma'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \tau'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \varphi'} \right) & \frac{\partial}{\partial \sigma_\epsilon^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \mathbf{u}'} \right) \end{bmatrix} \\ &= \begin{bmatrix} D & E \\ E^t & F \end{bmatrix} \end{aligned}$$

(1.19)

$$\begin{aligned}
&= \begin{bmatrix} \begin{bmatrix} \frac{\partial}{\partial \sigma_\varepsilon^2} \left(\frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} \frac{\partial}{\partial \sigma_\varepsilon^2} \left(\frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^N \mathbf{X}_i \mathbf{Z}_i \right) \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} \frac{\partial}{\partial \sigma_\varepsilon^2} \left(\frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^N \mathbf{X}_i \mathbf{Z}_i \right) & 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} -\frac{1}{(\sigma_\varepsilon^2)^2} \sum_{i=1}^N \mathbf{Z}_i' \mathbf{Z}_i \end{bmatrix} \end{bmatrix} \\
&= \begin{bmatrix} \begin{bmatrix} -\frac{1}{(\sigma_\varepsilon^2)^2} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} -\frac{1}{(\sigma_\varepsilon^2)^2} \sum_{i=1}^N \mathbf{X}_i \mathbf{Z}_i \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} -\frac{1}{(\sigma_\varepsilon^2)^2} \sum_{i=1}^N \mathbf{X}_i \mathbf{Z}_i & 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} -\frac{1}{(\sigma_\varepsilon^2)^2} \sum_{i=1}^N \mathbf{Z}_i' \mathbf{Z}_i \end{bmatrix} \end{bmatrix}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{\partial W}{\partial \sigma_u^2} &= - \begin{bmatrix} \begin{bmatrix} \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \beta \partial \beta'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \beta \partial \gamma'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \beta \partial \tau'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \beta \partial \varphi'} \right) \\ \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \gamma \partial \beta'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \gamma \partial \gamma'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \gamma \partial \tau'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \gamma \partial \varphi'} \right) \\ \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \tau \partial \beta'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \tau \partial \gamma'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \tau \partial \tau'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \tau \partial \varphi'} \right) \\ \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \varphi \partial \beta'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \varphi \partial \gamma'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \varphi \partial \tau'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \varphi \partial \varphi'} \right) \\ \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \beta'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \gamma'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \tau'} \right) & \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \varphi'} \right) \end{bmatrix} & \begin{bmatrix} \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \beta'} \right) \\ \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \gamma'} \right) \\ \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \tau'} \right) \\ \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \varphi'} \right) \\ \frac{\partial}{\partial \sigma_u^2} \left(\frac{\partial^2 \ell_h}{\partial \mathbf{u} \partial \mathbf{u}'} \right) \end{bmatrix} \end{bmatrix} \\
&= \begin{bmatrix} G & H \\ H^t & I \end{bmatrix}
\end{aligned}$$

(1.19)

$$= \begin{bmatrix} \begin{bmatrix} \frac{\partial}{\partial \sigma_u^2} (\frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} \frac{\partial}{\partial \sigma_u^2} (\frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^N \mathbf{X}_i \mathbf{Z}_i) \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} \frac{\partial}{\partial \sigma_u^2} (\frac{1}{\sigma^2} \sum_{i=1}^N \mathbf{X}_i \mathbf{Z}_i) & 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} -\frac{1}{(\sigma_u^2)^2} \sum_{i=1}^N \mathbf{Z}_i^t \mathbf{Z}_i \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} -\frac{N}{(\sigma_u^2)^2} \end{bmatrix} \end{bmatrix}.$$

Then

$$\frac{\partial W}{\partial \boldsymbol{\xi}} = \frac{\partial W}{\partial \sigma^2} + \frac{\partial W}{\partial \sigma_u^2}.$$

As result, score function of adjusted profile h-likelihood (A.5) becomes:

$$\begin{aligned} S(\boldsymbol{\xi}) &= \frac{\partial \ell_h}{\partial \boldsymbol{\xi}} - 0.5 \frac{1}{|H^*(\boldsymbol{\theta}, \mathbf{u})|} \frac{\partial |H^*(\boldsymbol{\theta}, \mathbf{u})|}{\partial \boldsymbol{\xi}} \\ &= \begin{bmatrix} \frac{\partial \ell_h}{\partial \sigma^2} \\ \frac{\partial \ell_h}{\partial \sigma_u^2} \end{bmatrix} - 0.5 \frac{1}{|W|} \begin{bmatrix} \frac{\partial |W|}{\partial \sigma^2} \\ \frac{\partial |W|}{\partial \sigma_u^2} \end{bmatrix}. \end{aligned} \tag{1.20}$$

Appendix B

R Codes of Simulation Study for Joint Model

Joint Model Under ML Method

```
#===== Generated survival times with Weibull Distribution =====  
surv.dat.frailty <- function(k=300, gamma=c(-6,1.5, 0.50), lambdaC=-6.5,  
Tau=0.8, sigmasq.u=0.5, sigmasq.e =0.5, phi=0.5)  
{  
  lambdaT <- gamma[1]  
  id <- c(1:k)  
  x1 <- ifelse(id==1:(k/2), 0, 1)  
  x2 <- round(rnorm(k, mean= 2, sd=0.5), digits=4)  
  vi <- round(rnorm(k, mean=0, sd=sqrt(sigmasq.u)), digits=4)  
  ei <- round(rnorm(k, mean =0 , sd= sqrt(sigmasq.e)), digits=4)  
  #===== true survival time =====  
  lambda0 <- exp(lambdaT + gamma[2]*x1 + gamma[3]*x2 + phi*vi)  
  Ti <- rweibull(k, shape=Tau, scale=1/lambda0^(1/Tau))  
  lambda1 <- exp(lambdaC + gamma[2]*x1 + gamma[3]*x2 + phi*vi)  
  C <- rweibull(k, shape=Tau, scale=1/lambda1^(1/Tau))
```



```

Ci <- pmin(C, 180) # censoring time
#===== observed survival time =====
si <- pmin(Ti, Ci)
event <- ifelse(si==Ti, 1, 0)
fol.si <- as.integer(si/15)
round(data.frame(id, si, fol.si, event, x1, x2, vi, Ti, Ci), digits=4)
}

surv.dat0 <- surv.dat.frailty(k=300, gamma = c(-6, 1.5, 0.5), lambdaC=-6.5,
Tau=0.8, sigmasq.u=0.5, sigmasq.e=0.5, phi=0.5)
#===== Generated longitudinal data using LME model =====
long.data <- function(dat=surv.dat0, beta=c(10,-0.2, 0.6), sigmasq.e=0.5)
{
  fol.Ti <- dat$fol.si
  ni<- length(fol.Ti)
  id <- NULL
  tj <- NULL
  x1 <- NULL
  x2 <- NULL
  ui <- NULL
  eij <- NULL
  Time <- NULL
  event <- NULL
  for(j in 1:ni)
  {
    tij <- c(0:fol.Ti[j])
    tj <- c(tj, tij)
    l.j <- length(tij)
    id <- c(id, c(rep(dat$id[j], l.j )))
    x1 <- c(x1, c(rep(dat$x1[j], l.j )))

