Deanship of Graduate Studies

Al-Quds University

JOINT MODELLING OF LONGITUDINAL AND SURVIVAL DATA

Nareman Mahmoud Amro

M.Sc. Thesis

Jerusalem-Palestine

1434/2013

JOINT MODELLING OF LONGITUDINAL AND SURVIVAL DATA

By

Nareman Mahmoud Amro

B.sc.: College of Science and Technology

Al-Quds University/Palestine

Supervisor: Khalid Ali Salah, PhD.

This thesis is submitted in Partial fulfillment of requirements of the degree of Master of Science, Department of Mathematics / Program of Graduate studies.

Al-Quds University

2013

The Program of Graduate Studies /Department of Mathematics
Deanship of Graduate Studies

JOINT MODELLI.NG OF LONGITUDINAL AND SURVIVAL DATA

By

Student Name: Nareman Amro

Registration Number: 21011455

Supervisor: Dr. Khalid Ali Salah

Master thesis submitted and accepted data:

The names and signatures of the examining committee members are as follows

1. Dr. Khalid Salah Head of Committee, Signatures……
2. Dr. Bader Al-Jawadi External Examiner, Signatures……
3. Dr. Abdulhakeem Eideh Internal Examiner, Signatures……

Al-Quds University

1434/2013

Declaration

I certify that the thesis, submitted for the degree of master, is the result of my own research except where otherwise acknowledged, and that the thesis (or any part of the same) has not been submitted for a higher degree to any other university or institution.

Signed……

Nareman Amro

Data:

22/12/2013

Dedication

To my father, Mahmoud.

To my mother, Latifa.

To my brothers, Ali, Haytham, Thaer, Mohammad, Saed, Ismail, and Ala’a.

To my sisters, Eman, Neda’a, Feda’a, Afnan, and Ezyah.

To my brother and sister sons, Mahmoud, Rahmeh, Anan.

To my friend

To my colleagues "teachers"

Acknowledgments

This work in this thesis has been carried out at the Department of Mathematics at Al-Quds University. I cannot fully express my gratitude to my supervisor Dr. Khalid Ali Salah. He is an exceptional academic advisor, supporting, extraordinary leader and ready any time to devote his time and efforts to help his students. In the same time he set an example for me by being just, open and honest, kind and gentle person, devoted and caring constant source of wisdom and experience. I really appreciate his patience, encouragement and valuable suggestions.

I am grateful to all the following doctors who taught me during the MA degree Dr. Yousef Zahalqia, Dr. Ibrahim Qrouz, Dr. Jameel Jamal, Dr. Taha Abu Kaf, Dr. Abdulhakeem Eideh, Dr. Mohammed Kaleel, Dr. Yousif Bedar, and Dr. Tahseen Almougrabi.

Abstract

Analysis involving longitudinal and time-to-event data are quite common in medical research. The primary goal of such studies is to simultaneously study the effect of treatment on both the longitudinal covariate and survival.

The early part of the thesis focuses on common model of longitudinal and the survival data, it is called joint model. In this joint model, the longitudinal submodel is a combination of a random mixed effect model. A semi-parametric submodel is also proposed to incorporate the true longitudinal trajectories and other baseline time covariates. Moreover, we consider a Bayesian approach which is motivated by the complexity of the model. Posterior and prior specification needs to accommodate parameter constraints due to the non-negativity of the survival function.

Bayesian approach was applied to the data using a conjugate and non conjugate prior families to obtain parameter estimates. Gibbs sampling is used to find value for estimates parameters $\sigma_\varepsilon^2, \sigma_a^2, \mu_a, (h_{01}, \dots, h_{0K})$ in the joint model and Metropolis Hasting are used to update the Markov chain to estimate parameters β_l, b, γ, ω whose full conditional densities cannot be sampled efficiently from the existing methods, leading us to propose efficient proposal densities.

The simulation studies demonstrate that the joint modeling method results in quite accurate, efficient and small biases for all the parameters. The analysis of real data indicates that when ignoring the association between the longitudinal and the survival data would lead to biased estimates for the most important parameters than joint model.

النمذجة المشترك للبيانات الطولية والبقاء

اعداد: ناريمان محمود عمرو

اشراف: د. خالد علي صلاح

الملخص

التحليل التي تشمل longitudinal and time-to-event dataهي شائعة لحد كبير في الابحاث الطبية. الهدف الأساسي من هذه الدراسات في وقت واحد لدراسة تأثير العلاج على survival و longitudinal covariate .

في الجزء الاول من الرسالة يركز علي survival dataو longitudinal بدمجهما يعطي ميسميjoint model .

في هذا joint model، النموذج الفرعي longitudinal ومزيج من random mixed effect model و stochastic process. النموذج الفرعي semi-parametric ايضا يقترح لدمج longitudinal trajectories صحيح و baseline time covariates . حسابات هذا النموذج لاحتمال موضوع متي يتم ألشفاء لطبيعة الوحيدة longitudinal data. علاوة علي ذلك ندرسBayesian approach التي تعمل علي تعقيد النماذج. الذي يحتاج الي مواصفات prior و posterior لاستيعاب معوقات الباروميتر نظرا لعدم السلبية من survival function.تم تطبيق Bayesian approach للبيانات باستخدام non-conjugate prior وconjugate وذلك للحصول علي تقدير للباروميتر في النموذج المستخدم للدراسة.Gibbs sampling يستخدم لإيجاد قيمه لتقدير الباروميترات $\sigma_\varepsilon^2, \sigma_a^2, \mu_a, (h_{01}, \dots, h_{0K})$ في joint model . Metropolis Hasting يستخدم لتحديث سلسلة ماركوف لتقدير الباروميترات $\beta_l, b, \gamma, \omega$ التي full conditional densities لا يمكن اخذ عينات بكفاءة من الاساليب المستخدمة مما يؤدي بنا الي اقتراح اقتران كثافة يكون كفء.

تثبت الدراسات أن نتائج محاكاة أسلوب joint modeling. دقيقه جدا وفعاله وصغير التحيز لكل البارومترات. تحليل البيانات الحقيقيه تشير إلى أنه عندما تجاهل الرابطة بين the survival data و longitudinal من شئنه ان يؤدي الي تقديرات منحازة للباروميترات الأكثر أهمية مقارنة مع joint model .

TABLE OF CONTENTS

List of Tables

List of Figures

LIST OF SYMBOLS

All unknown parameter	Ω
All unknown parameter in Ω except c	Ω_{-c}
Any one unknown parameter	c
Baseline hazard function	h_0
Candidate point	θ^*
Censoring time	C_i
Cumulative distribution function	$F(.)$
Cumulative hazard function	H
Current value	θ
Design matrix of fixed effects	$\boldsymbol{A_i}$
Design matrix of random effects	$\boldsymbol{B_i}$
Deviance information criterion function	D
Different subject	i
Effect of the variables	β_l
Error term	ε_i
Estimated survival function	$\hat{S}$
Estimation Monte Carlo	$\hat{I}$
Expectation	$E(.)$
Explanatory variable covariates (independent)	X_i
Exponential	e
Fixed effects variable	β
Fixed slop	b
Gamma distribution	$G(,)$
Hazard function	h_i
Indicator	δ_i
Initial value	θ_0
Initial value all unknown parameter	Ω^0
Intercept mean	μ_a
Inverse Gamma distribution	$IG(,)$
Likelihood function	L
Location parameter	λ
Mean	μ
Measurements of continuous time dependent covariate	Y
Minimum of the true even time and censoring time	V_i
Monte Carlo integration	$I(.)$
Normal distribution	$N(,)$
Number of estimated parameter	p
Number of individual at risk at time	r_j
Number of individual who experience the event of interest at time	d_j
Observation on the subject	j
Observed value of the marker for subject	Y_i
Order time	$t_{(}$
Point estimated of deviance function	$\widehat{D}$

Posterior distribution	$[./.]$
Posterior mean of deviance function	$\overline{D}$
Potential covariates	P_i
Prior distribution	$[.]$
Probability density function	$f(.)$
Proportional to	$\propto$
Random effects variable	b_i
Random intercept	a_i
Ratio of density	r
Regardless of the time	ϑ
Regression coefficient of disease risk	γ
Regression coefficient of potential covariate	ω
Remaining covariates after l^{th} covariate X_{il}	$X(-l)$
Set observation explanatory variables	x_i
Shape parameter	α
Survival function	S_i
Survival time	s_i
The observed the multivariate response with q components	$\boldsymbol{Y_i}$
Time at observation	t_{ij}
True observed failure time for subject	T_i
True un observed time dependent value	Z
True value of the marker	Z_i
True value of the multivariate	$\boldsymbol{Z_i}$
Uniform distribution	$U(,)$
Unknown parameter	β_k
Variance	σ^2
Variance error	σ_ε^2
Variance intercept	σ_a^2
Vector of measurement error	$\boldsymbol{\varepsilon_i}$

List of Abbreviations

ABBRVIATIONS	MEANING
AIC	Akaike Information Criterion
BP	Bias Percentile
CDF	Cumulative Distribution Function
CCR	Confidence Converge Rate
CD4	Cluster of differentiation 4
Cov	Covariance of the response variable
DIC	Deviance Information Criterion
DPCA	Sub set of PBC
EM	Expectation Maximization
HIV	Human immunod efficiency virus
Icmm	Estimation of latent class mixed model
Jm	Joint modeling
Log	Logarithm
MCE	Monte Carlo Error
MCMC	Markov Chain Monte Carlo
MCSD	Monte Carlo Standard Deviation
M-H	Metropolis-Hastings
ML	Maximum Likelihood
MSE	Mean Square Error
PDF	Probability Density Function
PBC	Primary Biliary Cirrhosis
PL	Partial Likelihood
PSA	Prostate specific antigen
SE	Standard Error
Var	Variance

CHAPTER ONE
Introduction

1.1 Introduction

In this chapter, we give an overview of the longitudinal studies and some literature review of people who have worked in this field, addition to the objectives of this study. Researchers are often interested in the associations between the longitudinal and survival data. Often in applied statistics, after some empirical data have been collected, the purpose of the analysis is to construct a statistical model. Otherwise; said, we are interested in situations where the aim is to explain how an outcome, or response, variable of particular interest is related to a set of explanatory variables, or covariates.

In longitudinal studies in the medical research areas, patients are following over time. In this way researchers can make repeated measurements and observe the changes in certain markers of a disease and the time to certain clinical events. Hence the resulting data consist of both the longitudinal data in the usual sense and the survival (time to event) data. Clearly the changes in clinical markers and decline times may not be independent.

Longitudinal data is data in the form of repeated measurements on the same unit over time. Data are routinely collected in this fashion in a broad range of applications, including agriculture, life sciences, medical and public health research, and physical science and engineering. The main reason and advantage of longitudinal analysis is to study the change over time. That is also how longitudinal analysis differs from repeated measures analysis. In longitudinal analysis, we model both the dependence of the response on the covariates and the associations among responses.

Longitudinal study has the ability to distinguish the variation in the outcomes across time for an individual from the ones among the population. To model the random variability in the longitudinal models with continuous outcomes, Diggle, Liang, and Zeger (2002) distinguish among three components of variability: random effects, serial association and measurement errors.

In many studies, multivariate outcomes are observed and hence multivariate longitudinal models are necessary. Many studies have discrete outcome variables which renders traditional likelihood-based methods that require the multivariate normality assumptions and cumbersome with time-varying covariates inapplicable. Three modern analyses approaches have been developed over the years for the analysis of longitudinal repeated measures study with discrete outcome variables. They are the marginal model, the nonlinear mixed effect model, and the transition model. Survival analysis examines and models the time it takes for events to occur. The event can be death, occurrence of a disease, marriage, divorce, etc. The time to event or survival time can be measured in days, weeks, years, etc. For example, if the event of interest is heart attack, then the survival time can be the time in years until a person develops a heart attack. Is widely used in many fields of studies such as medicine, biology engineering, pharmacology, epidemiology, and economics. In medical research, the events could be death, disease or failure of treatment.

Survival analysis focuses on the distribution of survival times and the estimation of the relationship between survival and one or more predictors.

Joint modeling of longitudinal and survival data can provide more efficient estimates and reduce bias in some situations.

In this research we first contrast the method for modeling the association between aspects of a longitudinal process and a time-to-event outcome. A naive approach, where repeated measurements are used as time-dependent covariates in a Cox (Cox, D. R., 1972) or parametric survival model, is compared to a shared random effects joint modeling approach. We then describe how a shared random effects joint model can be fit using Bayesian Markov Chan Monte Carlo (MCMC) methods in R software. Such software is flexible in allowing a number of possible associations between the longitudinal and time-to-event processes to be investigated, while predicted survival curves and future longitudinal values can easily be obtained for individual using posterior predictive distributions. Nevertheless, the Bayesian model implemented in R is currently not restricted to models in which the cumulative hazard has a closed form.

Finally, we apply the models to the primary biliary cirrhosis (PBC) data, which collected by the Mayo Clinic from 1974 to 1984, various biomarkers such as bilirubin, prothrombin time and albumin were collected longitudinally, our aim was to develop a model first to estimate the association between the risk of PBC disease and these biomarkers, and second to make predictions about the future rate of biomarkers and probability of PBC disease.

1.2 Objectives of The Thesis

The objectives of research can be summarized as follows:

1. To solidify the motivation and explanation of the survival regression specification of discrete and continuous time models for single event data, to formally examine and assess the association between the longitudinal and survival data.
2. To examine an appropriate form of a longitudinal model for describing the behavior of the longitudinal measures over time
3. To investigate and develop the statistical inference procedures based on the Bayesian approach, and then to using algorithms that utilize sampling scheme for fitting the survival model.
4. In order to illustrate the application of the univariate and multivariate survival models a simulation study and real data sets will be presented.

1.3 Structure of The Thesis

This thesis consists of four chapters. The first chapter is an introductory chapter in which definition longitudinal analysis, survival (time to event), joint modeling and Bayesian approaches.

In the second chapter, some background information about the longitudinal analysis is presented. Summarizes the common models in survival analysis. At the end of chapter, adapted joint modeling for longitudinal and survival data will present.

In the third chapter, the likelihood of the joint model and the full conditional distributions were presented. Also, Bayesian approaches through Markov Chain Monte Carlo Method (MCMC) for parameter estimations were considered in the same chapter.

A comparison study through a simulation study and real data analysis followed by conclusion of this study in the final chapter.

1.4 Literature Review

Methods jointly modeling longitudinal data and survival data have been studied in the literature. Tsiatis and Davidian (2004) reviewed the development of joint models, and described and contrasted some of earlier proposals for implementation and inference. Tseng, Hsieh, and Wang (2005) explored the joint modeling method under the accelerated failure time assumption when covariates are assumed to follow a linear mixed effects model with measurement errors. Their joint modeling method is based on maximizing the joint likelihood with random effects treated as missing data, with a Monte Carlo EM algorithm used to estimate all the unknown parameters. In Tseng, Hsieh, and Wang (2005), the two models shared unobserved error-free variables.

Faucett and Thomas (1996) took a Bayesian approach to joint models and developed and demonstrated implementation via Markov chain Monte Carlo (MCMC) techniques. Xu and Zeger (2001) used joint models as a framework in which to make more efficient inference on the marginal (given, say, baseline covariates such as treatment) event-time distribution by incorporating the longitudinal data as auxiliary information. Salah (2008) took a Bayesian approach to joint longitudinal and time to event with survival fraction.

Chapter Two

Joint Modeling of Longitudinal and Survival Data

2.1 Introduction

In this chapter focuses on modeling the longitudinal and the survival data. When association between the two models it is called joint model. In this joint model, the longitudinal submodel is a combination of a random mixed effect model and a semi-parametric submodel is also, a incorporate the true longitudinal trajectories and other baseline time covariates.

Longitudinal studies often produce two types of outcome, namely a set of longitudinal response measurements and the time to an event of interest, such as death, development of a disease or dropout from the study. Two typical examples of this setting are HIV and cancer studies. In HIV studies patients who have been infected are monitored until they develop AIDS or die, and they are regularly measured for the condition of the immune system using markers such as the CD4 lymphocyte count or the estimated viral load. Similarly in cancer studies the event outcome is the death or metastasis and patients also provide longitudinal measurements of antibody levels or of other markers of carcinogenesis, such as the PSA levels for prostate cancer.

These two outcomes are often separately analyzed using a mixed effects model for the longitudinal outcome and a survival model for the event outcome. However, in mainly two settings a joint modeling approach is required. First, when interest is on the event outcome and we wish to account for the effect of the longitudinal outcome as a time-dependent covariate, traditional approaches for analyzing time-to-event data (such as the

partial likelihood for the Cox proportional hazards models) are not applicable. In particular, standard time-to-event models require that time-dependent covariates are external, that is, the value of this covariate at time point t is not affected by the occurrence of an event at time point u, with $t > u$ (Kalbeisch and Prentice 2002, Section 6.3). However, the type of time-dependent covariates encountered in longitudinal studies do not satisfy this condition, due to the fact that they are the output of a stochastic process generated by the subject, which is directly related to the failure mechanism. Therefore, in order to produce valid inferences a model for the joint distribution of the longitudinal and survival outcomes is required instead. The second setting in which joint modeling is required is when interest is on the longitudinal outcome. In this case the occurrence of events causes dropout since no longitudinal measurements are available at and after the event time. When this dropout is nonrandom (i.e., when the probability of dropout depends on unobserved longitudinal responses, simple example concerning the occurrence of mastitis in dairy cows, in which the occurrence of mastitis can be modelled as a dropout process. It is shown through sensitivity analysis how the conclusions concerning the dropout mechanism depend crucially on untestable distributional assumptions. This example is exceptional in that from a simple plot of the data two outlying observations can be identified that are the source of the apparent evidence for non-random dropout and also provide an explanation of the behaviour of the sensitivity analysis. It is concluded that a plausible model for the data does not require the assumption of non-random dropout), then bias may arise from an analysis that ignores the dropout process, see Little and Rubin (2002, Chapter 15). To avoid this problem and obtain valid inferences the longitudinal and dropout process must be jointly modeled.

One of the modeling frameworks that have been proposed in the missing data literature for handling nonrandom dropout is the shared parameter model Follmann and Wu (1995), which postulates a mixed effects model for the longitudinal outcome and time-to-dropout model for the missingness process. In both settings the joint distribution of the event times and the longitudinal measurements is modeled via a set of random effects that are assumed to account for the associations between these two outcomes. Excellent overviews of this area of biostatistics and statistics research are given by Tsiatis and Davidian (2004), and Yu, Law, Taylor, and Sandler (2004).

The joint model describe in this study is very comprehensive and is applicable to a data set containing the following information for each $(i = 1, \dots, m)$ subjects:

1. Measurements $Y(t_{ij})$ $(j = 1, \dots, n_i)$ of a continuous time-dependent covariate (or disease marker), possibly measured with error at observation times t_{ij} relative to some baseline time (e.g. time of entry the study, or time of infection).
2. A disease status indicator δ_i with $\delta_i = 1$ indicating development of disease (time to event) and $\delta_i = 0$ indicating absence of disease at the end of follow up (censored).
3. Time V_i from baseline to the onset of disease or total length of follow up if a subject is censored.
4. Time fixed or time-dependent measurements $X_i = X_i(t_{ij})$ of other variables, such as treatment status, age, weight, etc. If it is time fixed $X_i(t_{ij})$ is constant and does not depend on t_{ij}.

