Joint Modeling of Longitudinal Ordinal Data on Quality of life and Survival

by

Kemmawadee Preedalikit

A thesis submitted to Victoria University of Wellington in fulfilment of the requirements for the degree of Doctor of Philosophy in Applied Statistics.

Victoria University of Wellington 2012
Abstract

Joint models for longitudinal and survival data have been widely discussed in the literature. This thesis proposes a joint model using a stereotype model for the longitudinal ordinal responses and a Cox proportional hazards model for survival time. Our current joint model has a new feature since no literature has examined the joint model under the stereotype model. The stereotype model can improve the fit by adding extra score parameters, but it still has the advantage of requiring only a single parameter to describe the effect of a predictor on the item response levels. We give an example to model longitudinal ordinal data and survival data for patients being followed up after treatments. The main focus is on modeling both the quality of life data and the survival data simultaneously with a goal of understanding the association between the two processes over time. These two models are linked through a latent variable that characterizes the quality of life of an individual and is assumed to underlie the hazard rate. In other words, the latent variable serves as a shared variable in the joint model. We present the joint model in two different aspects: one based on a Bayesian approach and the other one a semiparametric approach using the EM algorithm. For the Bayesian approach, the latent variable is treated as a continuous variable and is assumed to have a multivariate
normal distribution. The partial survival likelihood function is used in the survival component of the Bayesian joint model, while the full likelihood function is considered in the semiparametric joint model. In the latter approach the baseline hazard is assumed to be a step function and has no parametric form. The latent variable in the semiparametric joint model is then treated as a discrete variable. We illustrate our methodologies by analyzing data from the Staccato study, a randomized trial to compare two treatment methods, for Human Immunodeficiency Virus (HIV) infection of Thai patients on Highly Active Antiretroviral Therapy (HAART), in which the quality of life was assessed with a HIV Medical Outcome Study (MOS-HIV) questionnaire. Furthermore, we extend the study further to the case of multiple failure types in the survival component. Thus, the extension of the joint model consists of the stereotype model and the competing risks model. The Bayesian method is employed to estimate all unknown parameters in this extended joint model. The results we obtained are consistent for both the Bayesian joint model and the semiparametric joint model. Both models show that patients who had a better quality of life were associated with a lower hazard of HIV progression. Patients on continuous treatment also had a lower hazard of HIV progression compared with patients on CD4-guided interruption treatment.
Acknowledgments

This thesis would not have been possible without the kind support and help of various individuals. It is an honour for me to give my sincere thanks to all of them.

I am heartily thankful to my primary supervisor, Dr Ivy Liu, whose encouragement, supervision and support from the preliminary to the concluding level, has enabled me to develop an understanding of the research. She always kindly grants me her precious time for answering some of my questions and supports me in various ways. Without her guidance and persistent help this thesis would not have been possible. Next, I would like to show my gratitude to Dr Nokuthaba Sibanda, my secondary supervisor, for her supervision, advice, and guidance from the very early stage of this research as well as giving me extraordinary experiences throughout the work. Thank you for supporting me throughout my study. I also gratefully acknowledge Dr Yuichi Hirose, my co-supervisor, for his advice, crucial contribution and his expertise in semiparametric approach. Many thanks to Assoc Prof Megan Clark for kindly assistance since I first came and during my study.
I gratefully thank the Phayao University, Thailand for awarding me with the scholarship and special thanks to all people from the Mathematics Statistics and Operations Research (MSOR) programme, Victoria University for giving me the opportunity to be teaching assistance. These people have supported my financial affairs very efficiently during my three years PhD here.

I also would like to make a special reference to Dr Jintanat Ananworanich, deputy director in Scientific Affairs at HIV-NAT, for providing the dataset of the Staccato study. My gratitude to Assoc Prof Chulapan Engchanil and Burabha Pussadhamma, MD, for their help in medical area. Without their corporation I could not have gotten such relevant data.

I am particularly grateful for the assistance given by Liz O’Connor, learning advisor, for revising the English of my manuscript.

I would like to express my very great appreciation to three examiners, Dr Richard Arnold, Professor Alastair Scott and Professor David Collett, for their encouragement and their critical comments about it.

Finally, I owe my deepest gratitude to my parents who have given me the opportunity of an education from the best institutions and support throughout my life. Thanks to my beloved sister for being a supportive and caring sibling and all my friends who have always helped me and believed that I could do it.

To them I dedicate this thesis.
# Contents

1 Introduction 1
   1.1 Quality of Life Analysis ........................................... 5
   1.2 Survival Analysis ..................................................... 10
   1.3 Joint Model .......................................................... 13
   1.4 Outline of the thesis ............................................... 19

2 Stereotype Model 23
   2.1 Introduction ......................................................... 23
   2.2 The Stereotype Model ................................................. 25
      2.2.1 The Stereotype Model with Fixed Scores ...................... 29
      2.2.2 The Ordered Stereotype Model ................................. 34
   2.3 Model Interpretation ............................................... 36
   2.4 Model Fitting ....................................................... 38
   2.5 Multi-item Stereotype Model ...................................... 43

3 A Bayesian Joint Model 47
   3.1 Introduction ........................................................ 47
   3.2 Model Structure .................................................... 52
      3.2.1 Item Response Models ....................................... 54
      3.2.2 The Cox Proportional Hazards Model ....................... 59
      3.2.3 The Regression Model .................................... 64
      3.2.4 The Full Likelihood Function ............................... 66
   3.3 Parameter Estimation ............................................... 68
3.4 Convergence Diagnostics ..................................... 80
  3.4.1 Visual Inspection ........................................ 81
  3.4.2 Statistical Diagnostic Tests ............................... 82
3.5 Bayesian Model Comparison ................................. 86
  3.5.1 The Bayesian Information Criterion (BIC) ............. 86
  3.5.2 The Deviance Information Criterion (DIC) ............. 87

4 A Semiparametric Joint Model ................................. 91
  4.1 Introduction ................................................. 91
  4.2 Finite Mixture Models ....................................... 93
  4.3 Semiparametric Model ....................................... 101
  4.4 Model Structure ............................................. 106
    4.4.1 The Stereotype Model .................................. 107
    4.4.2 The Semiparametric Proportional Hazards Model .... 110
    4.4.3 The Full Likelihood Function .......................... 112
  4.5 Parameter Estimation ....................................... 114
    4.5.1 Baseline Hazard Estimation ............................ 115
    4.5.2 The E-step ............................................ 117
    4.5.3 The M-step ........................................... 118
  4.6 Model Comparison ........................................... 121

5 Data Application ................................................ 123
  5.1 Introduction ................................................. 123
  5.2 Data Analysis : Bayesian Joint Model ..................... 129
    5.2.1 Convergence Diagnostics ............................... 132
    5.2.2 Bayesian Model Selection .............................. 143
    5.2.3 Posterior Summary .................................... 144
  5.3 Data Analysis : Semiparametric Joint Model ............. 153
    5.3.1 Model Comparison ..................................... 153
    5.3.2 Estimated parameters ................................. 155
## CONTENTS

6 The Extension of the Joint Model
   6.1 Introduction .................................................. 161
   6.2 Model Structure ............................................... 163
      6.2.1 The Competing Risks Model ......................... 164
      6.2.2 The Full Likelihood Function ..................... 171
   6.3 Parameter estimation ....................................... 172
   6.4 Application to Pseudo Data ............................... 175

7 Conclusion ......................................................... 181

A Convergence Diagnostics and Posterior Summary ............... 189

B EM procedure and Results ................................... 223

C MOS-HIV questionnaire ....................................... 229
Chapter 1

Introduction

Many biomedical and other follow-up studies are performed to collect repeated measurements of a response variable at multiple time points, as well as to record time to an event of interest, and additional covariate information. \textit{Time to an event of interest} can refer to time to death and can also include other common events of interest such as time to recurrence of symptoms, time to infection of a disease, etc. In a longitudinal study of quality of life in which there is also an endpoint of interest, one may be interested in finding any possible relationship between the quality of life and the survival time of subjects. Quality of life may be one of the most important factors that could be used to predict survival. It is also worth identifying factors that jointly affect survival and quality of life. Various approaches have been proposed in the literature for handling quality of life and survival data separately (e.g. Guo and Carlin, 2004). However, in the case when data regarding quality of life and survival are collected
CHAPTER 1. INTRODUCTION

from the same subject, it is not ideal to perform a separate analysis on each set of data. We need to consider the dependency between these two outcomes. Therefore, there are potential gains in power in considering the quality of life and survival time as two simultaneous processes. That is, the change in quality of life over time and the time to an event of interest can be modeled in a joint model in which parameters can be estimated simultaneously.