```

```

x2 <- c(x2, c(rep(dat$x2[j], 1.j )))
Time <- c(Time, c(rep(dat$si[j], 1.j )))
event <- c(event, c(rep(dat$event[j], 1.j )))
ui <- c(ui, c(rep(dat$vi[j], 1.j )))
ej <- round(rnorm(1.j, mean = 0 , sd = sqrt(sigmasq.e)), digits=4)
eij<- c(eij, ej)
yij <- beta[1] + beta[2]*tj + beta[3]*tj*x2 + ui + eij
}

data.frame(id=id, tj=tj, Time=Time, event=event, yij=yij, x1=x1, x2=x2,
  ui=ui, eij=eij)
}

long.dat0 <- long.data(dat=surv.dat0, beta=c(10,-0.2, 0.6),sigmasq.e=0.5)
#===== Simpson's Rule for integration =====
simpson <- function(vec, a, b) ### see simpson's rule pdf
{
  # used for numerical integration
  # uses Simpson's rule; function values are elements of vec
  # length of vec should be odd
  n <- length(vec)
  m <- n - 1
  h <- (b - a)/m
  d <- c(1, rep(c(4, 2), m/2 - 1), 4, 1)
  integral <- (h/3) * sum(d * vec)
  list(integral = integral)
}

#=== ML estimators of JointModel By using "optim" function===
JointModel.like <- function(dat1 = surv.dat0, dat2 = long.dat0,
  beta=c(10.00, -0.2, 0.6), gamma= c(-6.0, 1.5, 0.5), Tau= 0.8, phi=0.5,
  sigmasq.u=0.5, sigmasq.e=0.5)

```

```

# it uses the function "optim" to minimize log-likelihood
{
id <- dat2$id
y <- dat2$yij
x <- dat2$x2
tj <- dat2$tj
unique.id <- unique(dat2$id)
k <- length(unique.id)
a1 <- -5      # lower limit of u
a2 <-  5      # upper limit of u
u <- seq(a1, a2, length=151)
initial = c(beta, gamma, Tau, phi, sigmasq.u, sigmasq.e)
obj.fn <- function(theta)
{
beta <- theta[1:3]
gamma <- theta[4:6]
Tau <- theta[7]
phi <- theta[8]
sigmasq.u <- theta[9]
sigmasq.e <- theta[10]
fu <- dnorm(u, 0, sqrt(sigmasq.u))
log.like <- 0
for (i in 1:k)
{
#### f(y|u)
yi <- dat2$yij[dat2$id==unique.id[i]]
xi <- dat2$x2[dat2$id==unique.id[i]]
tj <- dat2$tj[dat2$id==unique.id[i]]
n <- length(yi)

```

```

f.yij.u <- 1
for(j in 1:n)
{
  yij <- yi[j]
  xij <- xi[j]
  tij <- tj[j]
  tij.xij <- tij * xij
  X <- c(1,tij, tij.xij)
  muij <- c(t(X) %*% beta) + u
  f.yij.u0 <- dnorm(yij, mean= muij, sd=sqrt(sigmasq.e))
  #print(f.yij.u0)
  f.yij.u <- f.yij.u * f.yij.u0
}

f.yi.u <- f.yij.u
#### f(T|u)

x1 <- dat1$x1[dat1$id==unique.id[i]]
x2 <- dat1$x2[dat1$id==unique.id[i]]
x <- c(1, x1, x2)
Time <- dat1$si[dat1$id==unique.id[i]]
delta <- dat1$event[dat1$id==unique.id[i]]
const <- c((Tau * (Time^(Tau-1)))^(delta))
eta.i <- c(t(x) %*% gamma) + phi * u
sum.i <- delta * eta.i - (Time^(Tau)) * exp(eta.i)
f.t.u <- const * exp(sum.i)
##### likelihood function f.t.u * f.y.u * fu
like.i <- simpson(f.t.u * f.yi.u * fu, a1, a2)$integral
log.like.i <- log(like.i)
log.like <- log.like + log.like.i
}

```

```

-log.like
}

like.fit <- optim(par = initial, fn=obj.fn, hessian=TRUE,
method=c("L-BFGS-B"), control = list(maxit=500),
              lower=c(9.5 , -0.3, 0.4, -7.5, 1, 0, 0.6, 0.10, 0, 0),
              upper=c(10.5, -0.1, 0.8, -4.5, 2, 1, 1.0, 0.90, Inf, Inf))

cat(like.fit$message, "\n")

estimate <- round(like.fit$par, digits=4)
std.err <- round(sqrt(diag(solve(like.fit$hessian))), digits=4)
objective <- like.fit$value

list(estimate=estimate, std.err=std.err, objective=objective)
}

fit3 <- JointModel.like(dat1 = surv.dat0, dat2 = long.dat0,
  beta=c(10.00, -0.2,0.6), gamma= c(-6.0, 1.5, 0.5), Tau= 0.8, phi=0.5,
  sigmasq.u=0.5, sigmasq.e=0.5)

#===== simulation study =====

JM.sim.try <- function(sim=2, k=300, beta=c(10.00, -0.2, 0.6),
gamma=c(-6.0, 1.5, 0.5), lambdaC=-6.5, Tau=0.8, sigmasq.e=0.5,
sigmasq.u=0.5, phi=0.5)
{
doit <- function(s=1)
{
cat("Sim.No:", s, "\n")

data0 <- surv.dat.frailty(k=k, gamma=gamma, lambdaC=lambdaC, Tau=Tau,
sigmasq.u =sigmasq.u, sigmasq.e=sigmasq.e, phi=phi)

Data0 <- long.data(dat=data0, beta=beta, sigmasq.e=sigmasq.e)

like.fit <- JointModel.like(dat1=data0, dat2=Data0, beta=beta,
gamma=gamma, Tau=Tau, phi=phi, sigmasq.u=sigmasq.u,
sigmasq.e=sigmasq.e)

```

```

par.estimate <- like.fit$estimate
par.std <- like.fit$std.err
list(par.estimate=par.estimate, par.std=par.std)
}

res <- lapply(1:sim, function(i) try(doit(s=i), TRUE))
res.out <- res[sapply(res, function(x) !inherits(x, "try-error"))]
length.res.out <- length(res.out)
cat("No. samples:", length.res.out, "\n")
Table.est <- NULL
for( i in 1: length.res.out)
{
estimate <- res.out[[i]]$par.estimate
Std <- res.out[[i]]$par.std
Table.est0 <- c(estimate, Std)
Table.est <- rbind(Table.est, Table.est0)
}

dimnames(Table.est) <- list(NULL, c("beta0", "beta1", "beta2", "gamma0",
"gamma1", "gamma2", "Tau", "phi", "sigmasq.u", "sigmasq.e", "std.beta0",
"std.beta1", "std.beta2", "std.gamma0", "std.gamma1", "std.gamma2",
"std.Tau", "std.phi", "std.sigmasq.u", "std.sigmasq.e"))
Table.est
}

simul <- JM.sim.try(sim=500, k=300, beta=c(10.00, -0.2, 0.6),
gamma=c(-6.0, 1.5, 0.5), lambdaC=-6.2, Tau=0.8, sigmasq.e=0.5,
sigmasq.u=0.5, phi=0.5)

```

Joint Model Under H-likelihood Method

```

# ==== Estimates of Joint Model based on H-likelihood method ====
# ===== Generating survival data using exponential model =====