The adapted joint model in this study contains two submodels: a covariate tracking model (longitudinal model) and a disease risk model (survival model). The longitudinal model

relates the measurements $Y(t_{ij})$ to the true unobserved time-dependent values $\mathrm{Z}(t_{ij})$ and describes how the true unobserved covariates proves evolves with time and in relation to other factors. The survival model relates the risk of disease to the true underlying covariate history and other factors.

Before formulating the joint model in detail, we review briefly some basics properties of longitudinal and survival models.

2.2 Longitudinal Data

Following Henderson et al. (2000) and Wang, Y. and Taylor, J.M.G. (2001), to describe the longitudinal model, the response measurements are modeled using a mixed-effects model, a popular and flexible choice for continuous longitudinal data. With the notation introduced at the beginning of this chapter the adapted model for longitudinal data can be written as

$$\begin{aligned} Y_i(t_{ij}) &= Z_i(t_{ij}) + \varepsilon_i(t_{ij}) \\ Z_i(t_{ij}) &= a_i + bt_{ij} + \beta X_i(t_{ij}) \end{aligned} \qquad (2.1)$$

Where $Y_i(t_{ij})$ denoted the observed value of the marker for subject i at time t_{ij}, $X_i(t_{ij}) = \left(X_{i1}(t_{ij}), \dots, X_{ip}(t_{ij})\right)^T$ a ($p \times 1$ vector) denoted the values of p variables (other covariates) for subject i at time t_{ij}, $Z_i(t_{ij})$ can be thought of as the true values of the marker at time t_{ij}, $\varepsilon_i(t_{ij})$ are error terms assumed to be independent and normally distributed with mean zero and variance σ_ε^2, that is, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, the random intercept a_i assumed to be independent and normally distributed with mean μ_a and variance σ_a^2,

that is, $a_i \sim N(\mu_a, \sigma_a^2)$, b is the fixed slope, and the $1 \times p$ vector of unknown parameters of dimension p, $\beta = (\beta_1, \ldots, \beta_p)^T$ represents the effect of the p variables on the marker.

Longitudinal data is the regression model has the assumption that the errors are normally and independently distributed in the population with constant variance. But sometimes, one should expect some dependence between the response variables.

Covariates are often used as an alternative name for explanatory variables , but perhaps more specifically to refer to variables that are not of primary interest in an investigation but are measured because it is believed that they are likely to affect the response variable and consequently need to be included in analyses and model building. For example age, gender, eye color, heir color, and etc.

2.2.1 Exploring Longitudinal Data

Longitudinal data are by their very nature complex: there are typically a large number of observations on many subjects, and there are often covariates and missing values to further confuse the issue. Before wholesale analysis begins, a profitable first step is always to ***explore*** the data: simple procedures such as plotting or calculating means can provide valuable insights into the structure and unusual features of the data.

In plotting longitudinal data, we seek to address a number of goals:

- To characterize ***subject*** behavior, that is, to observe how the data vary across subjects at points in time.

- To characterize *time* behavior, that is, to observe how the response changes with time.

- To characterize *covariate* behavior, that is, to observe how the response varies with one or more covariates.

- To identify unusual observations in the data.

- To display all of the data points themselves, for example the primary method of plotting is the parallel plot (Weiss & Lazaro, 1992) all the data points are plotted with time on the horizontal axis and the response measure on the vertical axis, and observations on individual subjects are connected by straight lines. Every subject thus appears as a connected sequence of observations on the plot. This allows one to observe the individual behaviour of each subject across time, and also to observe the variability of subjects within each time point.

Achieving all or even many of these goals in one plot is a difficult task, the complex nature of longitudinal data can often result in a plot which is overly busy with points and lines making it difficult to get a feel for the overall features of the data. Despite this, a judiciously designed plot or combination of plots can often make sense of complex data.

When all of the subjects can be assumed to be a sample from some population, valuable insight into the average behavior of the population can be achieved by smoothing over the subjects and over time in some way. To achieve such a goal, a full account of the principles underlying kernel smoothing may be found in Hastie and Tibshirani (1990).

2.2.2 Multivariate Longitudinal Data

In medical studies, it is common to observe multivariate outcomes repeatedly. For example, in Community Behavioral Health (CBH) subjects take psychological tests and we use the resulting scores to measure each subject's overall cognitive performance. There are 5 test batteries which measure different kind of performance, such as general memory, delayed recall, attention, working memory, etc. (Zamrini et al. 2004, Rusinek et al. 2004). These test batteries are correlated with each other and therefore multivariate analysis for the repeated measurement is necessary.

Let $\mathbf{Y}_i(t_{ij}) = \left(\mathrm{Y}_{1i}(t_{ij}),\ \mathrm{Y}_{2i}(t_{ij}), \dots, \mathrm{Y}_{qi}(t_{ij})\right)^T$, be the observed multivariate response with q components at time $t_{ij},\ j = 1, \dots, n_i,$ are the observation on the i^{th} subject $i = 1, \dots, m,$ and q-dimensional response vector. The model we proposed is a multivariate generalization of the model in (2.1) given by:

$$\begin{aligned} \mathbf{Y}_i(t_{ij}) &= \mathbf{Z}_i(t_{ij}) + \boldsymbol{\varepsilon}_i(t_{ij}) \\ \mathbf{Z}_i(t_{ij}) &= \mathbf{A}_i(t_{ij})\beta + \mathbf{B}_i(t_{ij})\mathrm{b}_i \end{aligned} \qquad (2.2)$$

Where

$\mathbf{Z}_i(t_{ij}) = \left(\mathrm{Z}_{1i}(t_{ij}),\ \mathrm{Z}_{2i}(t_{ij}), \dots, \mathrm{Z}_{qi}(t_{ij})\right)^T$ are the true values of the multivariate response, q-dimensional response vector.

$\mathbf{A}_i(t_{ij})$ is a $(n_i \times p_1)$ "design matrix" of fixed effects; that characterizes the systematic part of the response, and β is a dimension $(p_1 \times 1)$ fixed effects vector.

$\mathbf{B}_i(t_{ij})$ is a $(n_i \times p_2)$ "design matrix" of random effects; that characterizes random variation in the response attributable to among-unit sources, and b_i is a dimension $(p_2 \times 1)$ random effects vector.

$\boldsymbol{\varepsilon}_i$ is a vector of measurement errors.

The assumptions of all variables and parameters in model (2.2) are the same as in (2.1) with multivariate instead of univariate. For more details see Salah, k. A. (2008).

The challenge of multivariate longitudinal analysis, especially with high-dimensional response vector, is to choose a reasonable approximation to the variance covariance matrix of $\boldsymbol{\varepsilon}_i$, which takes the form.

$$Var(\varepsilon_i) = \begin{pmatrix} Var(\varepsilon_{i1}) & Cov(\varepsilon_{i1}, \varepsilon_{i2}) & \cdots & Cov(\varepsilon_{i1}, \varepsilon_{in_i}) \\ Cov(\varepsilon_{i2}, \varepsilon_{i1}) & Var(\varepsilon_{i2}) & \cdots & Cov(\varepsilon_{i2}, \varepsilon_{in_i}) \\ \vdots & \vdots & \cdots & \vdots \\ Cov(\varepsilon_{in_i}, \varepsilon_{i1}) & Cov(\varepsilon_{in_i}, \varepsilon_{i2}) & \cdots & Var(\varepsilon_{in_i}) \end{pmatrix}$$

Of dimension $(n_i \times n_i)$ estimate of longitudinal parameters will be discussed in details in the next chapter.

2.3 Survival Data Analysis

Broadly speaking, survival analysis is a set of statistical methods for examining not only event *occurrence* but also the *timing* of events. These methods were developed for studying death, hence the name survival analysis, and have been used extensively for that purpose; however, they have been successfully applied to many different kinds of events, across a range of disciplines. Examples include manufacturing or engineering: how long

it takes widgets to fail; meteorology: when will the next hurricane be hit the North Carolina coast; social: what determines how long a marriage will last; financial: the timing of stock market drops, the list goes on. Sometimes other names are used to refer to this class of methods, such as event history analysis, or failure time analysis or transition analysis, but many of the basic techniques are the same as is the underlying idea, understanding the pattern of events in time and what factors are associated with when those events occur.

In order to start with survival analysis, a simple example were considered. A schematic depiction of simple survival data for six subjects is shown in Figure 2.1. In this figure, all subjects start their survival time at the same point, the study baseline. Further, we assume that each person can have the event only once. Three of the six patients (lines ending in solid circles #1,3, and 6) have an "event", and we can ascertain how long each of them was in the study prior to their event, their "survival time". As noted above the event may be death, but it can also be any other endpoint of interest, where we can measure the date of onset. In the Figure, there also 3 subjects (#2, 4 and 5) who do not have an event at least not while they are in the study. Subject #5 is the only one who completed the entire study without having an event. In contrast, two of the cases (open circles, #2 and 4) are lost to the study before having an event and before the study follow-up ends; they are said to be ***censored***. Actually, #5 is censored also, in this context, censoring simply means that at the end of a given individual's follow-up (whether that was early or at the end of the study), he/she had not had the event of interest. Different things can cause censoring, depending on the study design. It may be that these study participants decided they did not want to continue in the study, and so all we know is that at the time they left the

study, they had not yet had the event of interest. If our event of interest is not death, then it may be that censoring is caused by death again, we know that at the time we stopped following that person (i.e. when she died), she had not had the event of interest. And as noted above, people who have not had the event when all follow-up ends for all subjects, are also censored. We can view this as a special type of censoring, because everyone who has not had the event or already been censored for some other reason, is censored at this time.

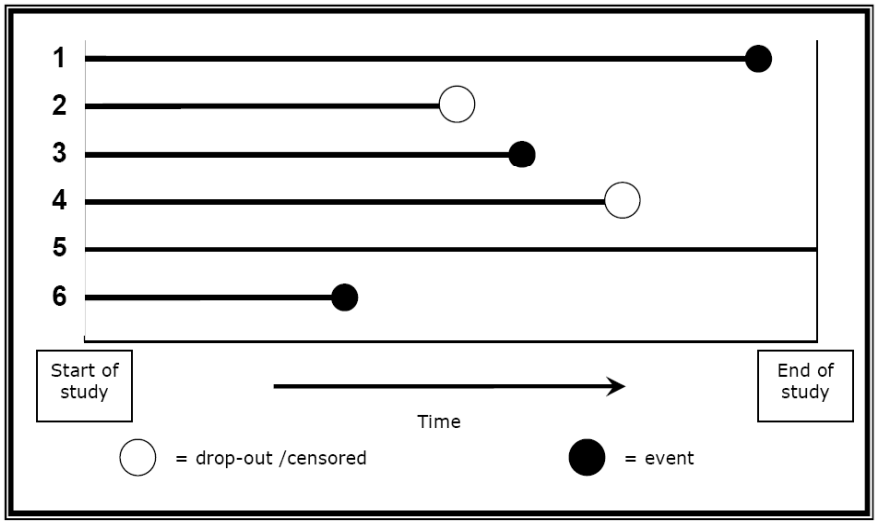


Figure 2.1: Hypothetical survival data for six patients.

As shown in the above example, three main issues were necessary for survival analysis:

- A beginning event. This needs to be well-defined.
- A scale or method for measuring time. This could be conventional methods such as minutes, days, months, years, ….
- An ending event. Again, this needs to be well-defined.

Missing data is endemic to longitudinal study settings; survival analysis is no exception. The various mechanisms for missing data in the survival context are usually grouped under the encompassing term, ***censoring***. Most generally, censoring occurs when the exact survival time is only known for a portion of the sample, with event times for the remaining subjects only known to occur in certain intervals. There are three categories of censoring: right, left, and interval censoring.

In survival analysis, we are interested in the random variable T, non negative, defined as the time from the starting event to the ending event, for example is that of human life. The date of birth is often the beginning event, calendar time is the usual time scale, and death is a common ending event.

In survival time modeling, let T_i denote the true observed failure time for the i^{th} subject $(i = 1, \dots, m)$, C_i denote the censoring time for the i^{th} subject, $V_i = min(T_i, C_i)$ which is taken as the minimum of the true event time T_i and the censoring time C_i. Furthermore, we define the event indicator as $\delta_i = I(T_i \leq C_i)$ where $I(\cdot)$ is the indicator function that takes the value 1 if the condition $T_i \leq C_i$ is satisfied, and 0 otherwise, that is

$$V_i = \begin{cases} T_i & \text{if } T_i \leq C_i \\ C_i & \text{if } T_i > C_i \end{cases}, \qquad \delta_i = \begin{cases} 1 & \text{if } T_i \leq C_i, \text{Time to Event} \\ 0 & \text{if } T_i > C_i, \text{Censored} \end{cases}$$

The probability of a subject's surviving beyond the time t is given by the survival function $S(t)$, in literature modeling of survival data usually take the form

$$S(t) = P(T > t) = 1 - F(t) = \int_{t}^{\infty} f(u)du \tag{2.3}$$

Where

$$F(t) = P(T \leq t) = \int_{0}^{t} f(u)du \tag{2.4}$$

or

$$f(t) = \frac{d}{dt}F(t) \tag{2.5}$$

where T is a continuous nonnegative random variable (defined on $(0, \infty)$) representing the survival time, $f(t)$ denotes the probability density function PDF of T, and $F(t)$ is the cumulative distribution function CDF of T represents the ending event occurs before time t.

$S(t) = P(T > t)$ is mean the ending event occurs past time t. This is the same as saying the event time for the random variable T occurs past t. This function $S(t)$ is non-increasing, a monotonic, nonnegative function, $\lim_{t\to 0} S(t) = 1$ and $\lim_{t\to \infty} S(t) = 0$.

The hazard function $h(t)$ is defined as the instantaneous probability the ending event occurs just past time t given that the event had not yet occurred at t.

$$h(t) = \lim_{\Delta t \to 0} \left[\frac{P[(t \leq T < t + \Delta t | T \geq t)]}{\Delta t} \right] \tag{2.6}$$

or , the ratio of the probability density function to the survival function $S(t)$

$$h(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} = -\frac{d}{dt} log\big(S(t)\big) \tag{2.7}$$

thus, $f(t)$ can be written as $f(t) = h(t)S(t)$.

Hence, from equation (2.7) the survival function can be expressed as

$$S(t) = exp\left(-\int_0^t h(u)du\right) \tag{2.8}$$

The cumulative hazard function $H(t)$ is the accumulated risk or hazard over time 0 up to time t. Defined as

$$H(t) = \int_0^t h(u)du$$

2.4 Censoring and Truncation of Survival Data

A survival time is censored when information on time to outcome event is not available for all study participants. Participant is said to be censored when information on time to event is not available due to loss to follow-up or non-occurrence of outcome event before the trial end. Broadly classifying three types of censoring are encountered:

a. Right censoring: a subject is right censored if it is known that the event of interest occurs some time after the recorded follow-up period. This is noted above in Figure 2.1.

b. Left censoring: a subject is left censored if it is known that the event of interest occurs some time before the recorded follow-up period. For example, you conduct a study investigating factors influencing days to first oestrus in dairy cattle. You start observing your population (for argument's sake) at 40 days after calving but find that several cows in the group have already had an oestrus event. These cows are said to be left censored at day 40. This is noted above in Figure 2.1.

c. Interval censoring: a subject is interval censored if it is known that the event of interest occurs between two times, but the exact time of failure is not known. In effect we say `I know that the event occurred between date A and date B: I know that the event occurred, but I don't know exactly when'. In an observational study of Enzootic Bovine Leucosis (EBL) seroconversion you sample a population of cows every six months. Cows that are negative on the first test and positive at the next are said to have seroconverted. These individuals are said to be interval censored with the first sampling date being the lower interval and the second sampling date the upper interval. This is noted above in Figure 2.1.

By contrast, truncated survival time data are those for which there is a systematic exclusion of survival times from one's sample, and the sample selection effect depends on survival time itself. We may distinguish two types of truncation:

a. Right truncation: a subject is right truncated if it leaves the population at risk some stage after the study start (and we know that there is no way the event of interest could have occurred after this date). For example, in a study investigating the date of first foot-and-mouth disease diagnosis on a group of farms, those

farms that are preemptively culled as a result of control measures are right truncated on the date of culling.

b. Left truncation: a subject is left truncated if it enters the population at risk some stage after the start of the follow-up period. For example, in a study investigating the date of first mad cow disease (BSE) diagnosis on a group of farms, those farms that are established after the start of the study are said to be left truncated (the implication here is that there is no way the farm can experience the event of interest before the truncation date).

If no censoring the empirical estimate of the survival function is the proportion of individuals with event times greater than t. With censoring if there are censored observations, then is not a good estimate of the true $S(t)$, so other non-parametric methods must be used to account for censoring.

In the following subsections, some parametric and non-parametric survival models were illustrated.

2.5 Non Parametric Survival Models

2.5.1 Kaplan-Meier Estimator (Product-Limit Estimator)

The Kaplan–Meier estimate (Kaplan, E. L. and P. Meier (1958)) of the survival function is an empirical or non-parametric method of estimating $S(t)$ from non- or right-censored data. It is extremely popular as it requires only very weak assumptions and yet utilizes the information content of both fully observed and right-censored data. It comes as standard

in most statistical packages (such as R) and can also be calculated by hand. Specifically the estimator is given by

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{t_j \le t} \left[1 - \frac{d_j}{r_j}\right] & \text{if } t > t_1 \end{cases}$$

Where $\hat{S}(t)$ is the estimated survival function at time $t, t_1 \le t_2 \le \cdots \le t_{n_i}$ are the ordered survival time, $j = 1, \ldots, n_i$, t_1 denotes the first observed failure time, r_j is the number of individual at risk at time t_j and d_j is the number of individuals who experience the event of interest at time t_j. (individuals censored at time t_j are included in r_j) the resulting estimates form a step function that can be plotted to given a graphical display of survival experience.

As illustrative example, consider the times of remission (weeks) for 21 leukemia patients receiving control treatment: 6+, 6, 6, 6, 7, 9+, 10+, 10, 11+, 13, 16, 17+, 19+, 20+, 22, 23, 25+, 32+, 32+, 34+, 35+. Note: times with + are right censored. It is useful to start the estimation of *S*(*t*) by writing the data in ascending order, including the censored times in the ranking but ensuring that they are distinguishable from times that are not censored.

Let $t_1 \le t_2 \le \cdots \le t_{16}$ is the set of 16 distinct death times observed in the sample and $t_0 = 0$, d_j is the number of deaths at t_j, r_j is the number of individuals "at risk" right before the j^{th} death time (everyone dead or censored at or after that time), and c_j is the

number of censored observations between the j^{th} and $(j+1)^{st}$ death times. Censorings tied at t_j are included in c_j.