An example we use to illustrate the joint model in this thesis is data from the Staccato study. Staccato is a multicentre trial carried out by a research collaboration from Switzerland, Australia and Thailand (Ananworanich and the Staccato study group, 2006). Their study was carried out for an average of approximately three years from January 2002 until November 2005. The majority of participants were Thai patients (about 80%) with Human Immunodeficiency Virus (HIV) infection and the remaining HIV patients came from Switzerland and Australia. The Staccato study is a randomized trial of continuous anti-retroviral treatment compared to CD4 guided interruption treatment. Patients on Highly Active Antiretroviral Therapy (HAART), with viremia below 50 copies/ml and CD4 cell count above 350 cells/µl were randomized into one of two different treatment strategies. They also had to have been on a stable HAART regimen for at least three months and have no evidence of pre-existing drug resistance. Note that CD4 cell count is a measure of the number of immune system cells, such as T-lymphocytes and monocytes, that have
CD4 receptors (Carter, 2009). HIV causes gradual CD4 decline over several years.

Data on demographics, medical history, clinical events and quality of life were obtained from 548 patients (439 in Thailand, 100 in Switzerland and nine in Australia). Only Thai patients were asked to complete the Medical Outcome Study HIV Health Survey (MOS-HIV) questionnaire (Wu et al., 1991, 1997) to evaluate the quality of life every 24 weeks beginning at the baseline visit and at weeks 24, 48, 72, 96, 120 and 144 of follow-up. Demographic information such as age, sex and date of HIV diagnosis was also collected. Patients were assessed for CD4 count, viral load, complete blood cell counts, clinical chemistry, clinical and laboratory adverse events, and HIV disease progression every four weeks over the first 12 weeks, and every 12 weeks thereafter. Patients were classified as being symptomatic if they manifested any symptoms listed in the classification system for HIV infection of the U.S. Center for Disease Control (CDC) (Centers for Disease Control and Prevention, 1992).

The study objective in the Staccato trial was to evaluate a strategy of CD4 guided interruption treatment, in comparison to continuous treatment. Evaluation of quality of life was also presented in the Staccato study using the data at baseline, week 24, week 48 and at the last visit. The study may have been designed to address the primary question of the treatment effect, but another interest could be to understand the association between the quality of life and survival time or how the patients’
quality of life would affect their survival. Quality of life assessment is one of the most important indicators for disease progression or any other endpoints in clinical trials because the quality of life may be predictive of survival. Therefore, to jointly summarize quality of life and survival (or time to an event of interest), we propose a joint model by combining an item response model and a survival model together. The quality of life data, in contrast, were analyzed separately from survival in the Staccato study (Nüesch et al., 2009). The main focus of this thesis is on modeling both the quality of life data and the survival data simultaneously with a goal of understanding the association between the two processes over time.

The item response model is used to model the quality of life data which were measured by the MOS-HIV questionnaire, whereas the Cox proportional hazards model is used to model time of progression to a symptom listed in CDC which is CDC-B (more details about CDC classification will be given in Chapter 5). These two models are linked through a latent variable that characterizes the quality of life of an individual. In other words, the longitudinal ordinal data (quality of life) and the survival data (time to CDC-B progression) will be jointly modeled across seven time points (baseline visit, 24, 48, 72, 96, 120 and 144 weeks).

The survival model described above focuses on only one endpoint event of interest. We extend the study further to the case of multiple failure types known as the competing risks model. The extension of the joint model then consists of the item response model and the competing risks model.
1.1. QUALITY OF LIFE ANALYSIS

This extension will illustrate the joint model approach in other aspects and will also provide an opportunity to establish if the effects of each covariate are different for the different failure types.

1.1 Quality of Life Analysis

The Constitution of the World Health Organization (WHO) defines quality of life as “individuals’ perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns”. Quality of life refers to the degree of excellence in an individual’s life at any given period of time and can also change over time as our living situations change. Quality of life measurements are important for assessing the impact on patients of chronic disease and its treatment.

Quality of life is usually measured using an instrument in the form of a questionnaire designed for patient completion. The questionnaire is generally composed of a set of questions or items relating to various dimensions of quality of life, such as mental, physical, social, etc. The format of responses to each item on a questionnaire may yield the binary categories “Yes/No” or frequently yields a series of ordered categories (e.g. “Better/Unchanged/Worse”). Furthermore, quality of life data are generally longitudinal in nature with a questionnaire administered at a baseline visit and subsequent repeated measurements at each follow-up time point. However, one of the main problems in analyzing longitudinal qual-
ity of life data is caused by missing data because patients might drop out or withdraw before the end of the study. This missing data problem is a common occurrence in a quality of life study where survival is also an issue.

Quality of life data are often summarized into a categorical scale that is either nominal or ordinal in nature. The standard methods for analyzing nominal data such as logistic regression models and the Pearson’s chi-squared test are adequate for providing results. Moreover, traditional analytical methods for quality of life data sometimes treat the ordinal scale as continuous by assigning scores to the ordinal outcomes (e.g. scores vary from 0 to 100, 100 indicating excellent quality of life). Likewise, the MOS-HIV questionnaire in the Staccato study was divided into 10 subscores ranging from 0 to 100 where greater values implied a better quality of life. The summary scores were transformed to standardized z-scores, with a mean of 50 and a standard deviation of 10. The results suggested that the mean quality of life score at baseline was significantly lower in the CD4 guided interruption treatment compared to the continuous treatment.

We know that quality of life assessments are measured principally using questionnaires in the form of an ordinal scale. Ordinal logistic regression models, such as the proportional odds model and the partial proportional odds model, are also well known for analyzing ordinal data (McCullagh, 1980). Among other models, the most popular and commonly used model for ordinal responses is the proportional odds model which is
also known as the cumulative logit model. The model can be written in logit form as

\[
\text{logit}[P(Y \leq j|X)] = \alpha_j - \beta'X, \quad j = 1, \ldots, J - 1
\]

This model is for a \( J \)-category ordinal response \( Y \) and \( X \) denotes a set of predictors with corresponding effect parameters \( \beta \). Note that the negative sign in the predictor term makes the sign of each component of \( X \) have the usual interpretation in terms of whether the effect is positive or negative. In the sense that if \( \beta > 0 \), higher levels of \( Y \) are more likely for higher values of \( X \). The parameters \( \{\alpha_j\} \) are the unknown intercept parameters and called the cut points. Furthermore, this model compares the probability of a response less than or equal to a given category \((j = 1, \ldots, J - 1)\) to the probability of a response greater than this category. To fit the model, maximum likelihood is the preferred method (Liu and Agresti, 2005). The weighted least squares approach is also used in some early applications of cumulative logit models (e.g. Williams and Grizzle, 1972) but this approach is designed for nonsparse contingency tables and cannot handle continuous predictors.

The proportional odds model has the same effects \( \beta \) for each category \( j \). In other words, it provides a single estimate of the log odds ratio over the cut points. McCullagh (1980) described this characteristic of the model as the proportional odds assumption, hence the name proportional odds model. Therefore, in the process of constructing the model, it is important to ver-
ify whether the proportional odds assumption is satisfied. Peterson and Harrell (1990) suggested that the score test can be used for testing this assumption but the test may perform poorly for sparse data. In cases where the proportional odds structure may be inadequate for model fit, alternative models are considered. Liu and Agresti (2005) concluded the strategies to improve the fit include (1) trying different link functions such as the log-log; (2) adding additional terms, such as interaction, to the linear predictor; (3) generalizing the model by adding dispersion parameters; (4) permitting separate effects for each logit for some but not all predictors; (5) using the ordinary model for a nominal response.

Although we can assume the response categories are ordered, we do not know the structure of this ordering with respect to a given explanatory variable. McKelveya and Zavoinab (1975), Winship and Mare (1984) and Scott (1997) stated that we never know the true interval between the adjacent response categories, even though the response categories are explicitly assumed to be equally spaced. Thus, in this thesis we generalize the stereotype model (Anderson, 1984) for item responses to solve such complicated measured scale problems and to capture the natural ordering of the data.

The stereotype model can be thought of as a multinomial logistic model constrained to respect the ordinal nature of the response categories (Lunt, 2005). Another motivation for the development of the stereotype model is to relax the strong assumption in the proportional odds model which
1.1. QUALITY OF LIFE ANALYSIS

assumes that each predictor variable has the same effect at every level of
the response variable. The proportional odds assumption in Agresti (2010,
ch.3) states that the log cumulative odds ratio for any two values of the
covariates is constant across response categories. The stereotype model,
however, has more flexibility when analyzing an ordinal response since
the proportional odds assumption is no longer used.

Moreover, the stereotype model has the ability to reduce the number
of parameters by imposing constraints, without reducing the adequacy
of the fit (Agresti, 2010, ch.4). This model could also provide the distin-
guishability of the adjacent response categories and estimate how close
adjacent response categories are, whereas the proportional odds model
cannot evaluate this information. For this reason, the stereotype model
can be used in more complicated situations.

Although the stereotype model has many advantages and some attrac-
tive features, the model has not been widely used. This is probably be-
cause of the complication of the model structure due to multiplicative pa-
rameters and this nonlinearity causes computational challenges for model
fitting. The stereotype model in the literature has been fit through the
procedure of maximum likelihood. However, Ahn et al. (2009) proposed
a Bayesian approach to estimate the parameters in the stereotype model.
This showed a new contribution to the literature. Similarly, in this thesis
the Bayesian method is employed for parameter estimation. In addition,
the contribution of this study is that the stereotype model is accommo-
dated in the joint model. This is a new application since no literature has examined the stereotype model within a joint model. More details on the stereotype model will be given in the next chapter.