```

```

surv.dat.frailty <- function(k=300, lambdaT=-6.0, lambdaC=-6.5,
  sigmasq.u=0.5, sigmasq.e=0.5, phi0=0.5)
{
  id <- c(1:k)
  x1 <- ifelse(id==1:(k/2), 0, 1)
  x2 <- round(rnorm(k, mean=2, sd=0.5), digits=4)
  vi <- round(rnorm(k, mean=0, sd=sqrt(sigmasq.u)), digits=4)
  ei <- round(rnorm(k, mean=0, sd=sqrt(sigmasq.e)), digits=4)
  # ===== true survival time (exponential) =====
  lambda0 <- exp(lambdaT + phi0*vi)
  Ti <- rweibull(k, shape=1, scale=1/lambda0)
  lambda1 <- exp(lambdaC + phi0*vi)
  C <- rweibull(k, shape=1, scale=1/lambda1)
  Ci <- pmin(C, 180) # censoring time
  # ===== observed survival time =====
  si <- pmin(Ti, Ci)
  event <- ifelse(si==Ti, 1, 0)
  fol.si <- as.integer(si/15)
  round(data.frame(id, si, fol.si, event, x1, x2, vi, Ti, Ci), digits=4)
}

surv.dat0 <- surv.dat.frailty(k=100, lambdaT=-4.5, lambdaC=-5.0,
  sigmasq.u=0.5, sigmasq.e=0.5, phi0=0.5)
# ===== Generating longitudinal data using LME model =====
long.data <- function(dat=surv.dat0, beta0=-0.2, sigmasq.e=0.5)
{
  fol.Ti <- dat$fol.si
  ni <- length(fol.Ti)
  id <- NULL
  tj <- NULL

```

```

x2 <- NULL
ui <- NULL
eij <- NULL
Time <- NULL
event <- NULL
for(j in 1:ni)
{
  tij <- c(0:fo1.Ti[j])
  tj <- c(tj, tij)
  l.j <- length(tij)
  id <- c(id, c(rep(dat$id[j], l.j)))
  x2 <- c(x2, c(rep(dat$x2[j], l.j)))
  Time <- c(Time, c(rep(dat$si[j], l.j)))
  event <- c(event, c(rep(dat$event[j], l.j)))
  ui <- c(ui, c(rep(dat$vi[j], l.j)))
  ej <- round(rnorm(l.j, mean =0, sd=sqrt(sigmasq.e)), digits=4)
  eij <- c(eij, ej)
  yij <- beta0*x2 + ui + eij
}

data.frame(id=id,tj=tj,Time=Time,event=event,yij=yij,x2=x2,ui=ui,eij=eij)
}

long.dat0 <- long.data(dat=surv.dat0, beta0=0.5, sigmasq.e=0.5)
# ===== H-likelihood method =====

hlike.fit <- function(dat = long.dat0, beta0 = -0.2, gamma0 = -6.0,
  sigmasq.e = 0.5, sigmasq.u = 0.5, phi0 = 0.5, iter = 20)
{
  # ===== predicting ui =====

  subID <- unique(dat$id)
  k <- length(subID)

```



```

u.hat0 <- rep(0, k)    # initial estimate
for(all.iter in 1:iter)
{
  cat(".")
  u.hat <- NULL
  for(it in 1:5)
  {
    y <- dat$yij
    delta <- dat$event
    time <- dat$Time
    x <- dat$x2
    score.u <- NULL
    fisher.u <- NULL
    for (i in 1:k)
    {
      ui.hat <- u.hat0[i]
      subID <- unique(dat$id)
      yi <- dat$yij[dat$id==subID[i]]
      xi <- dat$x2[dat$id==subID[i]]
      tj <- dat$tj[dat$id==subID[i]]
      timei <- dat$Time[dat$id==subID[i]][1]
      deltai <- dat$event[dat$id==subID[i]][1]
      X <- xi
      ni <- length(yi)
      I <- diag(ni)
      Z <- matrix(c(rep(1, ni)), ni, 1)
      V0 <- sigmasq.e * I
      V0.inv <- solve(V0)
      r <- yi - c(X * beta0) - c(Z * ui.hat)
    }
  }
}

```

```

exp.gam0 <- c(exp(gamma0 + phi0 * ui.hat))
score.u0 <- phi0 * deltai - timei * phi0 * exp.gam0 + c(t(Z)
%% V0.inv %% r) - (ui.hat/sigmasq.u)
fisher.u0 <- timei * (phi0^2) * exp.gam0 + c(t(Z) %% V0.inv %% Z)
+ (1/sigmasq.u)
score.u <- c(score.u, score.u0)
fisher.u <- c(fisher.u, fisher.u0)
}
u.hat0 <- u.hat0 + c(solve(diag(fisher.u)) %% score.u)
u.hat <- cbind(u.hat, u.hat0)
}

# ===== estimating regression parameters =====
theta0 <- c(beta0, gamma0, phi0)
theta <- NULL
for(it in 1:10)
{
subID <- unique(dat$id)
k <- length(subID)
y <- dat$yij
delta <- dat$event
time <- dat$Time
x <- dat$x2
score.theta <- matrix(c(0), 3, 1)
fisher.theta <- matrix(c(0), 3, 3)
second.term <- matrix(c(0), 3, 1)
for (i in 1:k)
{
subID <- unique(dat$id)
yi <- dat$yij[dat$id==subID[i]]

```

```

ui <- u.hat0[i]          # predicted value of ui
xi <- dat$x2[dat$id==subID[i]]
tj <- dat$tj[dat$id==subID[i]]
timei <- dat$Time[dat$id==subID[i]][1]
deltai <- dat$event[dat$id==subID[i]][1]

X <- xi
ni <- length(yi)
I <- diag(ni)
Z <- matrix(c(rep(1, ni)), ni, 1)
V0 <- sigmasq.e * I
V0.inv <- solve(V0)
r <- yi - c(X * beta0) - c(Z * ui)
dlogh.dbeta0 <- t(X) %*% V0.inv %*% r
exp.gam0 <- c(exp(gamma0 + phi0 * ui))
dlogh.dgamma0 <- - c(timei * exp.gam0) + c(deltai)
dlogh.dphi0 <- - c(ui * timei * exp.gam0) + c(deltai * ui)
score.theta0 <- matrix(c(dlogh.dbeta0, dlogh.dgamma0, dlogh.dphi0), 3, 1)
score.theta <- score.theta + score.theta0
fisher.beta0 <- t(X) %*% V0.inv %*% X
fisher.gamma0 <- c(timei * exp.gam0)
fisher.gamma0.phi0 <- c(ui * timei * exp.gam0)
fisher.phi0.gamma0 <- fisher.gamma0.phi0
fisher.phi0 <- c(ui^2 * timei * exp.gam0)
fisher.theta0 <- matrix(c(fisher.beta0 , 0, 0,
                          0, fisher.gamma0, fisher.gamma0.phi0,
                          0, fisher.phi0.gamma0, fisher.phi0), 3, 3)
fisher.theta <- fisher.theta + fisher.theta0
#===== second term in h-likelihood =====
Hi <- (timei*(phi0^2)*exp.gam0 + c(t(Z) %*% V0.inv%*% Z)+(1/sigmasq.u))

```

```

d.Hi.beta0 <- 0
d.Hi.gamma0 <- timei * (phi0^2) * exp.gam0
d.Hi.phi0 <- timei*(phi0^2)*exp.gam0*ui+timei*(2*phi0)*exp.gam0
d.Hi.theta0 <- matrix(c(d.Hi.beta0, d.Hi.gamma0, d.Hi.phi0), 3, 1)
second.term0 <- -0.5 * ((1/Hi) * d.Hi.theta0)
second.term <- second.term + second.term0
}

lh.score <- score.theta + second.term
theta0 <- theta0 + c(solve(fisher.theta) %*% lh.score)
theta <- cbind(theta, theta0)
beta0 <- theta0[1]
gamma0 <- theta0[2]
phi0 <- theta0[3]
}

#=== NEXT: estimating variance parameters (sigmasq.e, sigmasq.u) ===
zeta0 <- c(sigmasq.e, sigmasq.u)
zeta <- NULL
for(it in 1:15)
{
  subID <- unique(dat$id)
  k <- length(subID)
  y <- dat$yij
  delta <- dat$event
  time <- dat$Time
  x <- dat$x2
  score.zeta <- matrix(c(0), 2, 1)
  fisher.zeta <- matrix(c(0), 2, 2)
  Hi.diag.theta <- matrix(c(0), 3, 3)
  Hi.offdiag.beta <- NULL