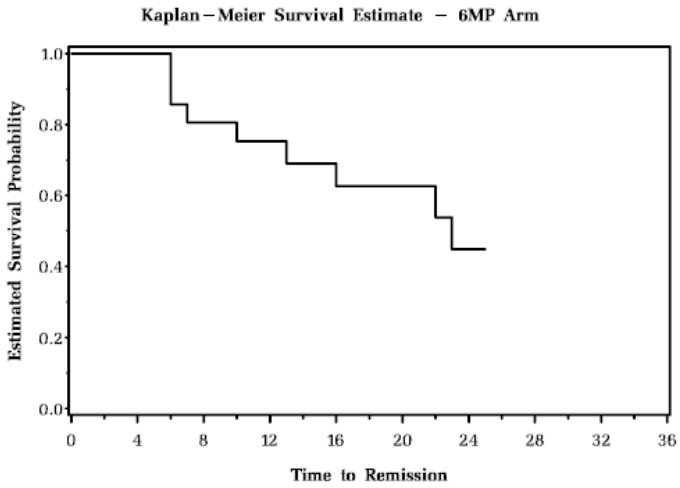
Two useful formulas are: (1) $r_j = r_{j-1} - d_{j-1} - c_{j-1}$, (2) $r_j = \sum_{l \geq j}(c_l + d_l)$

Hence, the Kaplan-Meier estimator of the survivorship function (or survival probability) summarized in the following table

Table 2.1 Calculating the KM Estimator

j	t_j	d_j	c_j	r_j	$\lambda_j = 1 - (d_j/r_j)$	$S(t_j) = \prod \lambda_j$
0	0	0	0	21	$1 - (0/21) = 1$	1
1	6	3	1	21	$1 - (3/21) = 0.857$	$1 \times 0.857 = 0.857$
2	7	1	0	17	$1 - (1/17) = 0.94$	$0.857 \times 0.94 = 0.8067$
3	9	0	1	16	$1 - (0/16) = 1$	$0.8067 \times 1 = 0.8067$
4	10	1	1	15	$1 - (1/15) = 0.933$	$0.8067 \times 0.933 = 0.7529$
5	11	0	1	13	$1 - (0/13) = 1$	0.7529
6	13	1	0	12	$1 - (1/12) = 0.917$	0.6902
7	16	1	0	11	$1 - (1/11) = 0.91$	0.6275
8	17	0	1	10	$1 - (0/10) = 1$	0.6275
9	19	0	1	9	$1 - (0/9) = 1$	0.6275
10	20	0	1	8	$1 - (0/8) = 1$	0.6275
11	22	1	0	7	$1 - (1/7) = 0.857$	0.5378
12	23	1	0	6	$1 - (1/7) = 0.833$	0.4482
13	25	0	1	5	1	0.4482
14	32	0	2	4	1	0.4482
15	34	0	1	2	1	0.4482
16	35	0	1	1	1	0.4482

Also, the KM plot for treated leukemia patients is given by



2.5.2 Turnbull Estimator

An estimator of the survival function is available for interval-censored data. In 1976 Richard Turnbull formulated an expectation maximization algorithm (EM) to estimated the nonparametric maximum likelihood estimator (NPMLE) for interval-censored data, but this estimator could accommodate truncated data too. We use the terms "Turnbull Estimator" and "NPMLE" interchangeably. The NPMLE for interval-censored data is based on m independent, arbitrary interval censored observations $(x_1, x_2, \ldots, x_m)$. We assume that x_i is interval censored, so that x_i is only known to lie in an interval $[\mathrm{L}_i, \mathrm{R}_i]$. The derivation of the Turnbull estimator will be introduced here, for example, suppose we have 5 failures which occur in a study. The survival times for the 5 patients in this hypothetical study are interval censored. The following data set shows the m = 5 interval censored data points, $[\mathrm{L}_1, \mathrm{R}_1] = [1, 2]$, $[\mathrm{L}_2, \mathrm{R}_2] = [2, 5]$, $[\mathrm{L}_3, \mathrm{R}_3] = [4, 7]$, $[\mathrm{L}_4, \mathrm{R}_4] = [3, 8]$, $[\mathrm{L}_5, \mathrm{R}_5] = [7, 9]$.

Before the NPMLE can be estimated using Turnbull's algorithm, equivalence classes must be defined to determine at what time points the survival function takes jumps. Equivalence classes are defined by each "L" that is immediately followed by "R" once the endpoints are ordered. In table 2.2 find the equivalence classes, we need to consider all of the $[L_i, R_i]$ intervals, for $i = 1, 2, \ldots, m$, and then order the $2m$ endpoints from smallest to largest.

Table 2.2 Construction of equivalence classes for an interval-censored data set

Initial Endpoints	1	2	2	5	4	7	3	8	7	9
Corresponding	L_1	R_1	L_2	R_2	L_3	R_3	L_4	R_4	L_5	R_5
Ordered Endpoints	1	2	2	3	4	5	7	7	8	9
Corresponding	L_1	L_2	R_1	L_4	L_3	R_2	L_5	R_3	R_4	R_5
Labels	L	L	R	L	L	R	L	R	R	R

Table 2.2 shows how we obtain the equivalence classes for the hypothetical data set. The first two lines in the table are the initial data. The next two lines represent the ordered endpoints and their corresponding labels. The fifth line denotes only whether the corresponding point is a left L or right R endpoint. Therefore, we have $n_i = 3$ equivalence classes which are $[2, 2], [4, 5], [7, 7]$. Only within these equivalence classes can the survival function have jumps. The *Turnbull estimator* of the CDF is given by

$$\hat{F}(x) = \begin{cases} 0 & \text{if } x < q_1 \\ \hat{s}_1 + \hat{s}_2 + \cdots + \hat{s}_j & \text{if } p_j < x < q_{j+1} \quad (1 \le j \le n_i - 1) \\ 1 & \text{if } x > p_{n_i} \end{cases}$$

where q_1 is the lower bound of the first equivalence class and p_{n_i} is the upper bound of the last equivalence class. The interval $[q_j, p_j]$ represents the j^{th} equivalence class.

Therefore, $\hat{F}$ is undefined for $x \in [\mathrm{q}_j, \mathrm{p}_j]$, for $1 \leq j \leq n_i$, which means that $\hat{F}$ is defined only in between the equivalence classes.

Before maximizing the likelihood of F, the CDF, it is necessary to be familiar with the α matrix. This is an $m \times n_i$ matrix of indicator variables. As stated earlier, each $[\mathrm{L}_i, \mathrm{R}_i]$ interval represents the censoring interval which contains the failure time for individual i and $[\mathrm{q}_j, \mathrm{p}_j]$ represents the j^{th} equivalence class. The $(ij)^{th}$ element of the α is defined as

$$\alpha_{ij} = \begin{cases} 1 & \text{if } [\mathrm{q}_j, \mathrm{p}_j] \subseteq [\mathrm{L}_i, \mathrm{R}_i] \\ 0 & \text{otherwise} \end{cases}$$

The maximum likelihood estimator for the CDF is represented as:

$$L(F) = \prod_{i=1}^{m} \left[F(\mathrm{R}_{ij} +) - F(\mathrm{L}_{ij} -) \right]$$

Once the CDF of the survival distribution is obtained, integration techniques can be used to calculate the PDF of the survival times, and the survival function in (2.5), $S(x) = 1 - F(x)$. As an alternative to maximizing the likelihood for the CDF, we can also maximize the equivalent likelihood given as

$$L(F) = \prod_{i=1}^{m} \left[\frac{\sum_{j=1}^{n_i} \alpha_{ij} \mathrm{s}_j}{\sum_{j=1}^{n_i} \mathrm{s}_j} \right]$$

where ij is the $(ij)^{th}$ element of the matrix defined earlier, and s_j represents the jump amount within the $(ij)^{th}$ equivalence class.

The process of maximizing the equivalent likelihood is the procedure used in Turnbull's method of finding the NPMLE. This technique involves finding the expected value of I_{ij}, a matrix that has the same dimensions as ij, and is defined as:

$$\mathrm{I}_{ij} = \begin{cases} 1 & \text{if } x_i \subseteq [\mathrm{L_i}, \mathrm{R_i}] \\ 0 & \text{otherwise} \end{cases}$$

The expected value of I_{ij} is first calculated under an initial s matrix, which is an $n_i \times 1$ matrix with elements equal to $\frac{1}{n_i}$. $\mathrm{E_s}[\mathrm{I}_{ij}]$ is the expected value of I_{ij} as a function of s, and is also denoted as $\mu_{ij}(\mathrm{s})$. When we treat μ_{ij} as an observed quantity, we can estimate the proportion of observations in the j^{th} equivalence class, π_{ij}, as the following:

$$\pi_j(\mathrm{s}) = \frac{1}{m}\sum_{i=1}^{m} \mu_{ij}\,(\mathrm{s})$$

The Turnbull method is an example of an Expectation-Maximization (EM) algorithm since the final s matrix of jump amounts is found by alternating between the calculation of $\mathrm{E_s}[\mathrm{I}_{ij}]$ and obtaining an s matrix which maximizes π_{ij}. The algorithm stops when the difference between successive values for s is less than a given tolerance, (see Zhao, 2008).

2.6 Semi-Parametric Model (Cox Proportional Hazards Model)

This model, proposed by Cox (1972), is perhaps the most-often cited article in survival analysis. The distinguishing feature of Cox's proportional hazard model, sometimes simply referred to as the. Cox model, is its demonstration that one could estimate the relationship between the hazard rate and explanatory variables without having to make

any assumptions about the shape of the baseline hazard function. Hence the Cox model is sometimes referred to as a semi-parametric model. The result derives from innovative use of the proportional hazard assumption together with several other insights and assumptions, and a partial likelihood (PL) method (a product of conditional likelihoods , used in certain situations for estimation and hypothesis testing) of estimation rather than maximum likelihood. Here follows an intuitive demonstration of how the model works, based on the explanation given by Allison (1984 and 1995). Cox proportional hazard regression model is a broadly applicable and the most widely used method of survival analysis Cox and Oakes (1984).The model involved is

$$\log\big(h_i(t)\big) = \ln \vartheta(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik},$$

where $x_{i1}, x_{i2}, \dots, x_{ik}$ are the explanatory variables of interest and covariates, $k = 1, \dots, p$, of p-dimension and $h_i(t)$ the hazard function. $ln\,\vartheta(t)$ is baseline hazard function, is an arbitrary function of time for any two individuals at any point in time the ratio of hazard function is a constant, since $\log\big(h_i(t)\big) = \ln \vartheta(t)$ when all of them x's are zero. Because the baseline hazard function does not have to be specified explicitly, the procedure is essentially a distribution free method. Estimates of the parameters in the model, i.e. $\beta_1, \beta_2, \dots, \beta_k$, are usually obtained by maximum likelihood estimation, and depend only on the order in which events occur, not on the exact time of their occurrence. The Cox proportional hazard model, instead, leaves the baseline hazard function unspecified.

The hazard model can be written as

$$h_i(t, x) = \vartheta(t)\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})$$

Let $\vartheta(t) = h_0(t)$, then we got

$$h_i(t,x) = h_0(t)\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}) \tag{2.9}$$

The hazard ratio for two observation i and j, where $i \neq j$ is

$$\frac{h_i(t,x)}{h_j(t,x)} = \frac{h_0(t)exp(\omega_i)}{h_0(t)exp(\omega_j)} = \frac{exp(\omega_i)}{exp(\omega_j)} = exp(\omega_i - \omega_j) \tag{2.10}$$

where $\omega_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$ and $\omega_j = \beta_1 x_{j1} + \beta_2 x_{j2} + \cdots + \beta_k x_{jk}$ are independent of time t, Cox model is proportional hazard model.

Remarkably, even though the baseline hazard is unspecified, the Cox model can still be estimated by method of partial likelihood Cox (1972). Although the resulting estimates are not as efficient as maximum-likelihood estimates for a correctly specified parametric hazard regression model, not having to make arbitrary.

2.7 Parametric Survival Models

Parametric models play an important role in survival analysis, since they offer straightforward modeling and analysis techniques. The goal is to estimate the unknown parameters to obtain inferences about the scientific problems at hand. Commonly used parametric models include the log normal, negative binomial, gamma, exponential and Weibull, which are attractive in their simplicity and the easy interpretability of their components.

2.7.1 Exponential Model

In the exponential model, the hazard rate is characterized as:

$$h(t, X) = \lambda \tag{2.11}$$

This implies that the conditional "probability" of an event is constant over time (and that events occur according to a Poisson process). In other words, the risk of an event occurring is flat with respect to time.

Modelling the dependency of the hazard rate on covariates entails constructing a model that ensures a non-negative hazard rate (or non-negative expected duration time). One way to do this is simply to exponentiate the covariates such that:

$$h(t, X) = \lambda_i = e^{X_i \beta} \tag{2.12}$$

Given the way that the hazard rate is specified in the exponential model, the cumulative hazard can be written as:

$$H(t, X) = \lambda t \tag{2.13}$$

Recall from the earlier notes that $H(t, X) = -\ln(S(t, X))$. As a result, we have:

$$\begin{aligned} S(t, X) &= e^{-H(t,X)} \\ &= e^{-\lambda t} \end{aligned} \tag{2.14}$$

This in turn means that the density is:

$$f(t, X) = h(t, X)S(t, X) = \lambda e^{-\lambda t} \tag{2.15}$$

This density means that the duration time T has an exponential distribution with mean $\frac{1}{\lambda} = E(t_i)$. In other words, the expected duration in an exponential model is:

$$E(t_i) = \frac{1}{\lambda} = \frac{1}{e^{X_i\beta}} = e^{-X_i\beta} \tag{2.16}$$

Having defined $h(t, X)$, $f(t, X)$, and $S(t, X)$, it is easy to construct the sample likelihood for the exponential model:

$$\begin{aligned} L(t, X) &= \prod_{i=1}^{m} [f(t, X)]^{\delta_i} \, [S(t, X)]^{1-\delta_i} \\ &= \prod_{i=1}^{m} \left[\lambda e^{-\lambda t}\right]^{\delta_i} \, \left[e^{-\lambda t}\right]^{1-\delta_i} \end{aligned} \tag{2.17}$$

2.7.2 Weibull Model

The Weibull model the definition density function is

$$f(t) = \frac{\alpha t^{\alpha-1}}{\beta^{\alpha}} exp\left[-\left(\frac{t}{\beta}\right)^{\alpha}\right] \quad 0 \leq t < \infty, \beta > 0 \quad \alpha > 0 \tag{2.18}$$

The Weibull model is more general and flexible than the exponential model and allows for hazard rates that are non-constant but monotonic. It is a two-parameter model (λ , and α), where λ is the location parameter and α is the shape parameter because it determines whether the hazard is increasing, decreasing, or constant over time. The distribution occurs in the analysis of survival data and has the important property that the corresponding hazard function can be made to increase with time ($\alpha > 1$), decrease with

time ($\alpha < 1$), or remain constant, by a suitable choice of parameter values. When $\alpha = 1$ the distribution reduces to the *exponential distribution*. So the Weibull distribution contain the exponential as special case.

Let $\lambda = log\left(\frac{1}{\beta^{\alpha}}\right)$, the distribution function can be reformatted to

$$f(t, X|\alpha, \lambda) = \alpha t^{\alpha-1} exp\big(\lambda - e^{\lambda} t^{\alpha}\big) \tag{2.19}$$

In the Weibull model, the hazard rate is characterized as:

$$h(t, X) = \alpha\lambda(\lambda t)^{\alpha-1} \tag{2.20}$$

The survival function for the Weibull is

$$S(t, X|\alpha, \gamma) = exp\big(-e^{\lambda} t^{\alpha}\big) \tag{2.21}$$

Having defined $h(t)$, $f(t)$, and $S(t)$, it is easy to construct the likelihood for the Weibull model:

$$\begin{aligned} L(t, X) &= \prod_{i=1}^{m} [f(t, X)]^{\delta_i} \, [S(t, X)]^{1-\delta_i} \\ &= \prod_{i=1}^{m} \big[\alpha t^{\alpha-1} exp\big(\lambda - e^{\lambda} t^{\alpha}\big)\big]^{\delta_i} \, \big[exp\big(-e^{\lambda} t^{\alpha}\big)\big]^{1-\delta_i} \end{aligned} \tag{2.22}$$

In this study Assume the survival time for the i^{th} subject follows a Weibull distribution

$t_i \sim Weibull\big(\gamma, \mu_i(t)\big)$, where $\mu_i(t) = exp\big(\gamma Z_i(t) + \omega P_i(t)\big)$ and $\gamma > 0$, then the hazards function in (2.20) becomes

$$h_i\big(t\big|Z_i(t), P_i(t)\big) = h_0(t) exp\big(\gamma Z_i(t) + \omega P_i(t)\big) \tag{2.23}$$

Where γ is the regression coefficient represents the effect of the marker on the disease risk, also, ω is the regression coefficient represents the effect of the other potential covariates, $P_i(t)$ denoted q-dimension other potential covariates (time fixed or dependent), which may include some or all of X_i covariates and $h_0(t)$ is baseline hazard.

We assume $h_0(t) = h_{0k}$, over some arbitrary partitioning $t_0^h = 0, t_1^h, \dots, t_K^h$ of the time scale into K intervals, $t_{k-1}^h < t \leq t_k^h$, t_k^h are not necessarily related to this times of the marker measurements for all subjects.

2.8 Joint Modeling

Along with time-to-event data, many studies also collect longitude. Instead of analyzing them separately, joint modeling of longitudinal and survival data has attracted great attention in the literature, two approaches have been proposed to carry out the joint modeling: a two-stage model (Raboud et al. 1993, Tsiatis et al. 1995, Bycott and Taylor 1998, Dafni and Tsiatis 1998) and a likelihood based joint model (DeGruttola and Tu 1994, Faucett and Thomas 1996,Wulfsohn and Tsiatis 1997, Xu and Zeger 2001, Wang and Taylor 2001, Salah, k. A. (2008)). It has been shown that by sharing information between the longitudinal model and the survival model, the likelihood based approach has the advantage of smaller bias and more efficiency in the parameter estimates. While applying the joint model to a real data set, we often observe nonlinear trends in the longitudinal trajectories and also a large variation in their shapes. Sometime we might want to model the survival time and the longitudinal measurement in the mean time, i.e., we might want to estimate the $[T|X]$ and $[Y|X]$ at the same time, where Y is a

longitudinal measurement which is a certain marker of the study. The final goal of joint modeling is to model $[T, Y|X]$.

Xu and Zeger (2001) use a latent variable model to describe the relationship between time-to-event data, longitudinal response, and covariates, in which covariates could only affect the response through its influence on an assumed latent process. The model below shows the relationship between event time T, biomarker response Y, and treatment indicator variable X, p-dimension, by assuming an underlying latent process ω corresponding to Y.

The recent approaches of joint modeling by Xu and Zeger (2001) proposed latent variable models where Y and T are modeled simutenously comparing to the two-stage model by Tsiatis, 1997.

$$Y = \omega + \varepsilon$$

And $h(t)$ for the survival time

$$h(t) = h_0(t) exp\big(\gamma(X\beta_1 + \omega\beta_2)\big)$$

Where β_1 and β_2 unknown parameter, γ is the regression coefficient, and the likelihood

$$[T, Y|X] = \int [T, Y|\omega, X]\,[\omega|X] d\omega = \int [T|\omega, X][Y|\omega]\,[\omega|X] d\omega$$

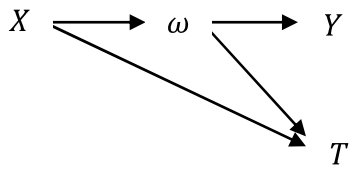


Figure 2.2.a The relationship in joint model between T, Y, X and ω

The model is established on the basis of three major assumptions

a. T and Y are conditionally independent given a latent process ω.
b. X can affect T either through ω or directly.
c. X can only affect Y through its influence on ω.