1.2 Survival Analysis

For the survival component, we consider the Cox proportional hazards model (Cox, 1972) to model time to an event of interest. The purpose of the model is to explore the relationship between the survival of patients and several explanatory variables. The Cox proportional hazards model for survival time \( t \) is usually written in the form of the hazard model shown below

\[
\begin{align*}
  h(t|X, \beta) &= h_0(t) \exp(X'\beta) \\
  (1.1)
\end{align*}
\]

This model provides the expression for the hazard at time \( t \) for an individual given a set of explanatory variables denoted by \( X \). That is, \( X \) presents the predictor variables which are being modeled to predict the hazard of an individual. The hazard function above is the product of two functions. The first one \( h_0(t) \), is called the baseline hazard function. The second one is the exponential of the linear term \( X'\beta \). The former function, \( h_0(t) \), characterizes how the hazard function changes as a function of survival time and the latter function, \( \exp(X'\beta) \), characterizes how the hazard function changes as a function of subject covariates. It can be seen that, the base-
1.2. **SURVIVAL ANALYSIS**

line hazard function is the function of \( t \) but does not involve \( X \). On the other hand, the exponential expression involves \( X \) but does not involve \( t \). Therefore, if all covariates are set to zero (\( \exp(0) = 1 \)), the formula is reduced to the baseline hazard function. This is the reason why \( h_0(t) \) is frequently referred to as the baseline hazard function. Another important point of the Cox proportional hazards model is that a parametric form of the baseline hazard function, \( h_0(t) \), does not have to be specified. Thus, the Cox proportional hazards model is considered a semiparametric model. The semiparametric nature is one of the reasons that makes the Cox proportional hazards model popular and widely used. The main reason for the popularity of the model is that, although the baseline hazard function is unspecified, the covariate effects can still be estimated reasonably from the exponential part of the model.

The Cox proportional hazards model parameters are usually derived by maximizing a likelihood function. The formula for such a likelihood function is called a *partial likelihood function* when the likelihood depends only on the parameter of interest and does not involve the baseline hazard. The partial likelihood function can be expressed as

\[
L(\beta | X) = \prod_{i=1}^{n} \left( \frac{\exp(X_i' \beta)}{\sum_{p \in R(t_i)} \exp(X_p' \beta)} \right)^{d_i}
\]

(1.2)

where \( d_i \) is an event indicator, if individual failed, \( d_i = 1 \), otherwise \( d_i = 0 \). \( R(t_i) \) is the set of individuals who are alive and uncensored at time \( t_i \) which is referred to as the risk set. Note that censored refers to those
subjects in whom the event of interest is not observed, possibly due to loss of follow-up, dropout of subjects or early termination of the study. As can be seen from the partial likelihood function (1.2), the baseline hazard, \( h_0(t) \), cancels out from the likelihood and does not play any role in the estimation. Thus, only the coefficient \( \beta \) can be estimated from this partial likelihood function.

Alternatively, we can derive a full likelihood function for \( h_0(t) \) and \( \beta \) in the Cox proportional hazards model (1.1). This allows us to derive the estimator for the baseline hazard \( h_0(t) \). The estimator may be used when combined with the estimated coefficients and specific covariate values to estimate the survival probabilities or the hazard rate of particular interest subjects. However, the baseline hazard is not easily calculated since it is a function of time. Fleming and Harrington (1991) and Andersen et al. (1993) demonstrated the counting process approach to derive the estimator. Recently, Ren and Zhou (2011) presented the empirical likelihood parameterization to obtain the full-profile likelihood function.

In this thesis, we present both the partial likelihood function and the full likelihood function by using different approaches for estimating parameters. The partial likelihood function is included in the joint model and estimation is done by using the Bayesian approach. The full likelihood function is included in another joint model and the parameters are estimated simultaneously based on the Expectation-Maximization (EM) algorithm (Dempster et al., 1977; McLachlan and Krishnan, 1997). Throughout
1.3. JOINT MODEL

In this thesis, we will call the joint model that is evaluated by the Bayesian approach as the Bayesian joint model. Another joint model implemented via the EM algorithm is called the semiparametric joint model since this approach is a semiparametric method with an unspecified baseline hazard function. That is, the baseline hazard is assumed to be a step function and has no parametric form.

1.3 Joint Model

We are interested in simultaneously investigating the relationship between quality of life and time to CDC-B progression as well as identification of factors that affect both variables. In order to do this, a joint model of longitudinal ordinal data and survival data is required. There is extensive literature on the joint model from numerous authors. Typically a mixed effects model and a Cox model or an accelerated failure time model have been considered to jointly model longitudinal data and survival data with shared random effects. The likelihood method based on the EM algorithm (Dempster et al., 1977; McLachlan and Krishnan, 1997) was often used with the joint model in literature, for instance, Tseng et al. (2005), Hsieh et al. (2006) and Ding and Wang (2008). They proposed the same concept of the joint modeling, that is to jointly model the survival time and its longitudinal covariates.

However, the main challenge is computation since the likelihood function can be complicated. Bayesian methods have also received much at-
CHAPTER 1. INTRODUCTION

tention. For example, Chi and Ibrahim (2007) presented the joint model of quality of life and breast cancer progression. In their research, the multidimensional longitudinal quality of life measurements were modeled using a hierarchical mixed effects model while disease-free survival was modeled by a parametric survival model. The quality of life data in their research were transformed into a scale of 0 to 10 with smaller values reflecting a better quality of life whereas we use the raw scores from the questionnaire to examine the stereotype model. Furthermore, Hu et al. (2009) illustrated use of Bayesian methods in joint analysis of longitudinal measurements and competing risks failure time data to compare the effectiveness between two treatments of lung disease. They also suggested that the Bayesian method can avoid high-dimensional integration and computational implementation.

Moreover, the joint model is becoming increasingly essential in AIDS studies. The early development of the joint model in AIDS studies was primarily motivated by characterizing the relationship between a longitudinal biologic marker such as CD4 count or immune response and the time to HIV/AIDS progression. For example, Huang et al. (2011) demonstrated the joint modeling of three components (response, covariates, and time-to-event processes) linked through random effects. The model was applied to AIDS clinical studies that explored the relationship between HIV dynamics and the time to decrease in CD4/CD8 ratio in the presence of CD4 process with measurement errors. The methodology they used is
the joint model with a skew-normal distribution which is the first attempt at working on such a distribution. Other references in the literature to joint models in AIDS studies include Lange et al. (1992), DeGruttola and Tu (1994), Hoover et al. (1992) and Tsiatis et al. (1995).

Although the longitudinal biomarker (CD4 or immune response) can be used as an indicator of a particular disease state, other factors such as quality of life may also be an important predictor of survival. Thus, we will alternatively perform the joint model in another application. That is, our joint model is motivated by characterizing the relationship between the longitudinal quality of life and time to HIV progression where the biomarkers are treated as covariates in the survival component. This is because living with HIV can impact on the quality of life not only in physical health but also in mental and social well-being. This could be the result of the comprehensive treatment strategies, various and consecutive care methods, and home care (Handford et al., 2006). A study of Saunders and Burgoyne (2002) found that patients with better immunologic/virologic outcomes showed slight improvements in the mean of quality of life, whereas those with poorer clinical outcomes showed a slight deterioration. Liu et al. (2006) stated that the quality of life of participants changed over the HIV disease course and HAART enhanced mental health functioning. In addition, many HIV/Acquired Immune Deficiency Syndrome (AIDS) patients struggle with social problems and cultural beliefs which can affect their quality of life (Aranda-Naranjo, 2004).
HIV has a significant impact on quality of life; hence the evaluation of quality of life has become an important issue in the study of HIV. Often in clinical trials where the primary endpoint is time to an event, patients are monitored longitudinally by a questionnaire to assess the quality of life after receiving a particular treatment. Although these longitudinal measures are frequently incomplete due to disease progression or treatment including death, these measurements are also important because they may be predictive of survival. The role of quality of life will become more prominent in the future especially in many chronic diseases. The improvements in survival depend not only on the given treatment, but are also associated with the quality of life. Therefore, methods that can model both the longitudinal quality of life and survival component jointly are very useful and essential in HIV clinical trials.

In the case of the joint model, the item response model and the survival model are usually linked through a latent variable. That is, the latent variable serves as the shared variable or shared parameter in the joint model. The latent variable is treated as a continuous variable and is assumed to have a multivariate normal distribution in the Bayesian joint model, whereas in the semiparametric joint model the latent variable is treated as a discrete variable. Consequently, the latent variable in this thesis is assumed to express all information on the quality of life of subjects. We will give a brief overview about the latent variable below.