```

```

Hi.offdiag.gamma <- NULL
Hi.offdiag.phi <- NULL
dHi.diag.theta.e <- matrix(c(0), 3, 3)
dHi.offdiag.beta.e <- NULL
dHi.offdiag.gamma.e <- NULL
dHi.offdiag.phi.e <- NULL
Hi.diag.b <- NULL
dHi.diag.b.e <- NULL
fisher.eu <- matrix(c(0), 2, 2)
for (i in 1:k)
{
  subID <- unique(dat$id)
  yi <- dat$yij[dat$id==subID[i]]
  ui <- u.hat0[i]          # predicted value of ui
  xi <- dat$x2[dat$id==subID[i]]
  tj <- dat$tj[dat$id==subID[i]]
  timei <- dat$Time[dat$id==subID[i]][1]
  deltai <- dat$event[dat$id==subID[i]][1]
  X <- xi
  ni <- length(yi)
  I <- diag(ni)
  Z <- matrix(c(rep(1, ni)), ni, 1)
  V0 <- sigmasq.e * I
  V0.inv <- solve(V0)
  r <- yi - c(X * beta0) - c(Z * ui)
  exp.gam0 <- c(exp(gamma0 + phi0 * ui))
  dlogh.sigmasq.e <- - ni * sigmasq.e + c(t(r) %*% r)
  dlogh.sigmasq.u <- - sigmasq.u + (ui^2)
  fisher.sigmasq.e <- ni

```

```

fisher.sigmasq.u <- 1
score.zeta0 <- c(dlogh.sigmasq.e, dlogh.sigmasq.u)
score.zeta <- score.zeta + score.zeta0
fisher.zeta0 <- matrix(c(fisher.sigmasq.e,0,0, fisher.sigmasq.u),2,2)
fisher.zeta <- fisher.zeta + fisher.zeta0
#===== second term in in h-likelihood =====
Hi.diag.theta0 <- matrix(c((1/sigmasq.e)*c(t(X)%*%X), 0, 0, 0,
  timei*exp.gam0, ui*timei*exp.gam0, 0, ui*timei*exp.gam0,
  (ui^2)*timei*exp.gam0), 3, 3)
Hi.diag.theta <- Hi.diag.theta + Hi.diag.theta0
Hi.offdiag.beta0 <- (1/sigmasq.e)*c(t(Z)%*%X)
Hi.offdiag.gamma0 <- phi0*timei*exp.gam0
Hi.offdiag.phi0 <- -deltai + phi0*ui*timei*exp.gam0 +timei*exp.gam0
Hi.offdiag.beta <- c(Hi.offdiag.beta, Hi.offdiag.beta0)
Hi.offdiag.gamma <- c(Hi.offdiag.gamma, Hi.offdiag.gamma0)
Hi.offdiag.phi <- c(Hi.offdiag.phi, Hi.offdiag.phi0)
Hi.diag.b0 <- (phi0^2)*timei*exp.gam0+(1/sigmasq.e)*c(t(Z)%*%Z)
+(1/sigmasq.e)
Hi.diag.b <- c(Hi.diag.b, Hi.diag.b0)
dHi.diag.theta.e0 <- matrix(c(-(1/sigmasq.e^2)*c(t(X)%*%X),
  0,0,0,0,0, 0,0,0), 3, 3)
dHi.diag.theta.e <- dHi.diag.theta.e + dHi.diag.theta.e0
dHi.offdiag.beta.e0 <- -(1/sigmasq.e^2)*c(t(Z)%*%X)
dHi.offdiag.gamma.e0 <- 0
dHi.offdiag.phi.e0 <- 0
dHi.offdiag.beta.e <- c(dHi.offdiag.beta.e, dHi.offdiag.beta.e0)
dHi.offdiag.gamma.e <- c(dHi.offdiag.gamma.e, dHi.offdiag.gamma.e0)
dHi.offdiag.phi.e <- c(dHi.offdiag.phi.e, dHi.offdiag.phi.e0)
dHi.diag.b.e0 <- -(1/sigmasq.e^2)*c(t(Z)%*%Z)

```

```

dHi.diag.b.e <- c(dHi.diag.b.e, dHi.diag.b.e0)
# ===== standard errors =====
fisher.e <- -0.5 * ni/(sigmasq.e^2) + 1/(sigmasq.e^3) * c(t(r) %*% r)
fisher.u <- -0.5 * 1/(sigmasq.u^2) + 1/(sigmasq.u^3) * (ui^2)
fisher.eu0 <- diag(c(fisher.e, fisher.u))
fisher.eu <- fisher.eu + fisher.eu0
}

Hi.offdiag <- rbind(Hi.offdiag.beta, Hi.offdiag.gamma, Hi.offdiag.phi)
Hi.mat <- rbind(cbind(Hi.diag.theta, Hi.offdiag), cbind(t(Hi.offdiag),
  diag(Hi.diag.b)))
dHi.offdiag.e <- rbind(dHi.offdiag.beta.e, dHi.offdiag.gamma.e,
dHi.offdiag.phi.e)
dHi.sig.e <- rbind(cbind(dHi.diag.theta.e, dHi.offdiag.e),
  cbind(t(dHi.offdiag.e), diag(dHi.diag.b.e)))
# Note: dHi.sig.u is c(0) matrix as " big.mat0" except one term
"-1/(sigmasq.u^2)"
big.mat0 <- matrix(c(0), c(k+3), (k+3))
big.mat0[(3+1):(3+k), (3+1):(3+k)] <- -1/(sigmasq.u^2) * diag(1, k)
dHi.sig.u <- big.mat0
# Note:(add (-) due to 1/|Hi.mat|) # (for "0.5" term, given
  both sigmasq.e,sigmasq.u=0.5, we have 0.5^2+ 0.5^2= 0.5)
second.term.e <- -(sigmasq.e^2) * sum(diag(solve(Hi.mat) %*% dHi.sig.e))
second.term.u <- -(sigmasq.u^2) * sum(diag(solve(Hi.mat) %*% dHi.sig.u))
second.term.zeta <- c(second.term.e, second.term.u)
lh.score.zeta <- score.zeta + second.term.zeta
zeta0 <- zeta0 + c(solve(fisher.zeta) %*% lh.score.zeta)
zeta <- cbind(zeta, zeta0)
sigmasq.e <- zeta0[1]
sigmasq.u <- zeta0[2]