To be more specific, $Y_i(t)$, the observed value of the process at time t is modeled as an independent observation from a generalized linear model (GLM) with linear predictor $\omega_i(t)$. where $\omega_i(t)$ is generally assumed to follow a Gaussian stochastic process. And the model allows different forms of conditional hazard to be specified for $[T|\omega, X]$. An application of this model is when the auxiliary variable Y is an imperfect surrogate end point for T.

Another approach to model both survival time and longitudinal marker through latent variables was proposed by DeGruttola and Tu (1994). The model is established on the basis of three major assumptions

a. T and Y are conditionally independent given a latent process ω and covariate X.
b. X can affect Y and T.
c. Random effect ω can affect Y and T.

The model can be specified as

$$Y = X_1\beta_1 + Z_1\omega + \varepsilon_1$$

and

$$T = X_2\beta_2 + Z_2\omega + \varepsilon_2$$

Where $\beta_i, i = 1,2$ are a p-vector of population average regression coefficients, called fixed effects, and where ω is a q-vector of subject-specific regression coefficients, called random effect. The matrices X_i and $Z_i, i = 1,2$ are $(n_i \times p)$ and $(n_i \times q)$ matrices of known covariates. Note that p and q are the numbers of fixed and subject-specific regression parameters in the model, respectively. The residual components ε_i, $i = 1,2$ are assumed to be independent $N(0, \sigma_{\varepsilon_i})$, where σ_{ε_i} depends on i only through its dimension n_i. Hence the likelihood can be written as

$$[T, Y|X] = \int [T|\omega, X][Y|\omega, X]\, d\omega$$

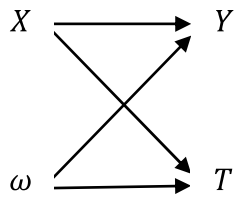


Figure 2.2.b The relationship joint model between T, Y, X and ω

Wang and Taylor (2001): include the longitudinal marker as a time-dependent covariate in the (proportional hazards) survival model.

In this study our joint model composed of two sub models: a covariate tracking model (longitudinal data) and a disease risk model.

For the longitudinal data we incorporate the model in (2.1).

$$\begin{aligned} Y_i(t_{ij}) &= Z_i(t_{ij}) + \varepsilon_i(t_{ij}) \\ Z_i(t_{ij}) &= a_i + bt_{ij} + \beta X_i(t_{ij}) \end{aligned} \tag{2.24}$$

By incorporating model (2.24) into hazard model in (2.23) we got the joint model of the form:

$$\begin{aligned} h_i\left(t\middle|Z_i(t), P_i(t)\right) &= h_0(t)exp\left(\gamma Z_i(t) + \omega P_i(t)\right) \\ &= h_0(t)exp\left(\gamma\left(a_i + bt_{ij} + \beta X_i(t_{ij})\right) + \omega P_i(t)\right) \end{aligned} \tag{2.25}$$

which describes the relationship between survival time and longitudinal data, it assumes that, given the true marker value and other covariates.

The joint model in (2.25) is established on the basis of four major assumptions

a. T and Y_i are conditionally independent given Ω, P_i and covariate X_i.
b. X_i can affect Y_i and T.
c. P_i can affect Y_i and T.
d. Unknown parameter Ω can affect Y_i and T.

The likelihood for the joint model in (2.25), $[Y_i, T\backslash X_i, P_i, \Omega]$ as

$$[Y_i, T\backslash X_i, P_i, \Omega] = \int [Y_i\backslash X_i, P_i, \Omega][T\backslash X_i, P_i, \Omega]dt$$

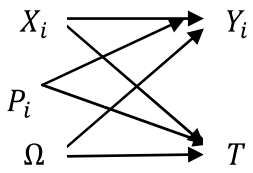


Figure 2.3 The relationship joint model between T, Y_i, X_i, Ω and P_i

2.9 Separate Modeling

From the joint model in (2.25)

$$h_i\left(t\middle|Z_i(t), P_i(t)\right) \quad = h_0(t)exp\left(\gamma\left(a_i + bt_{ij} + \beta X_i\left(t_{ij}\right)\right) + \omega P_i(t)\right)$$

Let $\gamma = 0$ we get the separate model (ignoring the association between the two sub-model).

$$h_i\left(t\middle|Z_i(t), P_i(t)\right) \quad = h_0(t)exp\left(\omega P_i(t)\right) \tag{2.26}$$

Clearly those markers (longitudinal data) and decline time (survival time) are not independent. There exist many methods for analyzing such data separately as introduced previously. However, joint modeling can provide more efficient estimates and reduce bias in many situations (see Xu and Zeger, 2001, and DeGruttola and Tu, 1994).

a. By using the separate models, when the covariates are measured with error, the estimate of the regression coefficient will be attenuated, that is, biased toward 0. Joint models can reduce this bias.
b. When studying the group effect on event time, conditioning on a time-varying covariates can mask the group effect. Joint models can overcome this problem.

Alternatively we might want to study the group effect on both the longitudinal data and the event time.

c. When censoring is informative about the event of interest, traditional statistical methods can give biased estimates of the event time distribution and group effect. Joint models may reduce this bias.

Chapter Three
Estimation of Parameters

3.1 Introduction

In chapter 2, we have a joint model of longitudinal data and survival data. It contains on the unknown parameters we need to estimations, wherefore we consider a Bayesian approach which is motivated by the complexity of the model. Posterior and prior specification needs to accommodate parameter constraints due to the non-negativity of the survival function.

The two well-known approaches for parameter estimation are maximum likelihood (ML) and Bayesian estimation. There exist a number of philosophical differences between the two approaches, including whether statistics are treated as the expected outcome from a large number of independent trials, usually associated with ML or a measure of subjective belief as in the Bayesian framework.

The fundamental idea behind maximum likelihood estimation is that a good choice for the estimate of a parameter of interest is the value of the parameter that makes the observed data most likely to have occurred. To do this, we need to establish some sort of function that gives us the probability for the data, and we need to find the value of the parameter that maximize this probability. This function is called the “likelihood function” in classical statistics, and it is essentially the product of sampling densities probability distributions for each observation in the sample.

Bayesian inference is the process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observation Gelman et al.,(1995). Bayes' Theorem for probability distributions is often stated as

$$\begin{aligned} Posterior \quad &\propto \ Likelihood \times Prior \\ [c\backslash y] \quad &\propto \ [y\backslash c][c] \end{aligned} \tag{3.1}$$

Where the symbol $\propto$ " means is proportional to."

Traditional maximum likelihood approach delivers only point estimates and associated asymptotic standard error estimates for the model parameters. This motivates the use of Bayesian analysis. As the development of computing power and improved scope for estimation via iterative sampling methods, Bayesian analysis of data in health, social and physical sciences has been greatly facilitated in the last decade. The new estimation methods Markov Chain Monte Carlo (MCMC) may be used to augment the data and this provides an analogue to the classical expectation maximization (EM) method, (see Gelman et al. 1995, 199).

Comparing to maximum likelihood, Bayesian estimations are more natural in parameter interpretation and easier to get true parameter densities. In the traditional maximum likelihood methods, variance estimates requires asymptotic assumptions which are sometimes not available. However in Bayesian framework, variance estimates and other posterior summary are produced by Gibbs sampler in the mean time samples of the posterior distribution are obtained, (see Chen, Ibrahim, and Sinha, 2004). However Bayesian inference has several other advantages over maximum likelihood methods in

model building and data analysis flexibilities. In traditional methods, there is no unified methodology for comparing non-nested models usually require asymptotic assumptions. While for Bayesian inference, model comparison and selections are easily completed via Gibbs sampler. Another area that Bayesian paradigm is superior to tradition likelihood method is its availability and flexibility of dealing missing values.

Also, unlike maximum likelihood we can consider with high-dimensional variables it is not easy to find.

3.2 The Likelihood of Some Distributions

Let $X = (x_1, \dots, x_m)$ be the random variables, $\hat{\mu}$ is mean, and $\hat{\sigma}^2$ is variance. If $X \sim N(\mu, \sigma^2)$, then the likelihood of the Normal distribution

$$\begin{aligned} L(\mu, \sigma^2 | x_i) &= \prod_{i=1}^{m} f(x_i | \mu, \sigma^2) \\ &= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(\frac{-(x_i - \mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{m/2}} exp\left(\frac{-\sum_{i=1}^{m}(x_i - \mu)^2}{2\sigma^2}\right) \end{aligned} \quad (3.2)$$

Where $\hat{\mu} = \frac{\sum_{i=1}^{m} x_i}{m}$ and $\hat{\sigma}^2 = \frac{\sum_{i=1}^{m}(x-\mu)^2}{m}$

If $X \sim G(\alpha, \beta)$, α is shape parameter and β scale parameter. Then the likelihood of the Gamma distribution

$$
\begin{aligned}
L(\alpha,\beta|x_i) &= \prod_{i=1}^{m} f(x_i|\alpha,\beta) \\
&= \prod_{i=1}^{m} \frac{\beta^{\alpha}}{\Gamma(\alpha)} {x_i}^{\alpha-1} exp\left(\frac{-\beta}{x_i}\right) \\
&= \left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}\right)^{m} \sum_{i=1}^{m} {x_i}^{\alpha-1}\, exp\left(-\sum_{i=1}^{m} \frac{\beta}{x_i}\right)
\end{aligned} \tag{3.3}
$$

Where $\hat{\mu} = \frac{\alpha}{\beta}$ and $\hat{\sigma}^2 = \frac{\alpha}{\beta^2}$

If $X \sim IG(\alpha,\beta)$. Then the likelihood of the Inverse Gamma distribution

$$
\begin{aligned}
L(\alpha,\beta|x_i) &= \prod_{i=1}^{m} f(x_i|\alpha,\beta) \\
&= \prod_{i=1}^{m} \frac{\beta^{a}}{\Gamma(\alpha)} {x_i}^{-(\alpha+1)} exp\left(\frac{-\beta}{x_i}\right) \\
&= \left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}\right)^{m} \sum_{i=1}^{m} {x_i}^{-(\alpha+1)}\, exp\left(-\sum_{i=1}^{m} \frac{\beta}{x_i}\right)
\end{aligned} \tag{3.4}
$$

Where $\hat{\mu} = \frac{\beta}{\alpha-1}$ and $\hat{\sigma}^2 = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$

3.3 Bayesian Estimation

Our Bayesian method involves a combination of direct sampling from the full conditional distribution, Bayesian model includes a suitable prior distribution hat summarizes information about parameter known or assumed at given time point, prior to obtaining further information from empirical data and posterior distribution that summarizes

information about a random variable or parameter after having obtained new information from empirical data.

The relationship can be written as

$$\begin{aligned} Posterior &\propto Likelihood \times Prior \\ [c \backslash y] &\propto [y \backslash c][c] \end{aligned}$$

Where $[c \backslash y]$ is posterior distribution, $[c]$ is prior distribution and $[y \backslash c]$ is the likelihood of the observed data y given parameter c.

Bayesian inference is then based on Monte Carlo samples drawn from the posterior distribution using an Markova Chain Monte Carlo (MCMC) algorithm such as the Gibbs sampler.

The Gibbs sampler, Geman and Geman (1984) is a special case of Metropolis-Hastings sampling. Metropolis et al. (1953), Hastings (1970), and the resulting Metropolis Hastings (M-H) algorithm, suppose that $\Omega = (\Omega_0, \Omega_1, \dots, \Omega_m)$ is $(m \geq 2)$ dimensional the complete conditional distributions.

$$\Omega_k \sim P\left(\Omega_k \middle| y, \Omega_1, \dots, \Omega_{k-1}, \Omega_{k+1}, \dots, \Omega_p\right)$$

Conditional distribution are much easier to simulate and usually have simple forms of prior distributions.

An alternative, and more general, updating scheme is as a form of generalized rejection sampling, where values are drawn from arbitrary (yet sensibly chosen) distributions and 'corrected' so that, asymptotically, they behave as random observations from the target

distribution. This is the motivation for methods such as the Metropolis-Hastings (M-H) updating scheme.

3.3.1 Markov Chain Monte Carlo method (MCMC)

Markov Chain Monte Carlo (MCMC) is a Monte Carlo method based on sampling from Markov Chain processes.

Monte Carlo integration is the original Monte Carlo approach was initially used to generate random number to compute integrals.

$$I(y) = E\big(p(x)\big) = \int_a^b f(x)p(x)dx,$$

Where $p(x)$ is a function of x , $f(x)$ is probability density function of x, defined over the interval (a, b), y is observed data, and x is random variable.

If we draw large number $x_1, \dots, x_m \sim f(x)$ then we can estimate Monte Carlo integration.

$$E\big(p(x)\big) \approx \frac{1}{m}\sum_{i=1}^{m} p(y|x_i)$$

Monte Carlo integration can be used to marginal posterior distribution for a Bayesian analysis.

Estimation of $I(y)$ is $\hat{I}(y)$, where

$$\hat{I}(y) = \sum_{i=1}^{m} p(y|x_i)$$

The estimated Monte Carlo standard error is given by

$$SE\left(\hat{I}(y)\right) = \sqrt{\frac{1}{m}\left(\frac{1}{m-1}\sum_{i=1}^{m}\left(p(y|x_i) - \hat{I}(y)\right)^2\right)}$$

Markov Chains is a stochastic process, $\{X_t\}$, $t = 0,1,2, \ldots$ where X_t takes values in the finite set $S = \{1,2, \ldots, M\}$, and is such that

$$P(next\ location|current\ and\ previous\ locations) = P(next\ location|current\ location)$$
$$P(X_m = i_m|X_0 = i_0, \ldots, X_{m-1} = i_{m-1}) = P(X_m = i_m|X_{m-1} = i_{m-1})$$

3.3.2 The Metropolis-Hasting Algorithm (M-H)

One problem with applying Monte Carlo integration is in obtaining samples from some complex probability distribution $P(x)$. Suppose our goal is to draw samples from some distribution $P(\theta)$ where $P(\theta) = \frac{f(\theta)}{K}$, where the normalizing constant K may not be known, and very difficult to compute. The Metropolis algorithm generates a sequence of draws from this distribution as follows:

1. Start with any initial value θ_0 satisfying $f(\theta_0) > 0$.
2. Using current θ value, sample a candidate point θ^* from some jumping distribution $q(\theta_1, \theta_2)$, which is the probability of returning a value of θ_2 given a previous value of θ_1. This distribution is also referred to as the proposal or candidate-generating distribution. The only restriction on the jump density in the Metropolis algorithm is that it is symmetric, i.e. $q(\theta_1, \theta_2) = q(\theta_2, \theta_1)$.

3. Given the candidate point θ^*, calculate the ratio of the density at the candidate (θ^*) and current (θ_{t-1}) points

$$r = \frac{P(\theta^*)}{P(\theta_{t-1})} = \frac{f(\theta^*)}{f(\theta_{t-1})}$$

4. If the jump increases the density ($r > 1$), then accept the candidate point (set $\theta_t = \theta^*$) and return to step 2.
5. If the jump decreases the density ($r < 1$), then with probability r accept the candidate point, else reject it and return to step 2.

This algorithm generates a Markov chain ($\theta_0, \theta_1, \ldots, \theta_k, \ldots$), as the transition probabilities from θ_t to θ_{t+1} depends only on θ_t and not ($\theta_0, \ldots, \theta_{t-1}$). Following a sufficient burn-in period (of, say, *k* steps), the chain approaches its stationary distribution and samples from the vector ($\theta_{k+1}, \ldots, \theta_{k+n}$) are samples from *P*(*x*). Hasting (1970) generalized the Metropolis algorithm by using an arbitrary transition probability function $q(\theta_1, \theta_2) = P(\theta_1 \longrightarrow \theta_1)$, and setting the acceptance probability for a candidate point as

$$r = \min\left\{\frac{f(\theta^*)q(\theta^*, \theta_{t-1})}{f(\theta_{t-1})q(\theta_{t-1},\ \theta^*)}, 1\right\}$$

Assuming that the proposal distribution is symmetric in M-H, recovers the original Metropolis algorithm.

3.3.3 Model Selection

Bayesian information criterion (BIC) is a criterion for model selection among a finite set of models. It is based, in part, on the likelihood function and it is closely related to the Akaike information criterion (AIC). When fitting models, it is possible to increase the

likelihood by adding parameters, but doing so may result in over fitting. Both BIC and AIC resolve this problem by introducing a penalty term for the number of parameters in the model; the penalty term is larger in BIC than in AIC.

BIC (Schwarz, 1978) and AIC (Akaike, 1987) are an index used in a number of areas as an aid to choosing between competing models. It is defined as

$$\begin{aligned} \mathrm{BIC} &= [p \times \log(\mathrm{m})] + \mathrm{D} \\ \mathrm{AIC} &= 2p + \mathrm{D} \end{aligned} \qquad (3.5)$$

Where D the deviance is function, and p is the number of estimated parameters. Deviance Information Criterion (DIC) is a hierarchical modeling generalization of the AIC and BIC. It is particularly useful in Bayesian model selection problems where the posterior distributions of the models have been obtained by Markov chain Monte Carlo (MCMC) simulation. Like AIC and BIC it is an asymptotic approximation as the sample size becomes large. It is only valid when the posterior distribution is approximately multivariate normal. DIC is a goodness of fit measure similar to Akaike's, Spiegelhalter et al. (2002) showed that p and D can be defined as

$$\begin{aligned} \mathrm{D} &= -2 \log \mathrm{P}(\mathrm{y}|\mathrm{c}) \\ p &= \overline{\mathrm{D}}(\mathrm{c}) - \widehat{\mathrm{D}}(\mathrm{c}) \end{aligned} \qquad (3.6)$$

Where c is unknown parameter, $\overline{\mathrm{D}}(\mathrm{c})$ is the posterior mean

$$\begin{aligned}\overline{\mathrm{D}}(\mathrm{c}) &= \frac{\sum \mathrm{D}(\mathrm{c})}{\mathrm{m}} \\ &= \frac{\sum(-2\log \mathrm{P}(\mathrm{y}|\mathrm{c}))}{\mathrm{m}}\end{aligned} \tag{3.7}$$

And $\widehat{\mathrm{D}}(\mathrm{c})$ is the point estimate of $\mathrm{D}(\mathrm{c})$

$$\widehat{\mathrm{D}}(\mathrm{c}) = -2\log \mathrm{P}(\mathrm{y}|\overline{\mathrm{c}}) \tag{3.8}$$

DIC Can be defined as

$$\mathrm{DIC} = 2p + \widehat{\mathrm{D}}(\mathrm{c}) \tag{3.9}$$

Lemma (3.1):

If $\mathrm{DIC} = 2p + \widehat{\mathrm{D}}(\mathrm{c})$, then $\mathrm{DIC} = \overline{\mathrm{D}}(\mathrm{c}) + p$

Proof

$$\begin{aligned}\mathrm{DIC} &= 2p + \widehat{\mathrm{D}}(\mathrm{c}) \\ &= 2p + 2\widehat{\mathrm{D}}(\mathrm{c}) - \widehat{\mathrm{D}}(\mathrm{c}) \\ &= 2\overline{\mathrm{D}}(\mathrm{c}) - \widehat{\mathrm{D}}(\mathrm{c}) \\ &= \overline{\mathrm{D}}(\mathrm{c}) + \left(\overline{\mathrm{D}}(\mathrm{c}) - \widehat{\mathrm{D}}(\mathrm{c})\right) \\ &= \overline{\mathrm{D}}(\mathrm{c}) + p\end{aligned}$$

3.4 Estimations of Model Specifications

From the model in (2.24)

$$\begin{aligned} Y_i(t_{ij}) &= Z_i(t_{ij}) + \varepsilon_i(t_{ij}) \\ &= a_i + bt_{ij} + \beta X_i(t_{ij}) + \varepsilon_i(t_{ij}) \end{aligned}$$

Assumed that $a_i \sim N(\mu_a, \sigma_a^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$.