There is a significant amount of information in medical studies that
cannot be observed directly but does have a measurable impact on observable characteristics, for example, a person’s intelligence or the strength of an attitude, behavioral intentions, quality of life and personality abilities. The information can only be measured indirectly by means of observable indicators such as a questionnaire. The responses on a questionnaire from an individual may be in the form of the level of agreement, health status or patient symptom scores. Various types of scaling techniques have been developed for deriving information on an unobservable variable from the indicators. The most widely used scaling method is a latent variable model.

A further use of the latent variable is to express the distribution of the observed variables \((X_1, X_2, \ldots, X_m)\) in terms of a smaller number of latent variables \((\theta_1, \theta_2, \ldots, \theta_n)\) where \(n < m\). The concern is to reduce the number of variables with as little loss of information as possible. Generally, a questionnaire contains a set of questions (e.g. 50 or 70 questions) and the number of subjects who complete the questionnaire may be 1000 respondents. This is because the more questions we ask, the more accurate results we might gain. If this is the case, we will have many variables to analyze and it is too difficult to see any pattern in their interrelationship. This problem can be solved by use of the latent variable model since this latent variable is designed to collect all the information in the entire set of questions into much fewer variables. Moustaki (2000) and Bartholomew (1987) found that the number of latent variables typically is much smaller than the number of questions.
The basic assumptions and objectives of latent variable modeling can be summarized as follows (Bartholomew et al., 2002):

- A small set of latent variables is assumed to explain the interrelationship in a set of observed response variables.

- Unobserved variables such as quality of life, intelligence, mathematical or verbal ability, racial prejudice, political attitude, consumer preferences, which cannot be measured by conventional means, can be quantified by assuming latent variables.

- Latent variable modeling is used to assign scores to sample units in the latent dimensions based on their questionnaire responses. This score is a numerical value that indicates an individual’s position on the latent variable scale. This score can also be used to classify subjects into groups by their characteristics.

The joint model not only allows investigation of both quality of life and survival components, but also incorporates all information simultaneously with valid and efficient inference. Hogan and Laird (1997), Xu and Zeger (2001) and Hsu et al. (2006) among others used the joint model to make more efficient inference on the survival model by incorporating the longitudinal data as auxiliary information. In addition, Tsiatis and Davidian (2004) reviewed a brief overview of earlier work on joint models. Recently, Wu et al. (2012) conducted a simulation study on joint modeling of longitudinal and survival data, and illustrated that
Bayesian methods have a similar performance to likelihood methods. Furthermore, Tsiatis and Davidian (2001), Song et al. (2002) and Song and Wang (2008) proposed the joint model in a semiparametric perspective. Their approaches are semiparametric in that there is no parametric density function assumed for random effects.

Throughout this study, our joint model is composed of the stereotype model and the Cox proportional hazards model. The stereotype model is used to model the quality of life data, whereas the Cox proportional hazards model is used to model the survival data. These two models share some latent variables together. We consider two approaches: a Bayesian method and a semiparametric method via an EM algorithm to obtain the parameter estimation. Our current joint model has a new feature since there is no literature on the joint model under the stereotype model. However, our joint model is highly complicated due to unobserved variables, censoring, missing data, semiparametric and nonlinear structure of the model. Therefore, a major difficulty in a joint likelihood is computation in which all parameters need to be estimated simultaneously.

1.4 Outline of the thesis

The proposed methods for analyzing the longitudinal ordinal responses and the survival data are discussed in this thesis. The methodologies and data applications are separately presented in different chapters.

Chapter 2 is a review of the stereotype model. Among other ordinal lo-
gistic regression models, the stereotype model is more parsimonious and more flexible for analyzing the ordered categorical data. Anderson (1984) indicated that this model is particularly appropriate when the categories are ordered. This chapter begins by reviewing the structure of the stereotype model that uses the reduced-rank regression idea to reduce the number of parameters. Yee and Hastie (2003) called it the reduced-rank multinomial logit model. The stereotype models with fixed scores and ordered scores are further described respectively. We review the methodologies to fit the model that have been proposed in the literature. Model interpretation is also described.

Chapter 3 presents model structure and estimation methods for the first approach which is the Bayesian joint model. The Bayesian joint model is composed of the stereotype model and the Cox proportional hazards model which are linked through a continuous latent variable. Because the questions in the MOS-HIV questionnaire are different in the number of response categories, we use the stereotype model to model the questions that have ordered responses and use the binary model to model the questions that have binary responses. The partial survival likelihood function is used in the survival component and the latent variable is modeled using a linear regression model. Since the Bayesian approach is used to evaluate the joint posterior density function, convergence diagnostic procedures are given using both visual inspection and statistical diagnostic tests. We construct many models so Bayesian model comparison is also considered.
Chapter 4 presents model structure and estimation methods for the second approach which is the semiparametric joint model. The semiparametric joint model consists of the stereotype model and the Cox proportional hazards model which are linked through a discrete latent variable. The unspecified baseline hazard function is also included in the joint model. EM algorithm is used to estimate the model parameters by iterative computation between the expectation step and the maximization step. The baseline hazard is iteratively estimated together with other parameters in the EM steps by its nonparametric maximum likelihood. The profile Akaike information: pAIC (Xu et al., 2006) is used for model comparison.

Chapter 5 illustrates the applications of two joint models: the Bayesian joint model and the semiparametric joint model. The methods are applied to data from the Staccato study.

Chapter 6 presents the extension of the joint model in the case of multiple failure types in the survival component. The model is called the competing risks model when multiple failure types are considered. Therefore, the joint model in this chapter is composed of the stereotype model and the competing risks model which are linked through a continuous latent variable. The methods have a similarity to the Bayesian joint model in Chapter 3 since other structures of the model are similar and only the survival component is different. In this chapter we present the methodology and give an example by using pseudo data.

The last chapter (Chapter 7) is the conclusion and discussion.
Chapter 2

Stereotype Model

2.1 Introduction

Anderson (1984) proposed a multinomial logit model where the reduced-rank regression idea was applied to non-Gaussian errors and referred to as the stereotype model. He also indicated that this model is particularly useful when the categories are ordered.

The reduced-rank regression was first introduced by Anderson (1951). It is a multivariate linear regression method where the estimated matrix of regression coefficients is of reduced rank and where several response variables are related to the same set of explanatory variables.

The reduced-rank regression is suitable for situations where the response variables are related to the predictor variables in a similar way. The method can be useful when there are many responses and predictor variables. However, it has been restricted to data where the response variable
CHAPTER 2. STEREOTYPE MODEL

is continuous and its use outside the Gaussian family is not mentioned (Reinsel and Velu, 1998). This might be the main reason why reduced-rank regression has only been used in a few applications.

**Reduced-Rank Regression Model**: we consider the multivariate linear regression of $Y$ on $X$ of dimension $q$ and $p$ respectively. The multivariate regression model is

$$Y = BX + \epsilon$$

where $Y$ is a column vector of $q$ response variables, $X$ is $p$-dimensional column vector of the explanatory variables and $\epsilon$ is a $q$-dimensional column vector of errors, $\epsilon \sim N(0, \Sigma)$. Thus, $B$ is the regression coefficient matrix of dimension $q \times p$.

Commonly the regression coefficient matrix $B$ is estimated by ordinary least squares method and so far we have not considered the rank of true $B$ but implicitly assumed that the resulting estimate $\hat{B}$ will have a full rank, $m = \min(p,q)$. However, Izenman (1975) introduced the term ‘reduced-rank regression’ where in some situations the regression coefficient matrix may not have a full rank. The hypothesis that $B$ has reduced rank less than $m$ is expressed as $B = \phi \beta'$ where $\phi$ is $q \times r$ matrix and $\beta$ is $p \times r$ matrix, both having rank $r$ and $r < m$. This gives the reduced-rank model as

$$Y = \phi \beta' X + \epsilon$$
2.2. THE STEREOTYPE MODEL

The reduced-rank $r$ must be estimated from the data if it is not know from the context (Aldrin, 2002).

2.2 The Stereotype Model

When the response variable is a categorical variable and consists of more than two discrete outcomes, the multinomial logistic regression model is used. In this case, the explanatory variable can be a continuous variable, a categorical variable or both. The multinomial logistic regression model is of the form

$$P(Y = \ell | X) = \frac{\exp(a_\ell + B_\ell'X)}{1 + \sum_{k=2}^{L} \exp(a_k + B_k'X)}$$  \hspace{1cm} (2.1)

where the response variable $Y$ categories are $\ell = 1, 2, \ldots, L$ and category $\ell = 1$ is treated as the baseline category to make the model identifiable, i.e. $a_1 = 0$ and $B_1 = 0$. The $B_\ell$ is the regression coefficient vector or the covariate effects which has $L - 1$ parameters for each predictor $X_i, i = 1, 2, \ldots, p$. The logit model can be represented as

$$\log \left[ \frac{P(Y = \ell | X)}{P(Y = 1 | X)} \right] = a_\ell + B_\ell'X, \quad \ell = 2, \ldots, L$$

This model is appropriate when the responses are unordered categories. However, if the response is ordinal, this model has $L - 1$ parameters for
each predictor $X_i$. In some research, if $L$ and $p$ are large, then the total number of regression coefficients in the multinomial logistic regression model would also be large. This would be difficult to interpret. Thus, it is convenient to use the reduced-rank regression idea to reduce the number of parameters. Similarly, Lunt (2005) suggested that if we assume that the regression function is the same for all categories, we can consider the stereotype model as the non-linear form of the constrained multinomial model. This model compares each category of the response variable with the reference category.