```

```

}

}

std.theta <- sqrt(diag(solve(Hi.mat)))[1:3]
std.zeta <- sqrt(diag(solve(fisher.eu)))

list(lh.score=lh.score, fisher.theta=fisher.theta, theta=theta,
theta0=theta0, std.theta=std.theta, zeta=zeta, zeta0=zeta0, std.zeta=
std.zeta, fisher.eu=fisher.eu, u.hat=u.hat)
}

hl.fit <- hlike.fit(dat = long.dat0, beta0 = 0.5, gamma0 = -4.5,
sigmasq.e = 0.5, sigmasq.u = 0.5, phi0 = 0.5, iter = 20)
hl.fit
#=====simulation study =====
frailty.sim.try <- function(sim=2, k=100, lambdaT=-4.5, lambdaC=-5.0,
beta0=0.5, gamma0=-4.5, sigmasq.u=0.5, sigmasq.e=0.5, phi0=0.5)
{
Table.est <- NULL
for(s in 1:sim)
{
cat("Sim.No:", s, "\n")
data0 <- surv.dat.frailty(k=k, lambdaT=lambdaT, lambdaC=lambdaC,
sigmasq.u=sigmasq.u, sigmasq.e=sigmasq.e, phi0=phi0)
Data0 <- long.data(dat=data0, beta0=beta0, sigmasq.e=sigmasq.e)
hl.fit<- hlike.fit(dat = Data0, beta0 = beta0, gamma0 = gamma0,
sigmasq.e = sigmasq.e, sigmasq.u = sigmasq.u, phi0 = phi0)
par.estimate <- c(hl.fit$theta0, hl.fit$zeta0)
par.std <- c(hl.fit$std.theta, hl.fit$std.zeta)
Table.est0 <- c(par.estimate, par.std)
Table.est <- rbind(Table.est, Table.est0)
}
}

```



```

dimnames(Table.est) <- list( NULL, c("beta0","gamma0", "phi0",
"sigmasq.e", "sigmasq.u","std.beta0", "std.gamma0", "std.phi0",
"std.sigmasq.e", "std.sigmasq.u"))
Table.est
}

simul <- frailty.sim.try(sim=500, k=200, lambdaT=-4.5, lambdaC=-5.0,
beta0=0.5, gamma0=-4.5, sigmasq.u=0.5, sigmasq.e=0.5, phi0=0.5)
Bias.theta <- function(dat=simul, theta=c(0.5, -4.5, 0.5, 0.5, 0.5))
{
#===== EST =====
Est.theta <- round(apply(dat, 2 , mean)[1:5], digits=5)
#===== Bias =====
bias.theta <- round(apply(dat, 2 , mean)[1:5] - theta , digits=5)
Var.theta <- apply(dat, 2 , var)[1:5]
#===== MSE =====
MSE.theta <- round(Var.theta + (bias.theta^2), digits=5)
#===== coverage probabilities of 95% CI =====
n <- nrow(dat)
CP.theta.0.95CI <- abs(dat[, 1:5] - matrix(rep(theta ,c(n,n,n,n,n)),
n, 5)) <= 1.96 * dat[, 6:10]
covpr.theta <- round(apply(CP.theta.0.95CI, 2, mean) * 100, digits=1)
#===== The Length of 95% CI (L=U-L) =====
Length.0.95.CI <- 2 * 1.96 * dat[, 6:10]
#===== The average Length of 95% CI =====
ave.Length.theta <- round(apply(Length.0.95.CI, 2, mean), digits=4)
#===== Empirical Result =====
result1 <- cbind(Est.theta, bias.theta, MSE.theta, covpr.theta ,
ave.Length.theta)
dimnames(result1) <- list(c("beta0", "gamma", "phi", "sigmsq.e",

```

```

"sigmsq.u"), c("Est","Bias", "MSE", "Cov_Prob", "ave.length.CI" ))
}

sim.estimate <- Bias.theta(dat=simul, theta=c(0.5, -4.5, 0.5, 0.5, 0.5))

```

Joint Model of Misspecified random effects Under ML Method

```

#==== In This R Code: we need to set Phi=0.5
#==== we generate data under assumption: [ui ~ chi^2, df=3

#=== we want to estimate survival submodel parameters under assumption :
ui ~ N(0,sigmsqu)
#===== Generated survival times with t \& Weibull Distribution =====
surv.dat.frailty <- function(k=100, gamma=c(-6,1.5, 0.50), lambdaC=-6.5,
Tau=0.8, sigmasq.u2=4, sigmasq.u =0.5, sigmasq.e =0.5, phi=0.5,alpha=0.05)
{
lambdaT <- gamma[1]
id <- c(1:k)
x1 <- ifelse(id==1:(k/2),0,1)
x2 <- round(rnorm(k, mean= 2, sd=0.5), digits=4)
# Note: we can erase lines 5-8, we are not consider robust method
any more, instead chisq, df=3
df <- 3
vi <- sqrt(sigmasq.u) * (1/sqrt(2*df)) * c(rchisq(k, df) - df)
ei <- round(rnorm(k, mean =0 , sd= sqrt(sigmasq.e)), digits=4)
#===== true survival time =====
lambda0 <- exp(lambdaT + gamma[2]*x1 + gamma[3]*x2 + phi*vi)
Ti <- rweibull(k, shape=Tau, scale=1/lambda0^(1/Tau))
lambda1 <- exp(lambdaC + gamma[2]*x1 + gamma[3]*x2 + phi*vi)
C <- rweibull(k, shape=Tau, scale=1/lambda1^(1/Tau))
Ci <- pmin(C, 180) # censoring time

```

```

#===== observed survival time
si <- pmin(Ti, Ci)
event <- ifelse(si==Ti, 1, 0)
fol.si <- as.integer(si/15)
round(data.frame(id, si, fol.si, event, x1, x2, vi, Ti, Ci), digits=4)
}

surv.dat0 <- surv.dat.frailty(k=100, gamma=c(-6, 1.5, 0.5),
lambdaC=-6.5, Tau=0.8, sigmasq.u2=4, sigmasq.u=0.5, sigmasq.e=0.5,
phi=0.5, alpha=0.5)
#===== Generated longitudinal data based on LME model =====
long.data <- function(dat=surv.dat0, beta=c(10,-0.2,0.6), sigmasq.e=0.5)
{
  fol.Ti <- dat$fol.si
  ni<- length(fol.Ti)
  id <- NULL
  tj <- NULL
  x1 <- NULL
  x2 <- NULL
  ui <- NULL
  eij <- NULL
  Time <- NULL
  event <- NULL
  for(j in 1:ni)
  {
    tij <- c(0:fol.Ti[j])
    tj <- c(tj, tij)
    l.j <- length(tij)
    id <- c(id, c(rep(dat$id[j], l.j )))
    x1 <- c(x1, c(rep(dat$x1[j], l.j )))

```

```

x2 <- c(x2, c(rep(dat$x2[j], 1.j )))
Time <- c(Time, c(rep(dat$si[j], 1.j )))
event <- c(event, c(rep(dat$event[j], 1.j )))
ui <- c(ui, c(rep(dat$vi[j], 1.j )))
ej <- round(rnorm(1.j, mean = 0 , sd = sqrt(sigmasq.e)), digits=4)
eij<- c(eij, ej)
yij <- beta[1] + beta[2]*tj + beta[3]*tj*x2 + ui + eij
}

data.frame(id=id, tj=tj, Time=Time, event=event, yij=yij, x1=x1,
x2=x2, ui=ui,eij=eij)
}

long.dat0 <- long.data(dat=surv.dat0, beta=c(10,-0.2,0.6), sigmasq.e=0.5)
#===== Simpson's Rule for integration =====
simpson <- function(vec, a, b) ### see simpson's rule pdf
{
# used for numerical integration
# uses Simpson's rule; function values are elements of vec
# length of vec should be odd
n <- length(vec)
m <- n - 1
h <- (b - a)/m
d <- c(1, rep(c(4, 2), m/2 - 1), 4, 1)
integral <- (h/3) * sum(d * vec)
list(integral = integral)
}

#==== ML estimates of Joint Models under Misspecified Random
Effect (u_{i} ~N(0,sigmasq.u) =====
JointModel.like <- function(dat1 = surv.dat0, dat2 = long.dat0,
beta=c(10.00, -0.2, 0.6), gamma= c(-6.0, 1.5, 0.5), Tau= 0.8, phi=0.5,