The parameters $\hat{\mu}_a$, $\hat{\sigma}_a$, $\hat{b}$, and $\hat{\beta}$ are the estimators of the unknown parameters μ_a, σ_a, b and β respectively.

Also, from the model in (2.23)

$$h_i\big(t \big| Z_i(t), P_i(t)\big) = h_0(t) exp\big(\gamma Z_i(t) + \omega P_i(t)\big)$$

The parameters $\hat{h}_0$, $\hat{\gamma}$, and $\hat{\omega}$ are the estimators of the unknown parameters h_0, γ and ω respectively.

Faucett et. al. (1996), has successfully adopted the Bayesian approach to study the same model as Wulfsohn. We will use this approach in our modeling, focusing on the estimation of the joint posterior density of all unknown model parameters

$$\Omega = \{(a_1, \dots, a_m), \mu_a, \sigma_a^2, b, \beta, \gamma, \omega, \sigma_\varepsilon^2, (h_{01}, \dots, h_{0K})\}.$$

The joint posterior density of the parameters depends on their prior density and likelihood assumptions. We use $[.]$ to denote marginal and $[.\backslash.]$ to denote conditional densities.

$$h_i\big(t \big| Z_i(t), P_i(t)\big) = h_0(t) exp\left(\gamma\left(a_i + bt_{ij} + \beta X_i(t_{ij})\right) + \omega P_i(t)\right)$$

For the likelihood function, we assume

- The data from different subjects are independent.
- For each subject i, given all the unknown parameters in Ω and covariates (X_i, P_i), the longitudinal data is independent of the survival time.
- For each subject i, given $\{Z_i(t_{ij}), j = 1, \dots, n_i\}$, $\{Y_i(t_{ij})\}_{j=1}^{n_i}$ are independent and $Y_i(t_{ij})$ has normal distribution $N(Z_i(t_{ij}),\ \sigma_\varepsilon^2)$.

Thus, the contribution of subject i to the conditional likelihood is used by Salah, (2008)

$$\begin{aligned}
\left[Y_i(t_{ij}), (s_i, \delta_i) | \Omega, X_i, P_i\right] &= \left[Y_i(t_{ij}) | \Omega, X_i, P_i\right] [(s_i, \delta_i) \backslash \Omega, X_i, P_i] \\
&= \left(\prod_{j=1}^{n_i} \left[Y_i(t_{ij}) \backslash \Omega, X_i, P_i\right] \right) [(s_i, \delta_i) \backslash \Omega, X_i, P_i] \\
&= \left(\prod_{j=1}^{n_i} \left[Y_i(t_{ij}) \backslash a_i, b, \beta, \sigma_\varepsilon^2, X_i\right] \right) [(s_i, \delta_i) \backslash a_i, b, \beta, \gamma, \omega, X_i, P_i] \\
&= \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} exp\left(\frac{-\left(Y_i(t_{ij}) - \left(a_i + bt_{ij} + \beta X_i(t_{ij})\right)\right)^2}{2\ \sigma_\varepsilon^2} \right) \\
&\times h_i^{\delta_i}(s_i) exp\left(-\int_0^{s_i} h_i(t) dt \right),
\end{aligned}$$

where $h_i(t) = h_0(t) \exp\left(\gamma\left(a_i + bt_{ij} + \beta X_i(t)\right) + \omega P_i(t)\right)$.

For subject i, the survival probability at his survival time s_i, using model (2.8) is then given by

$$S_i(s_i) = exp\left(-\int_0^{s_i} h_i(t)dt\right)$$

And using model (2.23) we get

$$\begin{aligned}
S_i(s_i) &= exp\left(-\int_0^{s_i} h_0(t)exp\left(\gamma\left(a_i + bt_{ij} + \beta X_i(t_{ij})\right) + \omega P_i(t)\right)dt\right) \\
&= exp\left\{-\sum_{k=1}^{k_i} h_{0k} \int_{t_{k-1}^h}^{t_k^h} exp\left(\gamma Z_i(t) + \omega P_i(t)\right)dt\right. \\
&\left. -h_{0k_i} \int_{t_{k_i}^h}^{s_i} exp\left(\gamma Z_i(t) + \omega P_i(t)\right)dt\right\} \\
&= exp\left\{-\sum_{k=1}^{k_i} h_{0k} \sum_{j=1}^{J_i(k)} \int_{t_{k(J-1)}^i}^{t_{kJ}^i} exp\left(\gamma Z_i(t) + \omega P_i(t)\right)dt\right. \\
&\left. -h_{0k_i} \sum_{j=1}^{J_i(s_i)} \int_{t_{s_i(J-1)}^i}^{t_{s_iJ}^i} exp\left(\gamma Z_i(t) + \omega P_i(t)\right)dt\right\} \\
&= exp\left\{-\sum_{k=1}^{k_i} h_{0k} \sum_{j=1}^{J_i(k)} exp\left(\gamma\left(a_i + \beta X_i(t_{kJ}^i)\right) + \omega P_i(t_{kJ}^i)\right)\frac{e^{\gamma b t_{kJ}^i} - e^{\gamma b t_{k(J-1)}^i}}{\gamma b}\right. \\
&\left. -h_{0k_i} \sum_{j=1}^{J_i(s_i)} exp\left(\gamma\left(a_i + \beta X_i(t_{s_iJ}^i)\right) + \omega P_i(t_{s_iJ}^i)\right)\frac{e^{\gamma b t_{s_iJ}^i} - e^{\gamma b t_{s_i(J-1)}^i}}{\gamma b}\right\}
\end{aligned} \tag{3.10}$$

Where $k_i = max\{k: t_k^h \le s_i\}$; for $k = 1, \dots, k_i$, $t_{k0}^i = t_{k-1}^h$, $t_{kJ_i(k)}^i = t_k^h$ and $\left(t_{k1}^h, \dots, t_{k(J_i(k)-1)}^i\right)$ are all the grid points in interval $\left(t_{k-1}^h, \dots, t_k^h\right)$ for subject i; $\left(t_{s_i1}^i, \dots, t_{s_i(J_i(s_i)-1)}^i\right)$ are all the grid points in interval $\left(t_{k_i}^h, s_i\right)$ for subject i; $t_{s_i0}^i = t_{k_i}^h$ and $t_{k_iJ_i(s_i)}^i = s_i$.

3.5 Full Conditional Distribution

The Bayesian framework computes the probability of a parameter given the data, posterior distribution $[c \setminus y]$. For any component Ω_c of Ω, the full conditional density of Ω_c given remaining parameters Ω_{-c} is proportional to the posterior density of Ω with the remaining parameters treated as fixed. That is, $[\Omega_c \setminus \Omega_{-c}] \propto [\Omega \setminus data]$.

Based on the joint posterior distribution of the parameters in section 3.4, both Faucett and Thompson (1996) and Wang and Taylor (2001) derive the full conditionals, but we prefer Wang and Taylor's derivation because they do not specify the prior distribution. The full conditional distributions of the parameters appearing only in the longitudinal model can be derived as follows.

- The variance of the measurement error σ_ε^2 by posterior distribution

$$
\begin{aligned}
[\sigma_\varepsilon^2 \setminus .] \quad &\propto \prod_{i=1}^{m} \prod_{j=1}^{n_i} \left[Y_i(t_{ij}) \setminus a_i, b, \sigma_\varepsilon^2, X_i(t_{ij})\right] [\sigma_\varepsilon^2] \\
&\propto \prod_{i=1}^{m} \prod_{j=1}^{n_i} \frac{1}{\sqrt{\sigma_\varepsilon^2}} exp\left(\frac{-\left(Y_i(t_{ij}) - \left(a_i + bt_{ij} + \beta X_i(t_{ij})\right)\right)^2}{2\,\sigma_\varepsilon^2} \right) [\sigma_\varepsilon^2] \\
&\propto \frac{1}{(\sigma_\varepsilon^2)^{\frac{\sum_{i=1}^{m} n_i}{2}}} exp\left(\frac{-\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(Y_i(t_{ij}) - \left(a_i + bt_{ij} + \beta X_i(t_{ij})\right)\right)^2}{2\sigma_\varepsilon^2} \right) [\sigma_\varepsilon^2] \\
&\propto (\sigma_\varepsilon^2)^{-\left(\left(\frac{\sum_{i=1}^{m} n_i}{2} - 1\right) + 1\right)} exp\left(\frac{-\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(Y_i(t_{ij}) - \left(a_i + bt_{ij} + \beta X_i(t_{ij})\right)\right)^2 / 2}{\sigma_\varepsilon^2} \right) [\sigma_\varepsilon^2] \\
&\propto IG(\alpha_0, \beta_0)[\sigma_\varepsilon^2]
\end{aligned}
\tag{3.11}
$$

Where $\alpha_0 = \frac{\sum_{i=1}^{m} n_i}{2} - 1$, and $\beta_0 = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(Y_i(t_{ij}) - \left(a_i + bt_{ij} + \beta X_i(t_{ij})\right)\right)^2}{2}$

- The intercept variance σ_a^2

$$
\begin{aligned}
[\sigma_a^2 \backslash .] &\propto \prod_{i=1}^{m} [a_i \backslash \mu_a, \sigma_a^2]\,[\sigma_a^2] \\
&\propto \prod_{i=1}^{m} \frac{1}{\sqrt{\sigma_a^2}} exp\left(\frac{-(a_i-\mu_a)^2}{2\sigma_a^2}\right)[\sigma_a^2] \\
&\propto \frac{1}{(\sigma_a^2)^{\frac{m}{2}}} exp\left(\frac{-\sum_{i=1}^{m}(a_i-\mu_a)^2}{2\sigma_a^2}\right)[\sigma_a^2] \\
&\propto (\sigma_a^2)^{-((\alpha_0)+1)} exp\left(\frac{-\beta_0}{\sigma_a^2}\right)[\sigma_a^2] \\
&\propto IG(\alpha_0, \beta_0)[\sigma_a^2] \qquad (3.12)
\end{aligned}
$$

Where $\alpha_0 \quad = \frac{m}{2} - 1, and\ \beta_0 \quad = \frac{\sum_{i=1}^{m}(a_i-\mu_a)^2}{2}$

- The intercept mean μ_a

$$
\begin{aligned}
[\mu_a \backslash .] &\propto \prod_{i=1}^{m} [a_i \backslash \mu_a, \sigma_a^2]\,[\mu_a] \\
&\propto \prod_{i=1}^{m} \frac{1}{\sqrt{\sigma_a^2}} exp\left(\frac{-(a_i-\mu_a)^2}{2\sigma_a^2}\right)[\mu_a] \\
&\propto \prod_{i=1}^{m} \frac{1}{\sqrt{\sigma_a^2}} exp\left(\frac{-(a_i^2-2\mu_a a_i+\mu_a^2)}{2\sigma_a^2}\right)[\mu_a] \\
&\propto \frac{1}{(\sigma_a^2)^{m/2}} exp\left(\frac{-(\sum_{i=1}^{m} a_i^2 - 2\mu_a \sum_{i=1}^{m} a_i + m\mu_a^2)}{2\sigma_a^2}\right)[\mu_a] \\
&\propto \frac{1}{(\sigma_a^2)^{m/2}} exp\left(\frac{-(\sum_{i=1}^{m} a_i^2 - 2\mu_a m\bar{a}_i + m\mu_a^2)}{2\sigma_a^2}\right)[\mu_a] \\
&\propto exp\left(\frac{-\left(\frac{\sum_{i=1}^{m} a_i^2}{m} - 2\mu_a\bar{a}_i + \mu_a^2\right)}{2\,\frac{\sigma_a^2}{m}}\right)[\mu_a] \\
&\propto exp\left(\frac{-(\mu_a-\alpha_0)^2}{2\beta_0}\right)[\mu_a] \\
&\propto N(\alpha_0, \beta_0)[\mu_a] \qquad (3.13)
\end{aligned}
$$

Where $\alpha_0 = \frac{\sum_{i=1}^{m} a_i}{m}\ , and\ \beta_0 = \frac{\sigma_a^2}{m}$

- The random intercept a_i , for $i = 1,2, \dots m$

$$
\begin{aligned}
[a_i \backslash .] \quad &\propto \prod_{j=1}^{n_i} \left[Y_i(t_{ij}) \backslash a_i, b, \beta, \sigma_\varepsilon^2\right] [a_i \backslash \mu_a] \prod_{i=1}^{m} h_i^{\delta_i}(s_i) exp\left(-\int_0^{s_i} h_i(t)dt\right) \\
&\propto exp\left(\frac{-\sum_{j=1}^{n_i}\left(Y_i(t_{ij}) - \left(a_i + bt_{ij} + \beta X_i(t_{ij})\right)\right)^2}{2\sigma_\varepsilon^2}\right) \\
&* exp\left(\frac{-(a_i - \mu_a)^2}{2\sigma_a^2}\right) exp(\gamma \delta_i a_i) S_i(s_i) \\
&\propto exp\left(\frac{-\left[(a_i - \alpha_0)^2\right]}{2\,\beta_0}\right) exp(\gamma \delta_i a_i) S_i(s_i) \\
&\propto N(\alpha_0, \beta_0) \qquad (3.14)
\end{aligned}
$$

Where $\alpha_0 = \frac{\sum_{j=1}^{n_i}\left(Y_{ij} - \left(bt_{ij} + \beta X_i(t_{ij})\right)\right)}{\sigma_\varepsilon^2} + \frac{\mu_a}{\sigma_a^2}$ and $\beta_0 = \left(\frac{n_i}{\sigma_\varepsilon^2} + \frac{1}{\sigma_a^2}\right)^{-1}$

- The average rate of decline of the marker b

$$
\begin{aligned}
[b \backslash .] \quad &\propto \prod_{i=1}^{m}\prod_{j=1}^{n_i} \left[Y_i(t_{ij}) \backslash a_i, b, \beta, \sigma_\varepsilon^2\right] [b] \prod_{i=1}^{m} h_i^{\delta_i}(s_i) exp\left(-\int_0^{s_i} h_i(t)dt\right) \\
&\propto \prod_{i=1}^{m}\prod_{j=1}^{n_i} exp\left(\frac{-\left(Y_i(t_{ij}) - \left(a_i + bt_{ij} + \beta X_i(t_{ij})\right)\right)^2}{2\,\sigma_\varepsilon^2}\right) \\
&* exp\left(b\gamma \sum_{i=1}^{m} \delta_i\, s_i\right)\left(\prod_{i=1}^{m} S_i(s_i)\right)[b] \\
&\propto exp(-1/2\beta_0)(b - m_b)^2) \\
&* exp\left(b\gamma \sum_{i=1}^{m} \delta_i\, s_i\right)\left(\prod_{i=1}^{m} S_i(s_i)\right)[b] \\
&\propto N(m_b, \beta_0)[b] \qquad (3.15)
\end{aligned}
$$

Here $m_b = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} t_{ij}\left(Y_i(t_{ij}) - \left(a_i + \beta X_i(t_{ij})\right)\right)}{\sum_{i=1}^{m}\sum_{j=1}^{n_i} t_{ij}^2}$ and $\beta_0 = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{m}\sum_{j=1}^{n_i} t_{ij}^2}$

- The effect of the variables on the marker β_l, $l = 1, \dots, p$

$$
\begin{aligned}
[\beta_l \backslash .] \quad &\propto \prod_{i=1}^{m} \prod_{j=1}^{n_i} \left[Y_i(t_{ij}) \backslash a_i, b, \beta, \sigma_\varepsilon^2\right] [\beta_l] \prod_{i=1}^{m} h_i^{\delta_i}(s_i) exp\left(-\int_0^{s_i} h_i(t) dt\right) \\
&\propto \prod_{i=1}^{m} \prod_{j=1}^{n_i} exp\left(\frac{-\left(Y_i(t_{ij}) - \left(a_i + bt_{ij} + \beta X_i(t_{ij})\right)\right)^2}{2\,\sigma_\varepsilon^2}\right) \\
&* exp\left(\beta_l \gamma \sum_{i=1}^{m} \delta_i X_{il}(s_i)\right)\left(\prod_{i=1}^{m} S_i(s_i)\right)[\beta_l] \\
&\propto exp\left(-(1/2\beta_0)\left(\beta_l - m_{\beta_l}\right)^2\right) \\
&* exp\left(\beta_l \gamma \sum_{i=1}^{m} \delta_i X_{il}(s_i)\right)\left(\prod_{i=1}^{m} S_i(s_i)\right)[\beta_l] \\
&\propto N\left(m_{\beta_l}, \beta_0\right)[\beta_l] \qquad (3.16)
\end{aligned}
$$

Where

$$m_{\beta_l} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n_i} X_{il}(t_{ij}) \left(Y_i(t_{ij}) - \left(a_i + bt_{ij} + \beta(-l) X_i(-l)(t_{ij})\right)\right)}{\sum_{i=1}^{m} \sum_{j=1}^{n_i} X_{il}^2(t_{ij})}$$

and

$$\beta_0 = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{m} \sum_{j=1}^{n_i} X_{il}^2(t_{ij})}$$

$\beta(-l) = \left(\beta_1, \dots, \beta_{l-1}, \beta_{l+1}, \dots, \beta_p\right)$ and $X_i(-l)$ the remaining covariates after l^{th} covariate X_{il} is excluded from X_i.