If this is the case, the multinomial logistic regression model (2.1) can be simplified to

$$P(Y = \ell | X) = \frac{\exp(a_\ell + \phi_\ell \beta' X)}{1 + \sum_{k=2}^{L} \exp(a_k + \phi_k \beta' X)}$$  \hspace{1cm} (2.2)

where $\phi_\ell$ are scalars, $\ell = 1, \ldots, L$ and $a_1 = \phi_1 = 0$. Also, this model can be defined using logits for pairs of categories as

$$\log \left[ \frac{P(Y = \ell | X)}{P(Y = 1 | X)} \right] = a_\ell + \phi_\ell \beta' X, \quad \ell = 2, \ldots, L$$

In this model, category $\ell = 1$ is treated as the baseline category, $a_\ell$ is the item intercept parameter, $\beta$ is the covariate effect and $\phi_\ell$ is the score of the response categories and $\phi_\ell \beta_i$ represents the log odds for response $Y = \ell$ versus $Y = 1$ per unit increase in $X_i$.

Model (2.2) is referred to as the reduced-rank multinomial logit model (Yee
2.2. THE STEREOTYPE MODEL

and Hastie, 2003). Anderson (1984) called it the stereotype model. For reasons of identification of the parameters, beside setting $a_1 = \phi_1 = 0$, Anderson (1984) also recommended setting $\phi_L = 1$, so that the model consists actually of $L - 1$ model equations. For example, if there are five predictors $(X_1, X_2, \ldots, X_5)$ and four outcome categories of a tumor grade: grade 1, grade 2, grade 3 and grade 4 ($Y = 1, 2, 3, 4$ respectively), this describes the tumor size. The higher the grade, the larger the tumor size and the faster the cancer grows. The stereotype model with a constraint $\phi_4 = 1$ is

$$
\log \left[ \frac{P(Y = 2 | X)}{P(Y = 1 | X)} \right] = a_2 + \phi_2(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5)
$$

$$
\log \left[ \frac{P(Y = 3 | X)}{P(Y = 1 | X)} \right] = a_3 + \phi_3(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5)
$$

$$
\log \left[ \frac{P(Y = 4 | X)}{P(Y = 1 | X)} \right] = a_4 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5
$$

On the other hand, the multinomial logistic regression model is

$$
\log \left[ \frac{P(Y = 2 | X)}{P(Y = 1 | X)} \right] = a_2 + \beta_{21} X_1 + \beta_{22} X_2 + \beta_{23} X_3 + \beta_{24} X_4 + \beta_{25} X_5
$$

$$
\log \left[ \frac{P(Y = 3 | X)}{P(Y = 1 | X)} \right] = a_3 + \beta_{31} X_1 + \beta_{32} X_2 + \beta_{33} X_3 + \beta_{34} X_4 + \beta_{35} X_5
$$
\[
\log \left[ \frac{P(Y = 4 | X)}{P(Y = 1 | X)} \right] = a_4 + \beta_{41}X_1 + \beta_{42}X_2 + \beta_{43}X_3 + \beta_{44}X_4 + \beta_{45}X_5
\]

It can be seen from the above that there are fewer parameters to be estimated in the stereotype model which is \(2L - 3 + p\) parameters while we have \((L - 1) + (L - 1) \times p\) parameters in the multinomial logistic regression model.

Model (2.2) is a one-dimensional stereotype model, (rank=1), since the relationship between \(Y\) and \(X\) involves only one function of \(X, \beta'X\). That is, one function of \(X\) discriminates between all categories. In this model, the \(\beta\) parameters no longer differ between the different categories of the outcome and the score of the response categories is given by the \(\phi_1\) parameter. The maximum possible dimension of matrix \(B\) is the rank of the \(L \times p\) matrix which is at most \(r = \min(L, p)\). For example, if \(p = 2\) (number of covariates = 2), the maximum dimension is two and a two-dimensional stereotype model can be written as

\[
P(Y = \ell | X) = \frac{\exp(a_\ell + \phi_1\beta'X + \psi_\ell\eta'X)}{1 + \sum_{k=2}^{L} \exp(a_k + \phi_k\beta'X + \psi_k\eta'X)}
\]

(2.3)

where \(\psi_\ell\) are scalars, \(\ell = 1, \ldots, L\) and \(a_1 = \phi_1 = 0\). The model now depends on the two functions of \(X\) which are \(\beta'X\) and \(\eta'X\). Furthermore, model (2.2) can be extended further for a higher dimension. Lunt (2005) justified the concepts of dimensionality stating that if one combination can distinguish between levels 1 and 2, but a different one is required to distinguish between levels 2 and 3, the relationship is a two-dimensional stereo-
type model. Examples of the use of multi-dimensional stereotype models were not shown in the literature until Lunt (2005) presented the simplest model and gave an example of a multi-dimensional stereotype model. He also suggested that when the maximum dimension of a stereotype model is used to evaluate probabilities, the likelihood function is exactly equivalent to a multinomial logistic regression even if their interpretations and the actual parameters are different. However, due to some of the computational complexities and lack of suitable software for fitting the multi-dimensional stereotype model, most studies in the literature showed one-dimensional stereotype model such as Anderson (1984) and Greenland (1994).

2.2.1 The Stereotype Model with Fixed Scores

The stereotype model has the same characteristic as the proportional odds model in which an explanatory variable has a similar effect for each logit and it is a special case of the adjacent-category logits. Also, the stereotype model can be used with both ordered and unordered categorical outcomes. This can be considered from the scores of the outcome categories, $\phi_\ell$. Greenland (1994) indicated that if the scores are fixed, it is not particularly realistic in an assumption and also could be regarded as a disadvantage of the stereotype model. Besides, if they are fixed as a constant and equally spaced for each covariate, such as $\{\phi_\ell = (\ell - 1)/(L - 1)\}$ for the constraints $\phi_1 = 0$ and $\phi_L = 1$, the stereotype model is equivalent to the
CHAPTER 2. STEREOTYPE MODEL

proportional odds version of the adjacent-categories logit model (Agresti, 2010). The relationship between these models can be shown below:

Suppose there are $L$ outcome categories with probabilities $P(Y = \ell)$, which can be written as $\pi_\ell$ and $\sum_\ell \pi_\ell = 1$ where $\ell = 1, 2, 3, \ldots, L$ and the first category is the baseline. The general adjacent-categories logit model without assuming the proportional odds assumption has the form

$$\log \left[ \frac{\pi_\ell(X)}{\pi_{\ell-1}(X)} \right] = a_\ell + \beta_\ell' X, \quad \ell = 2, 3, \ldots, L \quad (2.4)$$

This model gives the odds of responding in a higher category versus the lower adjacent category. It also can be seen from the model above that the linear combination of the predictor variables is required to discriminate among the logits or the predictor variables have a different effect for each logit. These logits also are a basic set equivalent to the baseline category logit model used to analyze nominal response outcomes. If we assume the first category is the baseline, the baseline category logit model can be expressed as

$$\log \left[ \frac{\pi_\ell(X)}{\pi_1(X)} \right] = a_\ell + \beta_\ell' X, \quad \ell = 2, 3, \ldots, L \quad (2.5)$$
2.2. THE STEREOTYPE MODEL

To see the connection between the adjacent-categories logit model and the baseline category logit model, we consider a reparameterized baseline category logit model and the connections are

\[
\log \left( \frac{\pi_\ell}{\pi_1} \right) = \log \left( \frac{\pi_\ell}{\pi_{\ell-1}} \right) + \log \left( \frac{\pi_{\ell-1}}{\pi_{\ell-2}} \right) + \ldots + \log \left( \frac{\pi_2}{\pi_1} \right) \tag{2.6}
\]

and

\[
\log \left( \frac{\pi_\ell}{\pi_{\ell-1}} \right) = \log \left( \frac{\pi_\ell}{\pi_1} \right) - \log \left( \frac{\pi_{\ell-1}}{\pi_1} \right)
\]