```

```

    sigmasq.u=0.5, sigmasq.e=0.5)
# it uses the function "optim" to minimize log-likelihood
{
  id <- dat2$id
  y <- dat2$yij
  x <- dat2$x2
  tj <- dat2$tj
  unique.id <- unique(dat2$id)
  k <- length(unique.id)
  a1 <- -5      # lower limit of u
  a2 <- 5       # upper limit of u
  u <- seq(a1, a2, length=151)
  initial = c(beta, gamma, Tau, phi, sigmasq.u, sigmasq.e)
  obj.fn <- function(theta)
  {
    beta <- theta[1:3]
    gamma <- theta[4:6]
    Tau <- theta[7]
    phi <- theta[8]
    sigmasq.u <- theta[9]
    sigmasq.e <- theta[10]
    fu <- dnorm(u, 0, sqrt(sigmasq.u))
    log.like <- 0
    for (i in 1:k)
    {
      #===== f(y|u)
      yi <- dat2$yij[dat2$id==unique.id[i]]
      xi <- dat2$x2[dat2$id==unique.id[i]]
      tj <- dat2$tj[dat2$id==unique.id[i]]

```

```

n <- length(yi)
f.yij.u <- 1
for(j in 1:n)
{
yij <- yi[j]
xij <- xi[j]
tij <- tj[j]
tij.xij <- tij * xij
X <- c(1,tij, tij.xij)
muij <- c(t(X) %*% beta) + u
f.yij.u0 <- dnorm(yij, mean= muij, sd=sqrt(sigmasq.e))
f.yij.u <- f.yij.u * f.yij.u0
}

f.yi.u <- f.yij.u
#==== f(T|u)

x1 <- dat1$x1[dat1$id==unique.id[i]]
x2 <- dat1$x2[dat1$id==unique.id[i]]
x <- c(1, x1, x2)
Time <- dat1$si[dat1$id==unique.id[i]]
delta <- dat1$event[dat1$id==unique.id[i]]
const <- c((Tau * (Time^(Tau-1)))^(delta))
eta.i <- c(t(x) %*% gamma) + phi * u
sum.i <- delta * eta.i - (Time^(Tau)) * exp(eta.i)
f.t.u <- const * exp(sum.i)
#===== likelihood function f.t.u * f.y.u * fu
like.i <- simpson(f.t.u * f.yi.u * fu, a1, a2)$integral
log.like.i <- log(like.i)
log.like <- log.like + log.like.i
}

```

```

-log.like
}

sigmasq.u=0.5, sigmasq.e=0.5

like.fit <- optim(par = initial, fn=obj.fn, hessian=TRUE,
method=c("L-BFGS-B"), control = list(maxit=500),
               lower=c(9.5 , -0.3, 0.4, -7.5, 1, 0, 0.6, 0.10, 0, 0),
               upper=c(10.5, -0.1, 0.8, -4.5, 2, 1, 1.0, 0.90, Inf, Inf))

cat(like.fit$message, "\n")

estimate <- round(like.fit$par, digits=4)

std.err <- round(sqrt(diag(solve(like.fit$hessian))), digits=4)

objective <- like.fit$value

list(estimate=estimate, std.err=std.err, objective=objective)
}

fit3 <- JointModel.like(dat1 = surv.dat0, dat2 = long.dat0,
  beta=c(10.00, -0.2, 0.6), gamma= c(-6.0, 1.5, 0.5), Tau= 0.8, phi=0.5,
  sigmasq.u=0.5, sigmasq.e=0.5)

#==== ML estimates of Joint Models under Misspecified
Random Effect ( $u_{\{i\}} \sim t\text{-dist}$  ====

#=== we generate data under assumption:  $[u_i \sim t\text{-dist}]$ :  $u_i \sim \text{Chi}^2$  ,df=3
#==+ we want to estimate survival submodel parameters under assumption:
 $u_i \sim t(0, v/v-2)$ 

JointModel.like <- function(dat1 = surv.dat0, dat2 = long.dat0, v=5,
  beta=c(10.00, -0.2, 0.6), gamma= c(-6.0, 1.5, 0.5), Tau= 0.8, phi=0.5,
  sigmasq.u=0.5, sigmasq.e=0.5)
# it uses the function "optim" to minimize log-likelihood
{
id <- dat2$id
y <- dat2$yij

```

```

x <- dat2$x2
tj <- dat2$tj
unique.id <- unique(dat2$id)
k <- length(unique.id)
a1 <- -5      # lower limit of u
a2 <- 5       # upper limit of u
u <- seq(a1, a2, length=151)
initial = c(beta, gamma, Tau, phi, sigmasq.u, sigmasq.e)
obj.fn <- function(theta)
{
  beta <- theta[1:3]
  gamma <- theta[4:6]
  Tau <- theta[7]
  phi <- theta[8]
  sigmasq.u <- theta[9]
  sigmasq.e <- theta[10]
  fu <- dt(u, v)  # density of u with v df
  log.like <- 0
  for (i in 1:k)
  {
    #===== f(y|u)
    yi <- dat2$yij[dat2$id==unique.id[i]]
    xi <- dat2$x2[dat2$id==unique.id[i]]
    tj <- dat2$tj[dat2$id==unique.id[i]]
    n <- length(yi)
    f.yij.u <- 1
    for(j in 1:n)
    {
      yij <- yi[j]

```



```

xij <- xi[j]
tij <- tj[j]
tij.xij <- tij * xij
X <- c(1,tij, tij.xij)
#=== u follows a t-distribution with v d.f.
muij <- c(t(X) %>% beta) + sqrt(sigmasq.u) * sqrt((v-2)/v) * u
f.yij.u0 <- dnorm(yij, mean= muij, sd=sqrt(sigmasq.e))
f.yij.u <- f.yij.u * f.yij.u0
}
f.yi.u <- f.yij.u
#===== f(T|u)
x1 <- dat1$x1[dat1$id==unique.id[i]]
x2 <- dat1$x2[dat1$id==unique.id[i]]
x <- c(1, x1, x2)
Time <- dat1$si[dat1$id==unique.id[i]]
delta <- dat1$event[dat1$id==unique.id[i]]
const <- c((Tau * (Time^(Tau-1)))^(delta))
#== u follows a t-distribution with v d.f.
eta.i <- c(t(x) %>% gamma) + phi * sqrt(sigmasq.u) * sqrt((v-2)/v)*u
sum.i <- delta * eta.i - (Time^(Tau)) * exp(eta.i)
f.t.u <- const * exp(sum.i)
#===== likelihood function f.t.u * f.y.u * fu
like.i <- simpson(f.t.u * f.yi.u * fu, a1, a2)$integral
log.like.i <- log(like.i)
log.like <- log.like + log.like.i
}
-log.like
}
like.fit <- optim(par = initial, fn=obj.fn, hessian=TRUE,

```

```

method=c("L-BFGS-B"), control = list(maxit=500),
      lower=c(9.5 , -0.3, 0.4, -7.5, 1, 0, 0.6, 0.10, 0, 0),
      upper=c(10.5, -0.1, 0.8, -4.5, 2, 1, 1.0, 0.90, Inf, Inf))
cat(like.fit$message, "\n")
estimate <- round(like.fit$par, digits=4)
std.err <- round(sqrt(diag(solve(like.fit$hessian))), digits=4)
objective <- like.fit$value
list(estimate=estimate, std.err=std.err, objective=objective)
}
fit4 <- JointModel.like(dat1 = surv.dat0, dat2 = long.dat0, v=5,
  beta=c(10.00, -0.2, 0.6), gamma= c(-6.0, 1.5, 0.5), Tau= 0.8,
  phi=0.5, sigmasq.u=0.5, sigmasq.e=0.5)