- The baseline hazard $(h_{01}, \dots, h_{0K})$

$$
\begin{aligned}
[(h_{01}, \dots, h_{0K})\backslash.] \quad &\propto \prod_{i=1}^{m} h_i^{\delta_i}(s_i) exp\left(-\int_0^{s_i} h_i(t)dt\right)[(h_{01}, \dots, h_{0K})] \\
&\propto \prod_{i=1}^{m} \left(h_0 exp\big(\gamma Z_i(s_i) + \omega P_i(s_i)\big)\right)^{\delta_i} \\
&* exp\left(-\int_0^{s_i} h_0 exp\big(\gamma Z_i(t) + \omega P_i(t)\big)dt\right)[(h_{01}, \dots, h_{0K})] \\
&\propto \prod_{k=1}^{K} h_{0k}^{d_k} exp\left(-h_{0k}\left\{\sum_{i:s_i \geq t_k^h} \int_{t_{k-1}^h}^{t_k^h} exp\big(\gamma Z_i(t) + \omega P_i(t)\big)dt\right.\right. \\
&\left.\left. + \sum_{i:s_i \in (t_{k-1}^h, t_k^h)} \int_{t_{k-1}^h}^{s_i} exp\big(\gamma Z_i(t) + \omega P_i(t)\big)dt\right)\right\}[(h_{01}, \dots, h_{0K})] \\
&\propto \prod_{k=1}^{K} h_{0k}^{d_k} exp\left(-h_{0k}\{\beta_{h_{0k}}\}\right) \\
&\propto \prod_{k=1}^{K} G\big(d_k, \beta_{h_{0k}}\big)[(h_{01}, \dots, h_{0K})]
\end{aligned} \tag{3.17}
$$

Where

$$
\begin{aligned}
\beta_{h_{0k}} \quad &= \sum_{i:s_i \geq t_k^h} \int_{t_{k-1}^h}^{t_k^h} exp\big(\gamma Z_i(t) + \omega P_i(t)\big)dt \\
&+ \sum_{i:s_i \in (t_{k-1}^h, t_k^h)} \int_{t_{k-1}^h}^{s_i} exp\big(\gamma Z_i(t) + \omega P_i(t)\big)dt
\end{aligned}
$$

and d_k is the number of event occurring in time interval $\left(t_{k-1}^h, t_k^h\right]$ $k = 1, \dots, K$.

- The regression coefficient of the effect of the marker on the risk γ

$$
\begin{aligned}
[\gamma \backslash .] \quad &\propto \prod_{i=1}^{m} h_i^{\delta_i}(s_i) exp\left(-\int_0^{s_i} h_i(t)dt\right)[\gamma] \\
&\propto \prod_{i=1}^{m} \left(h_0 exp\left(\gamma Z_i(t) + \omega_l P_{il}(t)\right)\right)^{\delta_i} exp\left(-\int_0^{s_i} h_i(t)dt\right)[\gamma] \\
&\propto (h_0)^{\sum_{i=1}^{m}\delta_i} exp\left(\omega_l \sum_{i=1}^{m} \delta_i P_{il}(s_i)\right) exp\left(\gamma \sum_{i=1}^{m} \delta_i Z_i(s_i)\right) \\
&* \prod_{i=1}^{m} exp\left(-\int_0^{s_i} h_i(t)dt\right)[\gamma] \\
&\propto exp\left(\gamma \sum_{i=1}^{m} \delta_i Z_i(s_i)\right)\left(\prod_{i=1}^{m} S_i(s_i)\right)[\gamma]
\end{aligned}
$$

- The regression coefficient of the effect of the other potential covariates ω

$$
\begin{aligned}
[\omega_l \backslash .] \quad &\propto \prod_{i=1}^{m} h_i^{\delta_i}(s_i) exp\left(-\int_0^{s_i} h_i(t)dt\right)[\omega_l] \\
&\propto \prod_{i=1}^{m} \left(h_0 exp\left(\gamma Z_i(t) + \omega_l P_{il}(t)\right)\right)^{\delta_i} exp\left(-\int_0^{s_i} h_i(t)dt\right)[\omega_l] \\
&\propto (h_0)^{\sum_{i=1}^{m}\delta_i} exp\left(\gamma \sum_{i=1}^{m} \delta_i Z_i(s_i)\right) exp\left(\omega_l \sum_{i=1}^{m} \delta_i P_{il}(s_i)\right) \\
&* \prod_{i=1}^{m} exp\left(-\int_0^{s_i} h_i(t)dt\right)[\omega_l] \\
&\propto exp\left(\omega_l \sum_{i=1}^{m} \delta_i P_{il}(s_i)\right)\left(\prod_{i=1}^{m} S_i(s_i)\right)[\omega_l]
\end{aligned}
$$

We do not have the form of a standard distribution for parameters γ, ω since their full conditional density have no conjugate prior. For each of these parameters, we propose using probability density function as normal proposal density.

We note that the full conditional distributions of the parameters $\sigma_\varepsilon^2, \sigma_a^2, \mu_a, (h_{01}, \dots, h_{0K})$ in our joint model are a product of its prior density and some standard distribution which are conjugate priors for these parameters.

While the conditional distribution of the parameters β_l, b, if the contributions from the survival data are ignored, then the normal distribution are conjugate priors.

The main difficulty which we will meet in the prior distributions is that when no standard form appears in the posterior distribution. In general, we do not have performance in choosing priors for the parameters γ, ω since their full conditional densities have no conjugate priors. One may use normal priors for γ, ω since they take values belong to the real line.

We can see no prior information about the unknown parameter, we can consider using diffuse priors. A diffuse prior plays a minimal role in the posterior distribution and often has no finite mean or finite variance the most often used diffuse priors are improper priors whose density has no finite integral. One of the most often used improper priors for $\sigma_\varepsilon^2, \sigma_a^2$ and h_{0K} has density (improper) $f(x) \propto \frac{1}{x}$ which corresponds to $IG(0,1)$ or $G(0,0)$. This improper prior assumes that log-transformation of the parameters is uniform over the real line. For the $b, \beta_l, \gamma, \omega_l$, the most natural and popular improper prior is

uniform over the real line, which corresponds to a $N(\mu, \infty)$ (see Gelman et al., 1995, and Salah, 2008).

Chapter Four
Applications

In this chapter, the aim is to evaluate the performance of the joint model simulation studies to apply in R language. We investigate how well the parameters in our joint model can be estimated of effects, bias and coverage rate. Moreover, we evaluate the performance of the joint model and separate model in real data, we apply the models to the primary biliary cirrhosis (PBC) data, which collected by the Mayo Clinic from 1974 to 1984, our aim was to develop a model first to estimate the association between the risk of PBC disease and these biomarkers, and second to make predictions about the future rate of biomarkers and probability of PBC disease, finally, we compare these results between the joint and separate modeling approach.

4.1 Simulated Data

4.1.1 Generation of The Data

We setup our simulation study by samples of size $m = 100$ subjects are randomized. Each longitudinal marker in model (2.24), $Y_i(t_{ij})$, $i = 1, \dots, m$; $j = 1, \dots, n_i$, was simulated as the sum of the trajectory function $Z_i(t_{ij})$ and the error terms $\varepsilon_i(t_{ij})$, each subject has his observed longitudinal measured $n_i = 10$ at time points $t_1 = 0.1, \dots, t_{10} = 1$, until the relapse or reaches the end of the study.

For the survival data, we consider a model in (2.23) to be for i, where P_i is binary baseline covariates with half of the subjects having one and the other half having zero. We took the mean of the Poisson process at time t as in (2.23) to be for $i = 1, \dots, 100$.

Estimated to get an initial estimate of the parameters $\Omega = \{\mu_a, \sigma_a^2, \sigma_\varepsilon^2, b, \beta, \gamma, \omega\}$, say $\Omega^0 = \left\{\mu_a{}^0, \sigma_a^{2\,0}, \sigma_\varepsilon^{2\,0}, b^0, \beta^0, \gamma^0, \omega^0\right\}$ and use them as initial values in MCMC sampler.

4.1.2 Summary Statistics

Salah, (2008,p111); used θ be ageneric notation for whatever parameter we are considering and its true value (with which we generate the data sets) be θ_0. Let $\hat{\theta}_i$, $\hat{\theta}_i(0.025), \hat{\theta}_i(0.975)$ and $\widehat{Var}(\hat{\theta}_i)$ be its point estimate, 2.5% quantile estimate, 97.5% quantile estimate, and standard error estimate, respectively, obtained from analyzing the i^{th} data set. The Estimate, Monte Carlo Standard Deviation (MCSD), Mean Squared Error (MSE), 95% Confidence Coverage Rate (CCR), and Bias in Percentage Terms (BPT) are calculated as follows:

$$\text{Estimate} = \hat{\theta} = \frac{\sum_{i=1}^{100} \hat{\theta}_i}{100} \tag{4.1}$$

$$\text{MCSD} = \sqrt{\frac{\sum_{i=1}^{100}(\hat{\theta}_i - \bar{\hat{\theta}})^2}{100-1}} \tag{4.2}$$

$$\text{Bias in Percentage} = \frac{\hat{\theta} - \theta_0}{\theta_0} * 100\% \tag{4.3}$$

$$\text{MSE} = \frac{\sum_{i=1}^{100}(\hat{\theta}_i - \theta_0)^2}{100} \tag{4.4}$$

$$\text{Coverage Rate} = \frac{\sum_{i=1}^{100} I\left(\theta_0 \in \hat{\theta}_i \mp 1.96\sqrt{\text{Var}(\hat{\theta}_i)}\right)}{100}, \text{or} \tag{4.5}$$

$$\text{Coverage Rate} = \frac{\sum_{i=1}^{100} I\left(\theta_0 \in \left(\hat{\theta}_i(0.025), \hat{\theta}_i(0.975)\right)\right)}{100} \tag{4.6}$$

With Bayesian approach, $\hat{\theta}_i$, $\hat{\theta}_i(0.025), \hat{\theta}_i(0.975)$ and $\widehat{Var}(\hat{\theta}_i)$ are the posterior median, the posterior (2.5%) quantile, the posterior 97.5% quantile, and posterior variance of θ,

respectively, estimated as the sample median, the sample (2.5%) quantile, the sample 97.5% quantile, and the sample variance, respectively, of a single series of SCMH sampler for θ. By doing so, we have used frequency evaluations of Bayesian inferences. According to Gelman et al. (1995,p104), frequency evaluations of Bayesian inferences hinge on the fact that if the true data distribution is included in the class of models, so that $f(y) \equiv p(y|\theta)$ for some θ, and under some regularity conditions, then in repeated

sampling with fixed θ, the posterior mode, mean, and median are asymptotically unbiased and efficient estimates for θ, and the mode is approximately normally distribution for large sample sizes. Our primary simulation study shows that the estimates based on the sample mean and the coverage rates of (4.5) are almost the same as the estimates based on the sample median and the coverage rates of (4.6), respectively, for all parameters. Our estimation results for the other parameters are based on the sample mean and the coverage rate of (4.5).

We use 100 data replications, thus the resulting estimates are subject to sampling variation (Monte Carlo Error). This variation for the point estimate can be calculated as

$$\hat{p} = \frac{MCSD}{\sqrt{100}} \tag{4.7}$$

For the approximate 95% coverage rate $\hat{p}$ the Monte Carlo error is

$$\text{MCE} = \sqrt{\frac{\hat{p}(1-\hat{p})}{100}} \tag{4.8}$$

Are almost the same as the estimates based on the sample median and the coverage rates. For the model (2.24), we fit a mixed model of the form

$$Y_i(t) \quad = a_i + bt + \beta X_i + \sigma_\varepsilon^2 * N(0,1)$$

by using R language packages “JM” and “lcmm”.

For the model (2.23), we use a Cox proportional hazards model and maximize the partial likelihood function.

4.1.3 Choosing Prior Distributions

We consider marginal uninformative prior by choosing uninformative prior for each parameter when assuming other parameters are known, and then using the product of these individual uninformative prior as our find prior. This approach leads to use of flat prior for μ_a, b, β, γ, ω and use of $1/\sigma_\varepsilon^2$ for σ_ε^2, $1/\sigma_a^2$ for σ_a^2 and of $1/h_{0k}$ for h_{0k}. With these prior, the corresponding full conditional distribution in chapter (3), making sampling based on these full conditional distribution straightforward, see Box &Tiao (1973).

In general, one should be cautious when using improper priors since this may lead to improper posterior distributions. As mentioned above, improper priors are assigned to variance parameters $\sigma_\varepsilon^2, \sigma_a^2$ and baseline hazard parameters h_{0k}.To ensure that the posterior densities for all parameter are proper and have finite mean and variance, we can consider our priors for the parameters μ_a, b, β, γ, and ω to be $U(-10^4, 10^4)$, for $\sigma_\varepsilon^2, \sigma_a^2$ and h_{0k} to be $[x] \propto \frac{1}{x} U(10^{-4}, 10^4)$.

4.1.4 Sampling Based on Full Conditional Densities

In this study, we discuss some issues related to implementing the SCMH sampler algorithm for our joint modeling approach. Throughout this study, the true values of the parameters, which are used to generate the 100 data sets, are in table 4.1.

Sampling for parameters σ_ε^2, σ_a^2, and μ_a is straight forward since their full conditional densities are of standard forms in model (3.11), (3.12), (3.13). For the parameters whose full conditional densities do not have a standard form, we use a Metropolis-Hastings (M-H) step to update the iterations of SCMH sampler. This method might cause slow convergence of the SCMH sampler if the proposal densities in the (M-H) step are not close to the target densities and hence the iterations do jump infrequently or jump with tiny steps. Therefore, we will use the full conditional density as a proposal density in Gibbs sampler algorithm, and in sampling process each updating step for these parameters, a new draw from the full conditional density is always accepted.

For the parameter in the longitudinal model a_i, b, and β cannot draw a random variate from these densities directly due to the terms from the survival data. We use the Metropolis-Hastings (M-H) algorithm to obtain the update in the Gibbs sampling sequence. We use the full conditional densities, which are obtained from the longitudinal model by ignoring the survival data, as the proposal densities (3.14), (3.15), and (3.16). These proposal densities are all normal and so sampling from them is simple. These proposal densities appear to be good choices since with or without the survival data, the behavior of the parameters appearing in the longitudinal model should remain relatively unchanged.

For the survival parameter γ and ω have no standard distribution, it is just an algebraic expression which come from the contribution of the longitudinal and survival data so that, for such parameters, one cannot draw random variates from their full conditional densities. For each of these parameters we propose using a normal density as a proposal density, and then using SCMH.

4.1.5 Simulation Results

All the parameters were assumed independent a prior and assigned non-informative prior. So, we specify for each parameter a reasonable and fast method to obtain an initial value (using the {lm} function in R, see Appendix 1). Estimates of parameters are based on a subsequent chain of 5,000 iterations. More iterations would of course be preferable, but limited computing power made these calculations very time-consuming.

The Gibbs sampler was run 5000 iterations after 1000 burn-in. The histogram, the time series plots of one sequence of Gibbs samples for different number of iterations and the average number of these iterations for all the parameter are presented in Figures 4.1.a. Visual inspection suggested that convergence was fairly rapid; as shown in these Figures.

Table 4.1 shows the true and estimated values of the model parameters from a simulation study. Moreover, MCSD, Mean Square Error, Bias as percent of true parameter, MCE, and 95% highest posterior density intervals for each parameter in the joint model, are represented in the same table. Parameter estimations in the joint modeling results have small standard errors for all parameters; the mean parameters are very mild and close to the true values. The small biases of the estimates are due to Monte Carlo simulation error. For the parameters in the longitudinal model, the estimates are less biased since they have

a closed distribution form as mentioned above. Due to including the survival information, the joint model produces slightly less biased estimates for mean intercept (μ_a) and slope (b). The estimates of all the parameters from the joint modeling analysis are quite accurate and efficient.

Figures 4.1.a Posterior histogram, time series and average values plots, respectively for all parameter values at 5000 iterations, using Gibbs sampler.

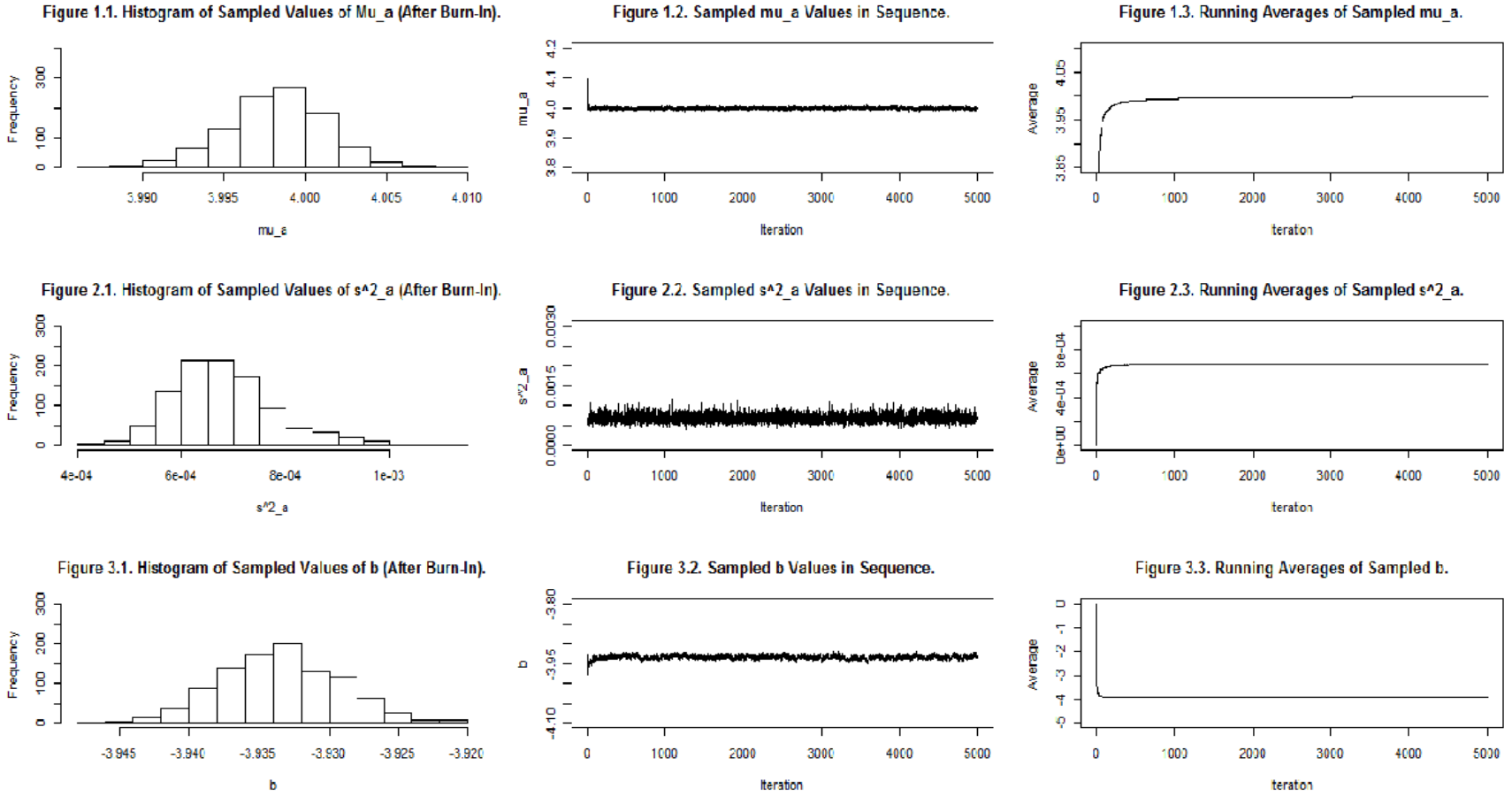


Figure 4.1. Histogram of Sampled Values of gamma (After Burn-In).

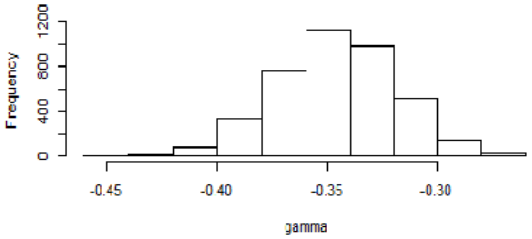


Figure 4.2. Sampled gamma Values in Sequence.