For example, suppose there are five outcome categories and we treat the first outcome category as a baseline. The baseline category \( \ell = 2, 3, 4, 5 \) can be written as

\[
\log \left( \frac{\pi_5}{\pi_1} \right) = \log \left( \frac{\pi_5}{\pi_4} \right) + \log \left( \frac{\pi_4}{\pi_3} \right) + \log \left( \frac{\pi_3}{\pi_2} \right) + \log \left( \frac{\pi_2}{\pi_1} \right)
\]

\[
\log \left( \frac{\pi_4}{\pi_1} \right) = \log \left( \frac{\pi_4}{\pi_3} \right) + \log \left( \frac{\pi_3}{\pi_2} \right) + \log \left( \frac{\pi_2}{\pi_1} \right)
\]

\[
\log \left( \frac{\pi_3}{\pi_1} \right) = \log \left( \frac{\pi_3}{\pi_2} \right) + \log \left( \frac{\pi_2}{\pi_1} \right)
\]

and for the category two against the baseline category is \( \log \left( \frac{\pi_2}{\pi_1} \right) \). As can be seen, we have \((\ell - 1)\) terms when we reparameterize the model. Furthermore, when we add these \((\ell - 1)\) terms as (2.6) in the general adjacent-
CHAPTER 2. STEREOTYPE MODEL

categories logit model,

$$\log \left[ \frac{\pi_\ell(X)}{\pi_{\ell-1}(X)} \right] = a_\ell + \beta_\ell' X, \quad \ell = 2, 3, \ldots, L$$

the equivalent baseline category logit model is

$$\log \left[ \frac{\pi_2(X)}{\pi_1(X)} \right] = a_2 + \beta_2' X$$

$$\log \left[ \frac{\pi_3(X)}{\pi_1(X)} \right] = (a_3 + \beta_3' X) + (a_2 + \beta_2' X)$$

$$= (a_2 + a_3) + (\beta_2' + \beta_3') X$$

$$\log \left[ \frac{\pi_4(X)}{\pi_1(X)} \right] = (a_4 + \beta_4' X) + (a_3 + \beta_3' X) + (a_2 + \beta_2' X)$$

$$= (a_2 + a_3 + a_4) + (\beta_2' + \beta_3' + \beta_4') X$$

$$\log \left[ \frac{\pi_5(X)}{\pi_1(X)} \right] = (a_4 + \beta_5' X) + (a_4 + \beta_4' X) + (a_3 + \beta_3' X) + (a_2 + \beta_2' X)$$

$$= (a_2 + a_3 + a_4 + a_5) + (\beta_2' + \beta_3' + \beta_4' + \beta_5') X$$

Therefore,

$$\log \left[ \frac{\pi_\ell(X)}{\pi_1(X)} \right] = a_\ell + \left( \sum_{\ell=2}^{L} \beta_\ell' \right) X$$

$$= a_\ell + \beta_\ell' X$$

The covariate effects $\beta^*$ are subscripted by $\ell$. This means that the effects vary according to the response paired with the baseline category. However, the general adjacent-categories logit model (2.4) is not more parsi-
monious than the baseline category logit model (2.5). Thus, from model (2.4) we can create a simple version by using a single parameter instead of $L - 1$ parameters to describe such an effect. The model becomes

$$
\log \left[ \frac{\pi_\ell(X)}{\pi_{\ell-1}(X)} \right] = a_\ell + \beta'X, \quad \ell = 2, 3, \ldots, L
$$

Model (2.7) has the proportional odds assumption that the covariate effects are invariant among the logits. This is why it is called the proportional odds for adjacent-categories logit model. The proportional odds assumption is often used for a model to analyze ordinal responses. The connection between the baseline category logit model and model (2.7) is

$$
\log \left[ \frac{\pi_\ell(X)}{\pi_1(X)} \right] = \sum_{\ell=2}^{L} a_\ell + (\ell - 1)\beta'X
$$

$$
= a^*_\ell + \beta'u_\ell
$$

where $u_\ell = (\ell-1)X$. Thus, the adjacent-categories logits are a constrained version of the baseline category logit model. Model (2.8) uses a single parameter for the covariate effects and the effects depend on the distance between categories, so this model recognizes the ordering of the response scale.

For the stereotype model,

$$
\log \left[ \frac{\pi_\ell(X)}{\pi_1(X)} \right] = \log \left[ \frac{P(Y = \ell | X)}{P(Y = 1 | X)} \right] = a_\ell + \phi_\ell\beta'X, \quad \ell = 2, \ldots, L
$$
if we treat \( \{ \phi \} \) as fixed and let \( \phi_\ell = (\ell - 1) \), then the model becomes the proportional odds version of the adjacent-categories logit model. Therefore, for fixed scores on \( \{ \phi \} \), the stereotype model has the form of a generalized-linear model in the exponential family (Greenland, 1994). We can easily use a generalized-linear model program via constrained polytomous logistic regression to fit a multinomial logistic regression model.

### 2.2.2 The Ordered Stereotype Model

Sometimes it is too difficult to assign fixed scores due to a lack of knowledge and data information. Thus, the scores need to be estimated and may be unequally spaced among the categories. Anderson (1984) also proposed the order constraint with the monotone increasing \( 0 = \phi_1 \leq \phi_2 \leq \ldots \leq \phi_{L-1} \leq \phi_L = 1 \) to capture the ordinal nature of the outcomes. As a result, the responses \( Y \) are treated as ordinal. The ordered stereotype model is defined as

\[
\log \left[ \frac{P(Y = \ell | X)}{P(Y = 1 | X)} \right] = a_\ell + \phi_\ell \beta'X, \quad \ell = 2, \ldots, L \tag{2.9}
\]

with \( 0 = \phi_1 \leq \phi_2 \leq \ldots \leq \phi_{L-1} \leq \phi_L = 1 \). When \( \{ \phi_\ell \} \) are treated as parameters with the non-decreasing constraints, we need to estimate not only \( \{a_\ell, \beta\} \), but also \( \{\phi_\ell\} \), satisfying the constraint for model (2.9). Such a model is referred to as the **ordered stereotype model**. Monotone decreasing \( \{\phi_\ell\} \) could be used but the sign of \( \beta \) would change. Since the scores of response categories, \( \phi_\ell \), are ordered, it could vary considerably between
2.2. THE STEREOTYPE MODEL

the two adjacent categories. When the two adjacent scores are indistinguishable, it is appropriate to constrain those adjacent scores to be equal and refit the model. This corresponds to collapsing the response scale by combining those two categories (Agresti, 2010). When $\phi_{\ell} = \phi_{\ell+1}$, this indicates that the predictors do not have any effects on the probability of having response $\ell + 1$ against response $\ell$. Consequently, for a unit increase in a predictor $X_p$, the log odds ratio $\phi_{1}\beta_p$ of category $\ell$ versus baseline category 1 becomes larger when category $\ell$ is further away from category 1. In other words, this ordered version describes a stochastic order of the response $Y$ subject to the linear combination $\beta'X$. Therefore, the higher the value of $\beta_pX_p$, the more $Y$ tends to move toward the high end of the response scale in a particular predictor $X_p$. However, the ordered version of the stereotype model is difficult to fit due to multiple constraints and lack of standard software.

At this point, under the ordered model, to handle the identifiability problem and conditional on ordering the parameters $0 = \phi_1 \leq \phi_2 \leq \ldots \leq \phi_{L-1} \leq \phi_L = 1$, reparameterization is needed in terms of the differences as

$$\gamma_{\ell} = \phi_{\ell+1} - \phi_{\ell}; \quad \text{for} \quad \ell = 1, \ldots, L - 1$$

Thus, $\phi_{\ell} = \sum_{s=1}^{\ell-1} \gamma_s$, for $\ell = 2, \ldots, L - 1$, since $\phi_1 = 0$ and $\phi_L = 1$. For the purpose of making $\{\phi_{\ell}\}$ increase in $\ell$ and be bounded by 1, we need to compute the set of parameters $\{\gamma_s\}$ to be positive and $0 < \sum_{s=1}^{\ell-1} \gamma_s < 1$. 


## 2.3 Model Interpretation

Interpretation of the parameter $\phi_\ell$ is by comparing response probabilities for two different response categories $\ell$ and $\ell'$. The response paired with the baseline can be expressed as

$$\log \frac{P(Y = \ell | X)}{P(Y = \ell' | X)} = \log \frac{P(Y = \ell | X)}{P(Y = 1 | X)} - \log \frac{P(Y = \ell' | X)}{P(Y = 1 | X)}$$

$$= a_\ell - a_{\ell'} + (\phi_\ell - \phi_{\ell'})\beta'X$$

when \{\phi_\ell\} are fixed and equally spaced, the interpretation of the parameters is similar to the multinomial logistic regression model. This means when $\phi_\ell = \phi_{\ell'}$, $\log [P(Y = \ell | X)/P(Y = \ell' | X)] = a_\ell - a_{\ell'}$ is a constant not dependent on $X$. In that case, the model still holds with the same scores if the response scale is collapsed by combining those two categories. If the difference $(a_\ell - a_{\ell'})$ is positive, this means the total number of outcomes in categories $\ell$ is greater than the total number of outcomes in categories $\ell'$. In addition, when $\phi_\ell = \phi_{\ell'}$, then $X$ does not distinguish between response $\ell$ and $\ell'$. This means that $X$ is not predictive between the two categories. For example, suppose that diseases are graded on scales: least severe, moderate and most severe. Patient’s age ($X$) is used to classify the disease staging. If $\phi_1 = 0, \phi_2 = \phi_3 = 1$, this means age does not help to predict how severe the damage is.