```

Bibliography

- [1] Donald I Abrams, Anne I Goldman, Cynthia Launer, Joyce A Korvick, James D Neaton, Lawrence R Crane, Michael Grodesky, Steven Wakefield, Katherine Muth, Sandra Kornegay, et al. A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection. *New England Journal of Medicine*, 330(10):657–662, 1994.
- [2] Per Kragh Andersen and Richard D Gill. Cox’s regression model for counting processes: a large sample study. *The annals of statistics*, pages 1100–1120, 1982.
- [3] Anton A Béguin and Ceec AW Glas. Mcmc estimation and some model-fit analysis of multidimensional irt models. *Psychometrika*, 66(4):541–561, 2001.
- [4] Daniel M Bolt and Venessa F Lall. Estimation of compensatory and noncompensatory multidimensional item response models using markov chain monte carlo. *Applied Psychological Measurement*, 27(6):395–414, 2003.
- [5] Elizabeth R Brown. Assessing the association between trends in a biomarker and risk of event with an application in pediatric hiv/aids. *The annals of applied statistics*, 3(3):1163, 2009.
- [6] Helen Brown and 1962-author Brown, Helen. *Applied Mixed Models in Medicine*. Statistics in practice. Wiley, 2014.
- [7] David Byar, Clyde Blackard, Veterans Administration Cooperative Urological Research

- Group, et al. Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage i bladder cancer. *Urology*, 10(6):556–561, 1980.
- [8] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.
- [9] Collett David. *Modelling Survival Data in Medical Research*. Chapman and Hall/CRC, 2003.
- [10] Peter Diggle, Peter J Diggle, Patrick Heagerty, Kung-Yee Liang, Patrick J Heagerty, Scott Zeger, et al. *Analysis of longitudinal data*. Oxford University Press, 2002.
- [11] Peter J Diggle, Inês Sousa, and Amanda G Chetwynd. Joint modelling of repeated measurements and time-to-event outcomes: The fourth armitage lecture. *Statistics in Medicine*, 27(16):2981–2998, 2008.
- [12] Il Do Ha, Youngjo Lee, and Jae-Kee Song. Hierarchical-likelihood approach for mixed linear models with censored data. *Lifetime data analysis*, 8(2):163–176, 2002.
- [13] Cox DR. Cox proportional-hazards regression for survival data. *J R Stat Soc*, 34(2):187–220, 1972.
- [14] Luc Duchateau and Paul Janssen. *The frailty model*. Springer Science & Business Media, 2007.
- [15] Martin R Edwards. An integrative review of employer branding and ob theory. *Personnel review*, 2010.
- [16] Cheryl L Faucett and Duncan C Thomas. Simultaneously modelling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in medicine*, 15(15):1663–1685, 1996.
- [17] Garrett M. Fitzmaurice. *Longitudinal data analysis*. CRC Press/Taylor Francis, 2012.

- [18] Garrett M. Fitzmaurice, Nan M. Laird, and James H. Ware. *Applied longitudinal analysis*. Wiley, 2011.
- [19] A.I. Goldman, B.P. Carlin, L.R. Crane, C. Launer, J.A. Korvick, L. Deyton, and D.I. Abrams. Response of cd4 lymphocytes and clinical consequences of treatment using ddi or ddc in patients with advanced hiv infection. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 11(2):161–169, 1996.
- [20] Victor De Gruttola and Xin Ming Tu. Modelling progression of cd4-lymphocyte count and its relationship to survival time. *Biometrics*, 50, 1994.
- [21] Il Do Ha and Youngjo Lee. Estimating frailty models via poisson hierarchical generalized linear models. *Journal of Computational and Graphical Statistics*, 12(3):663–681, 2003.
- [22] Il Do Ha, Youngjo Lee, and Jae-kee Song. Hierarchical likelihood approach for frailty models. *Biometrika*, 88(1):233–233, 2001.
- [23] Il Do Ha, Maengseok Noh, and Youngjo Lee. H-likelihood approach for joint modeling of longitudinal outcomes and time-to-event data. *Biometrical Journal*, 59(6):1122–1143, 2017.
- [24] N Haddad, A Robert, A Weeks, N Popovic, W Siu, and C Archibald. Hiv: Hiv in canada—surveillance report, 2018. *Canada Communicable Disease Report*, 45(12):304, 2019.
- [25] Robin Henderson, Peter Diggle, and Angela Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.
- [26] Robert S Hogg, Benita Yip, Charlotte Kully, Kevin JP Craib, Michael V O’Shaughnessy, Martin T Schechter, and Julio SG Montaner. Improved survival among hiv-infected patients after initiation of triple-drug antiretroviral regimens. *Cmaj*, 160(5):659–665, 1999.

- [27] Fushing Hsieh, Yi-Kuan Tseng, and Jane-Ling Wang. Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, 62(4):1037–1043, 2006.
- [28] Joseph G Ibrahim, Haitao Chu, and Liddy M Chen. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16):2796, 2010.
- [29] Jonathan M Kagan, Ana M Sanchez, Alan Landay, and Thomas N Denny. A brief chronicle of cd4 as a biomarker for hiv/aids: A tribute to the memory of john l. fahey. In *Forum on immunopathological diseases and therapeutics*, volume 6. Begel House Inc., 2015.
- [30] EL Kaplan and P Meier. Nonparametric estimation from incomplete observations j am stat assoc 53: 457–481. *Find this article online*, 1958.
- [31] John A Kellum, Lan Kong, Mitchell P Fink, Lisa A Weissfeld, Donald M Yealy, Michael R Pinsky, Jonathan Fine, Alexander Krichevsky, Russell L Delude, and Derek C Angus. Understanding the inflammatory cytokine response in pneumonia and sepsis: results of the genetic and inflammatory markers of sepsis (genims) study. *Archives of internal medicine*, 167(15):1655–1663, 2007.
- [32] John P Klein and Melvin L Moeschberger. *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2006.
- [33] Nan M Laird, James H Ware, et al. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.
- [34] Tony Lancaster. Econometric methods for the duration of unemployment. *Econometrica: Journal of the Econometric Society*, pages 939–956, 1979.
- [35] Jerald F Lawless. *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons, 2011.

- [36] Youngjo Lee and John A Nelder. Hierarchical generalized linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(4):619–656, 1996.
- [37] Youngjo Lee and John A Nelder. Modelling and analysing correlated non-normal data. *Statistical Modelling*, 1(1):3–16, 2001.
- [38] Youngjo Lee, John A Nelder, and Yudi Pawitan. *Generalized linear models with random effects: unified analysis via H-likelihood*. Chapman & Hall/CRC, 2006.
- [39] Menghan Li, Ching-Wen Lee, and Lan Kong. A latent class approach for joint modeling of a time-to-event outcome and multiple longitudinal biomarkers subject to limits of detection. *Statistical methods in medical research*, 29(6):1624–1638, 2020.
- [40] Ning Li, Robert M Elashoff, Gang Li, and Jeffrey Saver. Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the ninds rt-pa stroke trial. *Statistics in medicine*, 29(5):546–557, 2010.
- [41] K.Y. Liang and S.L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.
- [42] HJ Lim, Prosanta Mondal, and Stuart Skinner. Joint modeling of longitudinal and event time data: application to hiv study. *Journal of Medical Statistics and Informatics*, 1(1):1, 2013.
- [43] Haiqun Lin, Bruce W Turnbull, Charles E McCulloch, and Elizabeth H Slate. Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association*, 97(457):53–65, 2002.
- [44] Mary J. Lindstrom and Douglas M. Bates. Newton-raphson and em algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83:1014–1022, 1988.