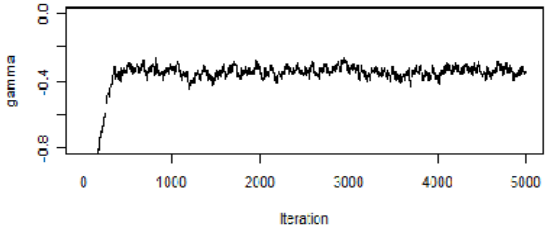


Figure 4.3. Running Averages of Sampled gamma.

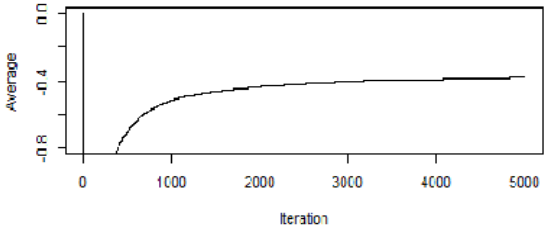


Figure 5.1. Histogram of Sampled Values of h (After Burn-In).

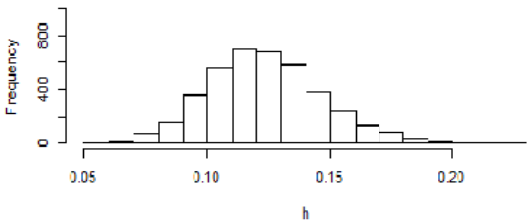


Figure 5.2. Sampled h Values in Sequence.

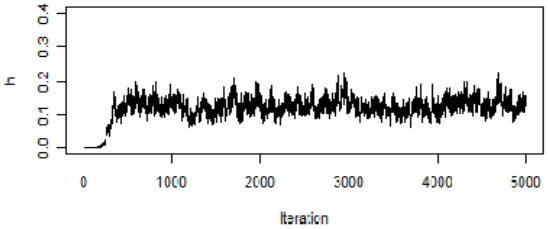


Figure 5.3. Running Averages of Sampled h.

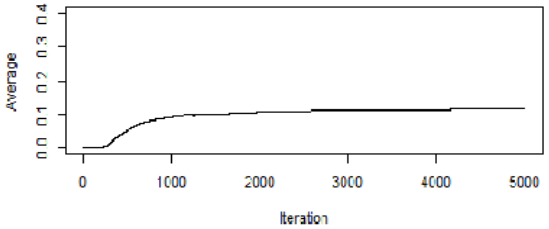


Figure 6.1. Histogram of Sampled Values of sigmase (After Burn-In).

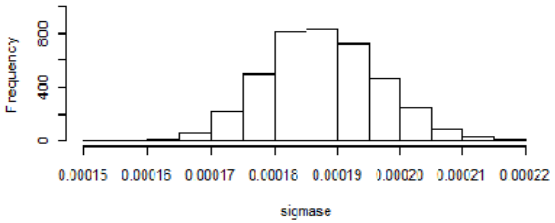


Figure 6.2. Sampled sigmase Values in Sequence.

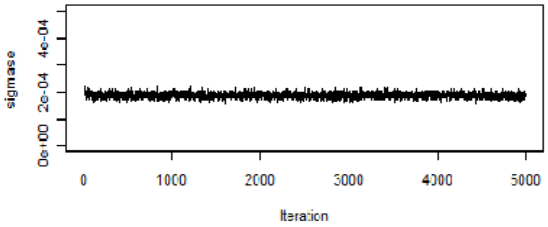


Figure 6.3. Running Averages of Sampled sigmase.

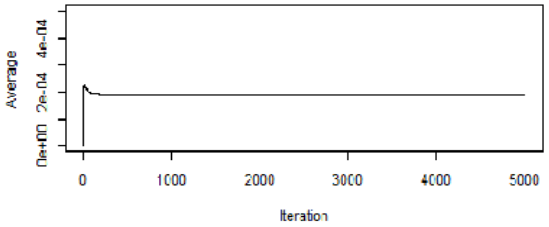


Table 4.1 Monte Carlo summary statistics of the parameter estimates

Parameter	True value	Estimate	MCSD	MSE	Bias%	MCE	95% C.I.		CCP
							Lower	Upper	
μ_a	4	3.952	9×10^{-3}	3.1×10^{-3}	1.2%	1.89×10^{-4}	3.9108	3.9932	0.964
σ_a^2	6×10^{-4}	5.89×10^{-4}	7.8×10^{-9}	8.8×10^{-5}	2.5%	1.7×10^{-7}	4.4×10^{-4}	7.8×10^{-4}	0.925
b	-4	-3.94	0.00206	0.0454	1.15%	9×10^{-5}	-3.864	-4.016	0.966
γ	-0.35	-0.346	7.7×10^{-4}	0.0277	1.09%	5.5×10^{-5}	-0.401	-0.2937	0.967
h^c	0.12	0.123	5.4×10^{-4}	0.0232	-2.8%	4.6×10^{-5}	0.0816	0.1722	0.916
σ_ε^2	18×10^{-5}	18.7×10^{-5}	$1.4 \times e^{-10}$	1.19×10^{-5}	-4.3%	$2.37 \times e^{-8}$	17×10^{-5}	20×10^{-5}	0.871

4.2 Real Data

4.2.1 Application to Real Data Set

As an illustrative application of joint modeling we consider the primary biliary cirrhosis (PBC) data collected by the Mayo Clinic from 1974 to 1984 [14]. PBC is a rare but fatal chronic liver disease of unknown cause characterized by inflammatory destruction of the small bile ducts within the liver, which eventually leads to cirrhosis of the liver.

Various biomarkers such as bilirubin, prothrombin time and albumin were collected longitudinally, and interest is on examining whether these biomarkers relate to the natural history of disease. In this study 312 patients are considered of whom 158 were randomly assigned to receive D-penicillamine whereas the other 154 were randomly assigned to the placebo group. Baseline covariates, for example age and gender, were measured at entry time of the study. Multiple repeated laboratory results from irregular follow-up visits are also available in this dataset. The original clinic protocol had specified visits at 6 months, 1 year, and annually thereafter, but 'extra' visits could occasionally occur due to worsening medical condition. The number of visits ranges from 1 to 16, and the median number of repeated measurements is 5. The median interval between visits is approximately 1 year. By the end of the study, 140 out of the 312 patients had died and the observed event time ranges from 41 to 5225 days.

In the original Mayo model, the prediction model for survival is based on age, total serum bilirubin value, serum albumin value, prothrombin time, and the presence or absence of edema and diuretic therapy. For simplicity, our joint model will only consider the repeated measurements of total serum bilirubin value as our longitudinal covariates, and

age, gender and treatment group as our survival covariates. Based on the clinical literature, we performed a logarithmic transformation of serum bilirubin to be used for our analysis and death is defined as our event of interest.

We are interested in testing for a treatment effect on survival after adjusting for the longitudinal bilirubin levels. Due to the right skewness of the observed serum bilirubin level, we will work with the natural logarithm of serum bilirubin for the remainder of this analysis.

As a descriptive analysis the Kaplan-Meier estimate for the time to death and the subject-specific longitudinal profiles for the two treatment groups are depicted in Figures 4.1.b and 4.1.c. We observe that in both groups patients show similar variability in their longitudinal profiles, whereas from the Kaplan-Meier estimate in Figure 4.1.b it seems that the DPCA group has slightly higher survival than the placebo group after the fourteen months of follow-up.

Figure 4.1.b (PBC data) The Kaplan-Meier estimate of the survival function for the two treatment groups.

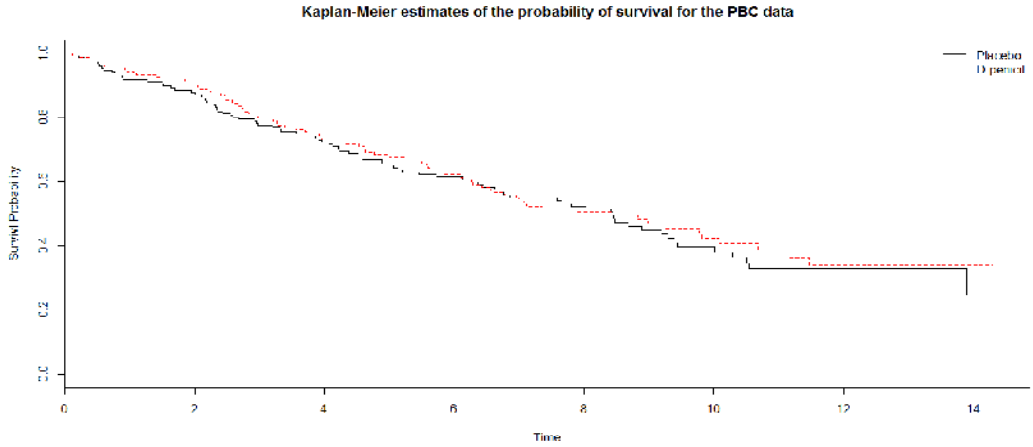
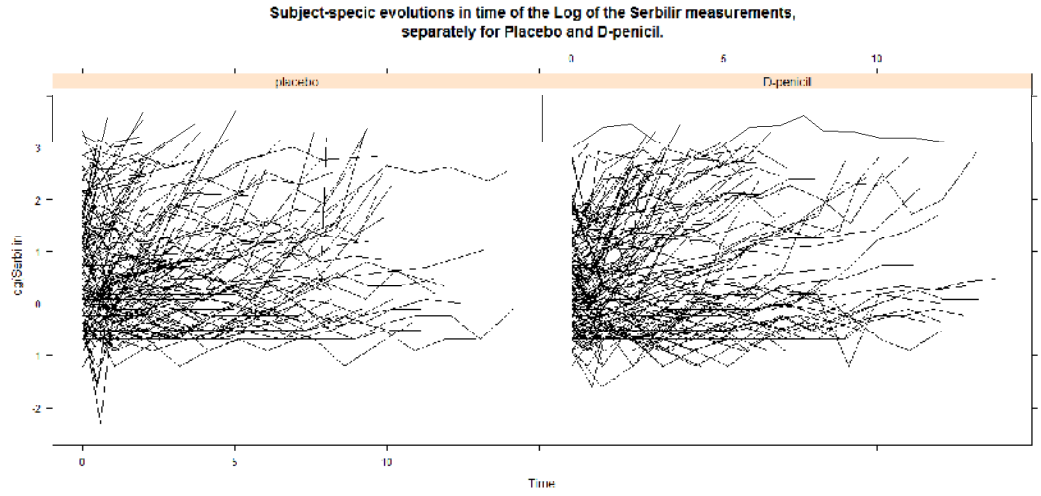


Figure 4.1.c (PBC data) The log longitudinal data for the two treatment groups.



4.2.2 The Joint Model

For our joint model on the PBC dataset, we need to specify a growth curve model and a survival model. Similarly, we will use a linear mixed model with a fixed slope and random intercept for our trajectory function:

$$Y_i(t_{ij}) = a_i + bt_{ij} + \varepsilon_i(t_{ij})$$

$$a_i \sim N(\mu_a, \sigma_a^2)$$

And

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2) \tag{4.9}$$

Where $Y_i(t_{ij})$ are the repeated log serum bilirubin measurements, and the random intercept and random error are approximated by a normal distribution. The mean and the variance of the random intercept, μ_a and σ_a^2 are unknown and has to be estimated too.

Consequently, our hazard function for the joint model will be:

$$\begin{aligned} h_i\left(t \middle| Z_i(t), P_i(t)\right) &= h_0 exp\left(\gamma Z_i(t) + \omega P_i(t)\right) \\ &= h_0 exp\left(\gamma\left(a_i + bt_{ij}\right) + \omega P_i(t)\right) \end{aligned} \tag{4.10}$$

where γ is the regression parameter for the log serum bilirubin trajectory and $\omega = \{\omega_{age}, \omega_{sex}, \omega_{drug}\}$ are the regression parameters for the survival covariates age, gender and treatment. For simplicity, we assumed that the baseline hazard is a constant. When $\gamma = 0$ in model (4.10) then we got separate model.

Table 4.2 shows the result of the joint modeling approach on the DPCA subset of the PBC data. In addition to that, we also performed separate survival model analyses using

available measurement of log serum bilirubin that is taken closest to the time of the event. The purpose of the separate model analyses is to illustrate how the application of our methodology serves our motivation of modeling the longitudinal outcome that occurs at the specific event time in order to have a more accurate representation of its effect on survival. Besides this measurement, covariates age, gender and treatment are also included to the Cox's model of our joint model.

A comparison between the separate model with the joint model reveals some interesting features. In particular, we observe that the regression coefficient for drug variable is larger in magnitude in the joint model, which results in a slightly stronger treatment effect. Much stronger bias is observed for the effect of the DPCA, with estimated regression coefficient $(\gamma_{bil} = 1.06275)$ for separate model and $(\gamma = 1.2558)$ for the joint model, which indicates that the DPCA measurements effected the survival under the joint model more than in the separate model.

Table 4.2 Parameter estimates from joint and separate modeling approaches

Parameter	Joint Model			Separate Model		
	Estimate	Std. Error.	P-Value	Estimate	Std. Error.	P-Value
μ_a	0.729	0.036	<0.001	--	--	--
σ_a	0.914	0.022	<0.001	--	--	--
b	0.116	0.005	<0.001	--	--	--
γ	1.2558	0.0809	<0.0001	(γ_{bil}) 1.06275	0.08437	0.0000
ω_{age}	2.0228	0.0693	<0.0001	1.36192	0.38189	0.0012
ω_{sex}	-0.0965	0.2086	0.6438	-0.09331	0.21343	0.6621
ω_{drug}	-0.1578	0.1238	0.3241	-0.10186	0.15871	0.5210
σ_ε	0.519	0.021	<0.001	--	--	--
Model fit	**AIC**	**BIC**	**Log.Lik**	**AIC**	**BIC**	**Log.Lik**
	5068.88	5106.31	-2524.44	3074.721	3108.153	-1531.36
	5068.88	5106.31	-2524.44	3074.721	3108.153	-1531.36

As an alternative to the Wald test, a likelihood ratio test (LRT) and AIC can be also used to test for a longitudinal effect. To perform this test we need to fit the joint model under the null hypothesis of no longitudinal effect in the survival sub model. The results are shown in Table 4.3.

Table 4.3 Model selection criteria

Model	AIC	BIC	log.Lik	LRT	df	p.value
Joint	4049.54	4113.17	-2017.77	0.15	1	0.00697
Separate	4071.39	4138.76	-2017.69			

It is clear that the joint model fit the data better than the separate model since $AIC_{Joint} = 4049.54$ is smaller than the $AIC_{Sparate} = 4071.39$. Moreover, the p-value indicates that there is significant deference between the two models due to longitudinal data.

4.3 Summary and Conclusions

Joint modeling of longitudinal and time-to-event data is one of the most rapidly evolving areas of current biostatistics research, with several extensions of the standard joint model presented here already proposed in the literature. These include, among others, handling multiple failure types, considering categorical longitudinal outcomes, assuming that several longitudinal outcomes affect the time-to-event, and associating the two outcomes via latent classes instead of random effects.

Furthermore, one very promising subfield which has emerged within the general joint modelling framework, is the use of these models in personalized medicine. In particular, there is lately a great need for tools that can help physicians take better informed decisions regarding their actions for the specific patients that they treat and not for an 'average' patient. Two features of joint models that allow them to become such a flexible dynamic tool, is the use of random effects and their time-dependent nature. For instance, as longitudinal information is collected for patients, we can continuously update the predictions of their survival probabilities, and therefore be able to discern between patients with low and high risk for an event.

One of the main practical limitations for joint modeling in finding its way into the toolbox of modern statisticians was the lack of free and reliable software. The R package JM (it can be downloaded from the CRAN website http://cran.r-project.org/) has been developed to fill this gap to some extent. JM has a user-friendly interface to fit joint models and also provides several supporting functions that extract or calculate various

quantities based on the fitted model (e.g., residuals, fitted values, empirical Bayes estimates, various plots, and others).

From the simulation results, we find that the joint model method usually has the smallest biases, quite accurate and efficient, and the largest coverage rates close to the nominal level (95%), joint model method produces less biased estimates and more reliable standard errors. Parameters μ_a, σ_a^2, b, γ, h^c and σ_ε^2 estimations in the joint modeling results 3.952, 5.89×10^{-4}, -3.94, -0.346, 0.123, and 18.7×10^{-5} respectively are very mild and close to the true values. The biases of the estimates 1.2%, 2.5%, 1.15%, 1.09%, -2.8%, and -4.3% respectively are small due to Monte Carlo simulation error. Also, we have small standard errors for all parameters. For the parameters in the longitudinal model, the estimates are less biased since they have a closed distribution form as mentioned above. Due to including the survival information, the joint model produces slightly less biased estimates for mean intercept μ_a and slope b.

From the real data results, we see that the regression coefficient for drug variable was larger in magnitude in the joint model, which results in a slightly stronger treatment effect. Much stronger bias is observed for the effect of the DPCA, with estimated regression coefficient (γ_{bil}=1.06275) for separate model and (γ =1.2558) for the joint model, which indicates that the DPCA measurements effected the survival under the joint model more than in the separate model. Compared to the joint model, the separate analysis produces relatively large bias in most model parameters, also the joint model in general produces more accurate point estimates than the separate analysis. Comparisons of the mean square errors and AIC between the two approaches ($AIC_{Joint} = 4049.54$ is

smaller than the $AIC_{Sparate} = 4071.39$) again suggest that the joint model performs superior to the separate analysis.

REFERENCES

Akaike, H. (1974). *A new look at the statistical model identification.* IEEE Transactions on Automatic Control, AC-19, 716-723.

Allison, P. D. (1984). *Event history analysis*. Thousand Oaks, CA: Sage.

Allison, P. D. (1995). *Survival analysis using SAS*: *A practical guide.* Cary, NC: SAS Institute.

Bycott, P. and Taylor, J. (1998). *A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model*. Statistics in Medicine, 17, 2061-2077.

Chen, M. H., Ibrahim, J. G. and Sinha, D. (2004). *A new joint model for longitudinal and survival data with a cure fraction,* J. Multivariate Anal. 91, no. 1, 18-34.

Cox, D. R. (1972). *Regression models and life-tables.* Journal of the Royal Statistical Society Series B-Statistical Methodology 34, 187-220.

Cox, Dm R. and Oakes, D. (1984). *Analysis of survival data*. London: Chapman and Hall.

Dafni, U. G. and Tsiatis, A. A. (1998). *Evaluating surrogate markers of clinical outcome measured with error*. Biometrics, 54, 1445-1462.

DeGruttola, V. and Tu, X.M. (1994). *Modeling progression of CD-4 lymphocyte count and its relationship to survival time*. Biometrics, 50, 1003-1014.

Diggle, P. J., P. Heagerty, K. Y. Liang, S. L. Zeger (2002). *Analysis of longitudinal data.* 2nd ed. Oxford University Press, 126-140, 214-216.

Faucett, C. J., and Thomas, D. C. (1996). *Simultaneously modeling censored survival data and repeatedly measured covariates: A Gibbs sampling approach*. Statistics in Medicine, 15, 1663-1685.