In the case that the scores are not equally spaced but two adjacent scores have almost the same values, then two such respective categories
might be collapsed and the estimation process should be fitted again. Moreover, when the difference \((φ_ℓ - φ_ℓ')\) is large, the odds of having response \(ℓ\) instead of \(ℓ'\) is strongly affected by the covariates. Also, the stereotype model gives a way of estimating how close the adjacent response categories \(ℓ\) and \(ℓ'\) are, based on how close \(φ_ℓ\) and \(φ_ℓ'\) are. In other words, the distinguishability and order of outcome categories could be addressed through these constraints.

When \(\{φ_ℓ\}\) are constrained to be ordered with monotone increasing form, \(0 = φ_1 ≤ φ_2 ≤ \ldots ≤ φ_{L-1} ≤ φ_L = 1\), these scores are directly related to the covariate effect. Therefore, the effect of the explanatory variables on the first odds ratio is less than the effect on the second, and so on. Furthermore, the conditional distributions of \(Y\) are treated as ordinal according to the value of \(βX\) when \(\{φ_ℓ\}\) are ordered. That is, the higher the value of \(βX\), the more \(Y\) tends to move toward the high end of the response scale. Thus, for a particular predictor \(X\), a value of \(β > 0\) means that the response \(Y\) tends to move toward higher values as \(X\) increases. Therefore, in this model the influence of covariates is assessed through the components of \(β\). Likewise, the strength of the effect of each covariate can be expressed through a single parameter \(β\). Note that the effect of the covariates is assumed constant across response categories. That is, the \(β\) parameter no longer differs between the different levels of the response.
CHAPTER 2. STEREOTYPE MODEL

The model simultaneously describes the effects of $X$ on these $L - 1$ logits, $\log \left[ \frac{P(Y = \ell | X)}{P(Y = 1 | X)} \right]$, $\ell = 2, \ldots, L$. From the $\ell^{th}$ logit, we have

$$\frac{P(Y = \ell | X = u)/P(Y = 1 | X = u)}{P(Y = \ell | X = v)/P(Y = 1 | X = v)} = \exp \left( \phi_\ell \beta (u - v) \right)$$

The explanatory variable $X$ has coefficient $\phi_\ell \beta$. This represents the log odds for response $Y = \ell$ versus $Y = 1$ with a unit increase in $X$. That is, when $X = u + 1$, the odds of response $\ell$ instead of $\ell = 1$ are $\exp(\phi_\ell \beta)$ times the odds when $X = u$. In other words, for a unit increase in a particular predictor $X$, the log odds ratio $\phi_\ell \beta$ for the response on level $\ell$ versus level 1 is larger in absolute value when category $\ell$ is farther from category 1.

2.4 Model Fitting

Since the stereotype model can be used with both ordered responses and unordered responses (nominal responses), model fitting for both cases will be discussed as follows. The stereotype model with unordered responses has been shown in model (2.2) and model (2.3). The estimation of parameters under these models can be fitted in a straightforward way by using the maximum likelihood method. The likelihood function can be expressed as

$$L_{\ell}(\cdot) = \prod_{i=1}^{N} \prod_{\ell=1}^{L} \left[ P(Y_i = \ell | X_i) \right]^{I(Y_i = \ell)}$$  (2.10)
where \( I(Y = \ell) \) denotes the indicator function of response \( Y = \ell \). Then models (2.2) and (2.3) are substituted into (2.10) and \( L_c(\cdot) \) is maximized with respect to the parameters of the model. Thus, the likelihood function needs to be optimized iteratively. Gill and Murray (1972) suggested that the Quasi-Newton algorithm be used. Anderson (1984) also presented the numerical optimization method to fit the model. Similarly, Holtbrugge and Schumacher (1991) used an iteratively reweighted least square (IRLS) method to maximize the likelihood function. The stereotype model in view of discriminant analysis was described by Feldmann (1993), who also presented the maximum likelihood parameter estimation. Yee and Hastie (2003) estimated the stereotype model by using the idea of the reduced-rank multinomial logistic models (RR-MLM) and used VGAM (Vector Generalized Additive Model) family function to estimate the parameters. However, VGAM ignores the ordered version of the stereotype model since this model has ordered constraint on the parameters. It is difficult to estimate the model with such a constraint.

The stereotype model has also been treated as a nonlinear model with multivariate response. The likelihood function was maximized directly by using SAS software which supplies unconstrained optimization methods (Kuss, 2006). In detail, he defined a \( L \)-dimensional response vector of dummy variable \( Y^* \) for every observation and wrote the stereotype model
as a nonlinear model through

\[ Y_i^* = f(X_i, a_\ell, \phi_\ell, \beta) + \varepsilon_i \]

with

\[ f(X_i, a_\ell, \phi_\ell, \beta) = \left( \frac{1}{1 + \sum_{k=2}^{L} \exp(a_k + \phi_k \beta X_i)} \right) \cdot \ldots \cdot \left( \frac{\exp(a_L + \phi_L \beta X_i)}{1 + \sum_{k=2}^{L} \exp(a_k + \phi_k \beta X_i)} \right) \]

where \( \varepsilon_i \) is a \( L \times 1 \) vector, \( E(\varepsilon_i) = 0 \), \( \text{Var}(\varepsilon_i) = \Sigma \), \( \Sigma \) has \( (L-1) \times (L-1) \) dimensions and \( a_1 = 0, \phi_1 = 0, \phi_L = 1 \). The subscript \( i \) is for the \( i^{th} \) observation. Two steps of the estimation procedure are required. Generalized least squares and the maximum likelihood estimation are possible to estimate the model parameters. Firstly, estimate \( \hat{\Sigma} \) for \( \Sigma \) from the residuals of the model using maximum likelihood approach by assuming \( \varepsilon_i \) as independent and identically distributed (i.i.d) and \( \varepsilon_i \sim N(0, \Sigma) \). The second step is to minimize the sum squared of error by using generalized least squares method.

\[ SSE(a_\ell, \phi_\ell, \beta, \hat{\Sigma}) = \sum_{i=1}^{N} [Y_i^* - f(X_i, a_\ell, \phi_\ell, \beta)]' \Sigma^{-1} [Y_i^* - f(X_i, a_\ell, \phi_\ell, \beta)] \]

iterating this process further until convergence. Kuss (2006) also suggested that to achieve reasonable estimates of the model parameters, good starting values are needed. Therefore, the set of starting values for the stereotype model (2.2) can be taken from the estimated parameters of the
multinomial logit model. Indeed, we can use the last category of response $\beta_L$ in the multinomial model as a starting point of $\beta$ in the stereotype model and $a_\ell$ can be taken from the multinomial model as well. In the same way, $\{\phi_\ell\}$ can be calculated from the relation $\beta_\ell = \phi_\ell \beta$ between the multinomial model and the stereotype model.

The stereotype model with ordered responses has been shown in model (2.9) with $0 = \phi_1 \leq \phi_2 \leq \ldots \leq \phi_{L-1} \leq \phi_L = 1$. The difficulty in fitting this ordered stereotype model is ordering constraints on the scores $\{\phi_\ell\}$. Greenland (1994) suggested that the appropriate scores can be decided or chosen in advance based on data from a pilot study or using generalized linear models that estimate the scores as additional parameters in the model. Lunt (2005) also indicated that it is not strictly necessary to estimate the $\{\phi_\ell\}$ parameters at all since they may already be known from background information or they may be constrained to take pre-specified values base on expected spacings. Both of their suggestions tried to avoid the difficulty in model fitting.

Since the stereotype model has the complicating factor of being multiplicative in the parameters, Greenland (1994) proposed an alternating algorithm by alternating fixing between $\beta$ and $\phi_\ell$ while the other parameter was estimated. He also suggested that this approach converged rapidly and reliably when started with scores derived from the unordered polytomous model. The proposed two-step iterative method consists of solving at each step problems of different natures: the first step for the stereotype
model (2.9) starts with selecting fixed scores \( \{ \phi_\ell \} \). If the set of \( \{ \phi_\ell \} \) has equally-spaced values, then this model is equivalent to the proportional odds version of the adjacent-categories logit model. In the second step, estimate \( \beta \) and \( \alpha_\ell \) with fixed \( \{ \phi_\ell \} \) using ordinary maximum likelihood estimation. In this step \( \phi_\ell X_p \) is treated as a predictor. On repeating this iterative process, treating the estimated \( \beta \) from the previous step as fixed (\( \beta'X \) being predictors) and \( \{ \phi_\ell \} \) are treated as unknown parameters. Then re-estimate \( \{ \phi_\ell \} \) and repeat the alternative fixing process until convergence occurs.

Furthermore, Ahn et al. (2009) applied Bayesian approach to estimate both unordered and ordered stereotype models. He further proposed the Monte Carlo based full Bayesian method and Expectation Conditional Maximization (ECM) algorithm for the estimation of model parameters in the presence of a completely general missingness mechanism (Ahn et al., 2011).