- [45] Benoit Liqueur, Jean-François Timsit, and Virginie Rondeau. Investigating hospital heterogeneity with a multi-state frailty model: application to nosocomial pneumonia disease in intensive care units. *BMC medical research methodology*, 12(1):79, 2012.
- [46] Lei Liu. Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. *Statistics in medicine*, 28(6):972–986, 2009.
- [47] Lisa M McCrink, Adele H Marshall, and Karen J Cairns. Advances in joint modelling: a review of recent developments with application to the survival of end stage renal disease patients. *International Statistical Review*, 81(2):249–269, 2013.
- [48] Charles E McCulloch. *Generalized, linear, and mixed models*. Wiley series in probability and statistics. Wiley, Hoboken, N.J., 2nd ed. edition, 2008.
- [49] Geert Molenberghs and Geert Verbeke. *Linear mixed models for longitudinal data*. Springer, 2000.
- [50] Geert Molenberghs and Geert Verbeke. *Models for discrete longitudinal data*. Springer Science & Business Media, 2005.
- [51] D.F. Moore. *Applied Survival Analysis Using R*. Use R! Springer International Publishing, 2016.
- [52] Gert G Nielsen, Richard D Gill, Per Kragh Andersen, and Thorkild IA Sørensen. A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian journal of Statistics*, pages 25–43, 1992.
- [53] Maengseok Noh and Youngjo Lee. Reml estimation for binary data in glms. *Journal of Multivariate Analysis*, 98(5):896–915, 2007.
- [54] Richard Peto and Julian Peto. Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society: Series A (General)*, 135(2):185–198, 1972.

- [55] Pete Philipson, Peter Diggle, Ines Sousa, Ruwanthi Kolamunnage-Dona, Paula Williamson, and Robin Henderson. joiner: Joint modelling of repeated measurements and time-to-event data. 2012.
- [56] Ross L Prentice. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69(2):331–342, 1982.
- [57] Cécile Proust-Lima, Mbéry Sène, Jeremy MG Taylor, and Hélène Jacqmin-Gadda. Joint latent class models for longitudinal and time-to-event data: A review. *Statistical methods in medical research*, 23(1):74–90, 2014.
- [58] Cécile Proust-Lima and Jeremy MG Taylor. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. *Biostatistics*, 10(3):535–549, 2009.
- [59] Hein Putter and Hans C van Houwelingen. Frailties in multi-state models: Are they identifiable? do we need them? *Statistical methods in medical research*, 24(6):675–692, 2015.
- [60] Elizabeth R. Brown and Joseph G. Ibrahim. A bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59(2):221–228, 2003.
- [61] Samuli Ripatti and Juni Palmgren. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56(4):1016–1022, 2000.
- [62] D. Rizopoulos. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Chapman & Hall/CRC Biostatistics Series. CRC Press, 2012.
- [63] Dimitris Rizopoulos. Jm: An r package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software (Online)*, 35(9):1–33, 2010.
- [64] Dimitris Rizopoulos and Pulak Ghosh. A bayesian semiparametric multivariate joint

- model for multiple longitudinal outcomes and a time-to-event. *Statistics in medicine*, 30(12):1366–1380, 2011.
- [65] Kristina E Rudd, Sarah Charlotte Johnson, Kareha M Agesa, Katya Anne Shackelford, Derrick Tsoi, Daniel Rhodes Kievlan, Danny V Colombara, Kevin S Ikuta, Niranjana Kissoon, Simon Finfer, et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the global burden of disease study. *The Lancet*, 395(10219):200–211, 2020.
- [66] Abdus Sattar and Sanjoy K Sinha. Joint modeling of longitudinal and survival data with a covariate subject to a limit of detection. *Statistical methods in medical research*, 28(2):486–502, 2019.
- [67] Daniel O Scharfstein, Andrea Rotnitzky, and James M Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120, 1999.
- [68] Mark D Schluchter. Methods for the analysis of informatively censored longitudinal data. *Statistics in medicine*, 11(14-15):1861–1870, 1992.
- [69] Steve Self and Yudi Pawitan. Modeling a marker of disease progression and onset of disease. In *AIDS epidemiology*, pages 231–255. Springer, 1992.
- [70] Steve Self and Yudi Pawitan. Modeling a marker of disease progression and onset of disease. In *AIDS epidemiology*, pages 231–255. Springer, 1992.
- [71] Mervyn Singer, Clifford S Deutschman, Christopher Warren Seymour, Manu Shankar-Hari, Djillali Annane, Michael Bauer, Rinaldo Bellomo, Gordon R Bernard, Jean-Daniel Chiche, Craig M Coopersmith, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *Jama*, 315(8):801–810, 2016.

- [72] Xiao Song, Marie Davidian, and Anastasios A Tsiatis. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, 58(4):742–753, 2002.
- [73] Inês Sousa. A review on joint modelling of longitudinal measurements and time-to-event. *Revstat Stat J*, 9:57–81, 2011.
- [74] Leonard A. Stefanski and Raymond J. Carroll. Conditional scores and optimal scores for generalized linear measurement- error models. *Biometrika*, 74:703–716, 1987.
- [75] Michael J Sweeting and Simon G Thompson. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53(5):750–763, 2011.
- [76] Terry Therneau and Patricia Grambsch. *Modeling Survival Data: Extending The Cox Model*, volume 48. 01 2000.
- [77] Terry M. Therneau and Patricia M. Grambsch. Modeling survival data: Extending the cox model. *Springer-Verlag, US*, 2000.
- [78] Anastasios A Tsiatis and Marie Davidian. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88(2):447–458, 2001.
- [79] Anastasios A Tsiatis and Marie Davidian. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, pages 809–834, 2004.
- [80] Anastasios A Tsiatis, Victor Degruttola, and Michael S Wulfsohn. Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, 90(429):27–37, 1995.

- [81] Anastasios A Tsiatis, Victor Degruttola, and Michael S Wulfsohn. Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, 90(429):27–37, 1995.
- [82] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.
- [83] James W Vaupel, Kenneth G Manton, and Eric Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454, 1979.
- [84] L. J. Wei, D. Y. Lin, and L. Weissfeld. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. 84(408):1065–1073, 1989.
- [85] Andreas Wienke. *Frailty models in survival analysis*. CRC press, 2010.
- [86] Eleanor MP Wilson and Irini Sereti. Immune restoration after antiretroviral therapy: the pitfalls of hasty or incomplete repairs. *Immunological reviews*, 254(1):343–354, 2013.
- [87] Lang Wu. *Mixed effects models for complex data*. CRC Press, 2010.
- [88] Lang Wu, Wei Liu, Grace Y Yi, and Yangxin Huang. Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 2012, 2012.
- [89] Lang Wu and Tingting Yu. Joint modeling of longitudinal and survival data. *Wiley StatsRef: Statistics Reference Online*, pages 1–9, 2016.
- [90] Michael S Wulfsohn and Anastasios A Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, pages 330–339, 1997.
- [91] Menggang Yu, Ngayee J Law, Jeremy MG Taylor, and Howard M Sandler. Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, pages 835–862, 2004.