Follmann D, Wu M (1995). *An approximate generalized linear model with random effects for informative missing data.* Biometrics, 51, 151-168.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin D. B. (1995). *Bayesian data analysis*. London: Chapman and Hall.

Geman, S., and Geman, D. (1984). *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721-741.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, Chapman and Hall, New York.

Hastings, W. K. (1970). *Monte Carlo sampling methods using Markov Chains and their applications*. Biometrika, 57, 97-109.

Henderson, R., Diggle, P., and Dobson, A. (2000). *Joint modeling of longitudinal measurements and event time data. Biostatistics* 4, 465-480.

Kalbeisch J, Prentice R (2002). *The Statistical analysis of failure time data.* 2nd edition. John Wiley & Sons, New York.

Larsen K (2005). *The Cox proportional hazards model with a continuous latent variable measured by binary indicators.* Biometrics 61:1049-1055.

Little R, Rubin D (2002). *Statistical analysis with missing data.* 2nd edition. John Wiley & Sons, New York.

Raboud, J., Reid, N., Coates, R. A., and Farewell, V. T. (1993). *Estimating risks of progressing to AIDS when covariates are measured with error*. Journal of the Royal Statistical Society, Series A(156), 396-406.

Salah, k. A. (2008). *Joint modelling of longitudinal and survival data in presence of cure fraction with application to cancer patients data*. PhD Thesis.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). *Bayesian measures of model complexity and fit (with discussion).* Journal of the Royal Statistical Society B.

Tseng, Y.K., Hsieh, F., and Wang, J.L. (2005). *Joint modeling of accelerated failure time and longitudinal data.* Biometrika, 92, 587-603.

Tsiatis, A. A., DeGruttola, V. and Wulfsohn, M. S. (1995). *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, 27-37.

Tsiatis, A.A. and Davidian, M. (2004). *An overview of joint modeling of longitudinal and time-to-event data.* Statistica Sinica, 14, 793-818.

Wang, Y. and Taylor, J. M. G. (2001). *Jointly modeling longitudinal and event time data: Application in AIDS studies*. Journal of the American Statistical Association, 96, 895-905.

Wang, Y. and Taylor, J.M.G. (2001). *Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome*. *Journal of the American Statistical Association* 96, 895-905.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). *A joint model for survival and longitudinal data measured with error*. Biometrics, 53, 330-339.

Xu, J. and Zeger, S. L. (2001). *Joint analysis of longitudinal data comprising repeated measures and times to events*. Applied Statistics, 50, 375-387.

Yu M, Law N, Taylor J, Sandler H (2004). *Joint longitudinal-survival-cure models and their application to prostate cancer.* Statistica Sinica, 14, 835-832.

Zhao, G. (2008). *Nonparametric and parametric survival analysis of censored data with possible violation of method assumptions.* North Carolina at Greensboro University.

APPENDIX A

This is data generating function for simulation (chapter 4)

#This appendix gives R code for generating observations from the joint model. These observations can be used in simulation study
#to study the properties of the estimators (Run these codes just once).
The following packes were needed to run these codes and later codes
#1. "mcmc" 2. "JM" 3. "lcmm" 4. "actuar"

```
##############################################################################
###########                    Program One                    ###################
###########Generate Observed and Initial values for all parameters ###################
##############################################################################
y <- numeric() #measurement vector
id <- numeric() #subject id vector
ni <- numeric() #number of measurements for each subject
t <- numeric() #measurement time vector
u <- numeric() #uniform for CDF method
censor <- numeric() # vector of censoring times
m <- 100 # number of subjects
a <- numeric() # random intercepts vector
b <- numeric() #random slopes vector
mualpha <- 4; mubeta <- -4; sigmaa <- 1; sigmab <- 1; sigmae <- 0.01
k <- 1 # dummy variable
g <- -2 # gamma
c <- 0.00005 #baseline hazard
d <- numeric() #failure time vector
end <- numeric() # combined event time vector
f <- 0
### generate slopes, intercepts, event times, id vector
f <- 0
for(i in 1:m){
u[i] <- runif(1,0,1)
a[i] <- rnorm(1,mualpha,sqrt(sigmaa)) #create intercepts
b[i] <- rnorm(1,mubeta, sqrt(sigmab)) # create slopes
#create event times by CDF method
d[i] <- (log(1-g*b[i]*log(1-u[i])/(c*exp(g*a[i])))/(b[i]*g))
censor[i] <- rexp(1,0.01) #create censoring times
if(d[i]>censor[i]) f <- f+1
end[i] <- min(d[i],censor[i]) #combined event time
# determines number of measurements per subject
ni[i] <- rpois(1,10)
for(j in k:(k+ni[i]-1)){
id[j] <- i}
k <- k+ni[i]}
r <- 0
k <- 1
# create measurement vectors
for(i in 1:sum(ni)){
if(i>1){
if((id[i]-id[i-1])==1) {r <- 0
k <- k+1}}
```

```
t[i] <- rexp(1,10)+r
r <- t[i]
y[i] <- rnorm(1,a[k]+b[k]*t[i],sqrt(sigmae))}
# create censoring indicator vector
s <- numeric()
for(i in 1:m){
                    s[i] <- 0
                    if(end[i]==d[i]) s[i] <- 1}
d <- end


#We specify for each parameter a reasonable and fast method to obtain an initial value.

#The random slopes and intercepts alpha and beta

alphas <- numeric()
betas <- numeric()
for(i in 1:m){
alphas[i] <- lm(y[id==i]~t[id==i])$coefficients[1]
betas[i] <- lm(y[id==i]~t[id==i])$coefficients[2] }

#The means and variances of  mu-alpha, mu-beta, segmas-alpha, segmas-beta
mualpha <- mean(alphas)
mubeta <- mean(betas)
sigmaa <- var(alphas)
sigmab <- var(betas)

#The variance of the measurement error semas-e
sigmae <- mean((y-(alphas[id]+betas[id]*t))^2)

#The regression coeffecient gamma
test1 <- list(a=alphas,b=betas,death=d,censor=s)
q <- coxph(Surv(death,censor)~a+b,data=test1)
gamma <- mean(q$coef[1],q$coef[2])

#The baseline hazard c
a <- 0
for(i in 1:m){
a <- a+exp(gamma*alphas[i])*(exp(gamma*betas[i]*d[i])-1)/gamma/betas[i]}
c <- rgamma(1,sum(s),a)



##############################################################################
############                Program Two                      ###################
############          Generate Survival Data                 ###################
##############################################################################
y <- numeric() #measurement vector
id <- numeric() #subject id vector
ni <- numeric() #number of measurements for each subject
t <- numeric() #measurement time vector
u <- numeric() #uniform for CDF method
censor <- numeric() # vector of censoring times
m <- 100 # number of subjects
a <- numeric() # random intercepts vector
b <- numeric() #random slopes vector
```

```
mualpha <- 4; mubeta <- -4; sigmaa <- 1; sigmab <- 1; sigmae <- 0.01
k <- 1 # dummy variable
g <- -2 # gamma
c <- 0.00005 #baseline hazard
d <- numeric() #failure time vector
end <- numeric() # combined event time vector
f <- 0
### generate slopes, intercepts, event times, id vector
f <- 0
for(i in 1:m){
u[i] <- runif(1,0,1)
a[i] <- rnorm(1,mualpha,sqrt(sigmaa)) #create intercepts
b[i] <- rnorm(1,mubeta, sqrt(sigmab)) # create slopes
#create event times by CDF method
d[i] <- (log(1-g*b[i]*log(1-u[i])/(c*exp(g*a[i])))/(b[i]*g))
censor[i] <- rexp(1,0.01) #create censoring times
if(d[i]>censor[i]) f <- f+1
end[i] <- min(d[i],censor[i]) #combined event time
# determines number of measurements per subject
ni[i] <- rpois(1,10)
for(j in k:(k+ni[i]-1)){
id[j] <- i}
k <- k+ni[i]}
r <- 0
k <- 1
# create measurement vectors
for(i in 1:sum(ni)){
if(i>1){
if((id[i]-id[i-1])==1) {r <- 0
k <- k+1}}
t[i] <- rexp(1,10)+r
r <- t[i]
y[i] <- rnorm(1,a[k]+b[k]*t[i],sqrt(sigmae))}
# create censoring indicator vector
s <- numeric()
for(i in 1:m){
                    s[i] <- 0
                    if(end[i]==d[i]) s[i] <- 1}
d <- end



##############################################################################
############                Program Three                 ####################
############          Main Program for Simulation         ####################
##############################################################################
#We create a matrix in which to store the values of the parameters at each
#iteration, and then begin the loop in which the parameters will be updated.
#K represents the total number of iterations, and burn represents the number
#of iterations to discard in order to ensure convergence.

K <- 5000; burn <- 1000; m <- 100; alphasm <- numeric(K)
storage <- matrix(NA, K, 8)
for(j in 1:K)
{
        #Updating the model parameters
```

```
	for (i in 1:m)
	{
		alphaprop <- rnorm(1,alphas[i],0.1)
		a <- -c*exp(gamma*alphas[i])*(exp(gamma*betas[i]*d[i])-1)/gamma/betas[i]
		b <- -c*exp(gamma*alphaprop)*(exp(gamma*betas[i]*d[i])-1)/gamma/betas[i]
		q <- -(alphas[i]-mualpha)^2/2/sigmaa
		e <- -sum((y[id==i]-(alphas[i]+betas[i]*t[id==i]))^2/2/sigmae)-(alphas[i]-
mualpha)^2/2/sigmaa
		f <- -sum((y[id==i]-(alphaprop+betas[i]*t[id==i]))^2/2/sigmae) -(alphaprop-
mualpha)^2/2/sigmaa
		qr <- -(alphaprop-mualpha)^2/2/sigmaa
		U <- runif(1,0,1)
		if(log(U)<b+f+qr+gamma*s[i]*(alphaprop-alphas[i])-a-e-q)
		alphas[i] <- alphaprop
	}
	#Note that we take the logarithm of the ratio in order to avoid exponentiating,
	#thus improving numerical stability.

	#The random slopes
	for (i in 1:m)
	{
		betaprop <- rnorm(1,betas[i],0.1)
		a <- -c*exp(gamma*alphas[i])*(exp(gamma*betas[i]*d[i])-1)/gamma/betas[i]
		b <- -c*exp(gamma*alphas[i])*(exp(gamma*betaprop*d[i])-1)/gamma/betaprop
		q <- -(betas[i]-mubeta)^2/2/sigmab
		e <- -sum((y[id==i]-(alphas[i]+betas[i]*t[id==i]))^2/2/sigmae)-(betas[i]-
mubeta)^2/2/sigmab
		f <- -sum((y[id==i]-(alphas[i]+betaprop*t[id==i]))^2/2/sigmae)-(betaprop-
mubeta)^2/2/sigmab
		qr <- -(betaprop-mubeta)^2/2/sigmab
		U <- runif(1,0,1)
		if(log(U)<b+f+qr+gamma*s[i]*d[i]*(betaprop-betas[i])-a-e-q)
		betas[i] <- betaprop
	}
	# The marker means and variances
	mualpha <- rnorm(1,mean(alphas),sqrt(sigmaa/m))
	mubeta <- rnorm(1,mean(betas),sqrt(sigmab/m))
	sigmaa <- rinvgamma(1,m/2-1,sum((alphas-mualpha)^2)/2)
	sigmab <- rinvgamma(1,m/2-1,sum((betas-mubeta)^2)/2)
	#The regression parameter
	gammaprop <- rnorm(1,gamma,0.01)
	a <- numeric()
	b <- numeric()
	for(i in 1:m)
	{
		a[i] <- -c*exp(gamma*alphas[i])*(exp(gamma*betas[i]*d[i])-1)/gamma/betas[i]
		b[i] <- -c*exp(gammaprop*alphas[i])*(exp(gammaprop*betas[i]*d[i])-
1)/gammaprop/betas[i]
	}
	U <- runif(1,0,1)
	if(log(U)<(gammaprop-gamma)*sum(s*(alphas+betas*d))+sum(b)-sum(a))
	{
		gamma <- gammaprop
	}
	#The baseline hazard c
	a <- 0
```

```
        for(i in 1:m)
        {
                a <- a+exp(gamma*alphas[i])*(exp(gamma*betas[i]*d[i])-1)/gamma/betas[i]
        }
        c <- rgamma(1,sum(s),a)

        #The measurement error
        expect <- function(alphas,beta,t,id)
        {
                k <- numeric()
                for(i in 1:length(y))
                {
                        k[i] <- alphas[id[i]]+betas[id[i]]*t[i]
                }
                k
        }
        sigmae <- rinvgamma(1,length(y)/2-1, sum((y-expect(alphas,betas,t,id))^2)/2)

        storage[j, ] <- c(j, mualpha, mubeta, sigmaa, sigmab, gamma, c, sigmae)
}
```

APPENDIX B

This is the R code on real data (chapter 4)

```
##############################################################################
############                  Program One                    ###################
############          Plot survival and longitudinal data    ###################
##############################################################################
 # Load packages JM,  lattice and lcmm
# indicator for the composite event for the PBC dataset
pbc2$status2 <- as.numeric(pbc2$status != "alive")
pbc2.id$status2 <- as.numeric(pbc2.id$status != "alive")
pbcSurv <- survfit(Surv(years, status2) ~  drug , data = pbc2.id)
plot(pbcSurv, mark.time = FALSE, main = "Kaplan-Meier estimates of the probability of
survival for the PBC data",  xlab = "Time", ylab = "Survivl Probability" , col = c("black",
"red"), lty = 1:2)
legend("topright", c("Placebo", "D-penicil"), lty = 1:2, col = c("black", "red"),bty = "n")
xyplot(log(serBilir)~ year  | drug , group = id , data = pbc2, xlab = "Time", ylab =
"log(Serbilir)", col = 1, type = "l", main = "Subject-specic evolutions in time of the Log of
the Serbilir measurements, separately for Placebo and D-penicil.")


##############################################################################
############                  Program Two                    ###################
############                  Fit Separate Model             ###################
##############################################################################


sex_1 = factor(pbc2.id$sex, labels=c("male", "female"))
coxFit.pbc <- coxph(Surv(years, status2) ~ drug + log(age)+sex_1+serBilir,
   data = pbc2.id, x = TRUE)
baseline<-basehaz(coxFit.pbc)
summary(coxFit.pbc)
anova(coxFit.pbc)
summary(baseline)


##############################################################################
############                  Program Three                  ###################
############                  Fit Joint Model                ###################
##############################################################################
sex_1 = factor(pbc2.id$sex, labels=c("male", "female"))
lmeFit.pbc <- lme(log(serBilir) ~ year, random = ~1+ year  | id, data = pbc2)
coxFit.pbc <- coxph(Surv(years, status2) ~ drug + log(age)+sex_1,
   data = pbc2.id, x = TRUE)
jointFit <- jointModel(lmeFit.pbc, coxFit.pbc,
          method = "piecewise-PH-aGH")
summary(jointFit)
anova(jointFit)
baseline<-basehaz(jointFit)
summary(baseline)
```

APPENDIX C

Table of real data (The first 20 Patient Observations)

Case	days	Death	Tx	Age	Sex	Asc	Hepa	Spider	Edema
1	400	2	1	21464	1	1	1	1	1
2	4500	0	1	20617	1	0	1	1	0
3	1012	2	1	25594	0	0	0	0	0.5
4	1925	2	1	19994	1	0	1	1	0.5
5	1504	1	2	13918	1	0	1	1	0
6	2503	2	2	24201	1	0	1	0	0
7	1832	0	2	20284	1	0	1	0	0
8	2466	2	2	19379	1	0	0	0	0
9	2400	2	1	15526	1	0	0	1	0
10	51	2	2	25772	1	1	0	1	1
11	3762	2	2	19619	1	0	1	1	0
12	304	2	2	21600	1	0	0	1	0
13	3577	0	2	16688	1	0	0	0	0
14	1217	2	2	20535	0	1	1	0	1
15	3584	2	1	23612	1	0	0	0	0
16	3672	0	2	14772	1	0	0	0	0
17	769	2	2	19060	1	0	1	0	0
18	131	2	1	19698	1	0	1	1	1
19	4232	0	1	18102	1	0	1	0	0.5
20	1356	2	2	21898	1	0	1	0	0

Case	Bili	Chol	Albu	Urine	Alka	Sgot	Tri	Plate	Proth	Hist
1	14.5	261	2.6	156	1718	137.95	172	190	12.2	4
2	1.1	302	4.14	54	7394.8	113.52	88	221	10.6	3
3	1.4	176	3.48	210	516	96.1	55	151	12	4
4	1.8	244	2.54	64	6121.8	60.63	92	183	10.3	4
5	3.4	279	3.53	143	671	113.15	72	136	10.9	3
6	0.8	248	3.98	50	944	93	63	.	11	3
7	1	322	4.09	52	824	60.45	213	204	9.7	3
8	0.3	280	4	52	4651.2	28.38	189	373	11	3
9	3.2	562	3.08	79	2276	144.15	88	251	11	2
10	12.6	200	2.74	140	918	147.25	143	302	11.5	4
11	1.4	259	4.16	46	1104	79.05	79	258	12	4
12	3.6	236	3.52	94	591	82.15	95	71	13.6	4

13	0.7	281	3.85	40	1181	88.35	130	244	10.6	3
14	0.8	.	2.27	43	728	71	.	156	11	4
15	0.8	231	3.87	173	9009.8	127.71	96	295	11	3
16	0.7	204	3.66	28	685	72.85	58	198	10.8	3
17	2.7	274	3.15	159	1533	117.8	128	224	10.5	4
18	11.4	178	2.8	588	961	280.55	200	283	12.4	4
19	0.7	235	3.56	39	1881	93	123	209	11	3
20	5.1	374	3.51	140	1919	122.45	135	322	13	4

DATA: Primary Biliary Cirrhosis (PBC)

OBSERVATIONS: 418

VARIABLES: 20

VARIABLE DESCRIPTION:

Case

Case number

Days

The number of days between registration and the earlier of death, liver transplantation, or study analysis time in July, 1986.

Death

2 = death
1 = censored due to liver treatment
0 = “Days” is time to censoring

Tx

Treatment Code
1 = D-penicillamine
2 = placebo

Age

Age in years. For the first 312 cases, age was calculated by
dividing the number of days between birth and study registration by 365.

Sex

0 = male
1 = female

Asc

Presence of ascites
0 = no
1 = yes

Hepa

Presence of hepatomegaly
0 = no
1 = yes

Spider

Presence of spiders

0 = no
1 = Yes

Edema

Presence of edema
0 = no edema and no diuretic therapy for edema
0.5 = edema present for which no diuretic therapy was given, or edema resolved with diuretic therapy
1 = edema despite diuretic therapy

Bili

Serum bilirubin, in mg/dl

Chol

Serum cholesterol, in mg/dl

Albu

Albumin, in gm/dl

Urine

Urine copper, in mg/day

Alka

Alkaline phosphatase, in U/liter

Sgot

SGOT, in U/ml

Tri

Triglycerides, in mg/dl

Plate

Platelet count; coded value is number of platelets per-cubic-milliliter of blood divided by 1000.

Proth

Prothrombin time, in seconds

Hist

Histologic stage of disease, graded 1, 2, 3, or 4

Missing values are denoted by "." in the listing of the data.