In this study, we present the ordered stereotype model and use an alternating algorithm which is an iterative two-step approach for selecting fixed between \( \phi_\ell \) and \( \beta \). More information about fitting parameters when the stereotype model is accommodated in the joint model will be discussed in Chapter 3 and Chapter 4.
2.5 The Stereotype Model with Multi-item and Longitudinal Data

Suppose we consider $J$ item responses, $Y_1, Y_2, \ldots, Y_J$, which are designed to serve as measures of the response frequencies of the quality of life items. These items might represent how patients are feeling and the kinds of things they are able to do. The possible answers correspond to the following options: poor, fair, good, very good and excellent. The direction of ordering can be recorded from poor to excellent where low responses imply a poor quality of life. This means each of $Y$ is supported on the ordered categorical set $\{1, 2, \ldots, L\} : \{\text{poor}, \ldots, \text{excellent}\}$. Patients’ answers would also help to understand the effects of the medication they are taking. However, the individual patient is the only person who can express a subjective concept about the quality of life and patients often overestimate their feelings. Information coming from family members, friends, and caregivers can also be important. For these reasons, it has been difficult to use a specific reference level to estimate optimal quality of life. Therefore, we could correct for the unreliability of these issues by using a latent variable model in which the latent trait essentially plays the role of the true quality of life. Throughout this thesis, we define $\theta$ to be a latent random variable and we consider longitudinal designs with $M$ protocol-specified time points for quality of life data collection, $t_1, t_2, \ldots, t_M$. We will present two different joint models: the Bayesian joint model and the
semiparametric joint model. More information of both joint models will be discussed in Chapter 3 and Chapter 4 respectively.

This thesis analyzes data from the Staccato study which examined Human Immunodeficiency Virus (HIV) infection in 354 patients. The response variable was the assessment of the quality of life by using the Medical Outcomes Study HIV Health Survey (MOS-HIV) which indicates how patients are feeling every 24 weeks beginning at the baseline visit until the end of the study. The MOS-HIV questionnaire is composed of many questions on physical and mental health. Each question is in the same ordered general format but there is a difference in the number of ordered outcome categories. Some questions have three responses \( (L = 3) \) and some questions have five responses \( (L = 5) \). However, we still can use the stereotype models to model all questions/items simultaneously.

For the Bayesian joint model (in Chapter 3), the stereotype model used for this example has the form

\[
\log \left( \frac{P(Y_{ijm} = \ell \mid \theta_i(t_m))}{P(Y_{ijm} = 1 \mid \theta_i(t_m))} \right) = a_{\ell} + \phi_{\ell j} \theta_i(t_m) \tag{2.11}
\]

where \( \ell = 2, \ldots, L_j \). A higher value of \( \ell \) indicates a better quality of life. In general, we can let each question has its own intercept \( (a_{\ell j}) \) and score \( (\phi_{\ell j}) \). However, it will result in too many parameters. Thus, a simplifying assumption is made to group the questions that have the same number of ordered outcome categories into the same group. This may not be a reasonable assumption since it assumes all items/questions with the same
2.5. MULTI-ITEM STEREOTYPE MODEL

number of categories have equal intercepts and response category scores regardless of the type of questions. This assumption could be relaxed by allowing different \( a_\ell \) and \( \phi_\ell \) values according to the type of questions and number of response categories. However, for simplicity we retain the assumption.

The response probabilities, \( P(Y_{ijm} = \ell \mid \theta_i(t_m)) \), are modeled through

\[
P(Y_{ijm} = \ell \mid \theta_i(t_m)) = \frac{\exp(a_\ell + \phi_\ell b_j \theta_i(t_m))}{1 + \sum_{k=2}^{L} \exp(a_k + \phi_k b_j \theta_i(t_m))}, \quad \ell = 1, 2, \ldots, L
\]

where \( Y_{ijm} \) denotes the response on question/item \( j \) for subject \( i \) at time \( t_m \) and \( \theta_i(t_m) \) is defined as the value of a continuous latent random variable for subject \( i \) at time \( t_m \). An item intercept parameter is \( a_\ell \) and the score of outcome categories is \( \phi_\ell \). In the previous section, \( \beta \) was defined as the covariate effect but henceforth the strength of effect of each covariate from item response \( j \) at different time point is specified by \( b_j \). This is because of the complicated structure of our joint model which is composed of three main models: the stereotype model, the regression model and the Cox proportional hazards model. We need to change some notation in some models in order to make them distinguishable.

From the stereotype model (2.11), we can conclude that exponentiating a \( \phi_\ell b_j \) value gives an odds ratio for the odds of response \( \ell \) instead of response “poor” (\( \ell = 1 \)) with a unit increase in \( \theta \). For example, the odds ratio comparing the response “good” (\( \ell = 3 \)) and response “poor” for a 1-unit increase in quality of life score is \( \exp(\phi_3 b_j) \).
For the semiparametric joint model (in Chapter 4), we construct the joint likelihood function by using an idea of the finite mixture model. The latent variable $\theta$ is treated as a discrete variable. Thus, the stereotype model with discrete latent random variable becomes

$$\log \left[ \frac{P(Y_{ijm} = \ell \mid \theta_r)}{P(Y_{ijm} = 1 \mid \theta_r)} \right] = a_\ell + \phi_\ell (b_j + \theta_r)$$

where $a_\ell$ is an item intercept parameter and we set $a_1 = 0$, $b_j$ is the item slope coefficient, $\phi_\ell$ is the score of the response categories and we consider the ordered stereotype model with $0 = \phi_1 \leq \phi_2 \leq \ldots \leq \phi_{L-1} \leq \phi_L = 1$. The latent variable $\theta_r$ is defined to be a discrete latent random variable taking one of the discrete values $1, 2, \ldots, R$. 
Chapter 3

A Bayesian Joint Model

3.1 Introduction

Quality of life (QOL) and length of survival of patients are often evaluated in clinical trials. Medical research also frequently focuses on the relationship between quality of life and survival time of subjects since two interventions can have very similar survival outcomes but have different effects on a patient’s quality of life. In the Staccato study, the quality of life assessments were measured regularly every 12 weeks after patients received the treatment. Most of the quality of life scales were ordinal such as good, fair, poor. This generated a sequence $Y = (Y(t_1), Y(t_2), \ldots, Y(t_M))$ of longitudinal quality of life data for each patient. Longitudinal data involve repeated measurements of a response variable at a number of time points. During the collection of the longitudinal quality of life data, time to HIV progression was also considered in parallel. This can be referred to
as the time-to-event data $T$.

Longitudinal data and time-to-event data can be analyzed as two separate end points. Alternatively, they can be jointly analyzed as one combined end point if the two data components are likely to be related. That is the $Y$ and $T$ processes may be dependent on each other. Although methods for separate analyses of the two data components are well established, joint analysis of the two endpoints is still under study and is being increasingly developed (Henderson, 2005). There are potential gains in power in considering two endpoints simultaneously, such as considering quality of life and survival time as two simultaneous processes. Many scientific studies generate both longitudinal data and survival data simultaneously. For example, Berridge and Whitehead (1971) presented a combination of a proportional hazards model for the failure times and a continuation ratio model for the ordinal category responses. Similarly, Ribaudo et al. (2000) demonstrated the joint analysis of quality of life response and log survival time. They indicated that the quality of life response might account for different survival patterns. The estimation of the joint distribution gives appropriately weighted estimates for the average quality of life response pattern over the observed distribution of survival times.

Moreover, Wang et al. (2002) suggested that joint modeling quality of life and survival could be used to combine information for summarizing the outcome in terms of a patient’s risk of dying and this might be predicted, in part, by his current quality of life. Chi and Ibrahim (2006)
proposed a joint likelihood approach to jointly model multidimensional quality of life and multivariate survival data to investigate the relationship between quality of life, disease-free survival and overall survival in the International Breast Cancer Study Group trial.

There are various approaches for the simultaneous analysis of quality of life and survival data in literature. The most widely used approach in clinical research is quality-adjusted survival analysis. This method combines the amount of time that patients spend in a number of different health states with weights reflecting the quality of life. The modeling results are used to create a composite measure of quality and quantity of life. Although quality-adjusted survival analysis is extensively used and researchers can overcome the problem of missing quality of life data due to death, it does not adequately deal with missing data resulting from dropout from the quality of life study prior to death. Alternatively, the joint modeling approach may provide the means to model the dropout processes explicitly (Billingham and Abrams, 2002).

This thesis investigates a joint model of longitudinal ordinal data and survival data. This is illustrated using an application to the joint modeling of quality of life measurements and survival time for HIV patients being followed up after treatment. The ordinal quality of life measurements are modeled using the stereotype model, while the survival time is modeled using the Cox proportional hazards model. They are linked through a continuous latent variable that is assumed to underlie the quality of life
and the hazard rate. A well known example for the joint modeling application is AIDS research and the primary motivation is to characterize the relationship between a longitudinal continuous variable and time to HIV/AIDS progression (Ibrahim et al., 2001). However, we demonstrate the joint model in another application which is longitudinal quality of life data and time to HIV/AIDS progression. Also, this shows a new feature of joint modeling because we use the stereotype model to model the longitudinal ordinal data. This is the first time that the stereotype model is used in a longitudinal ordinal and survival joint model.

We illustrate our methodology by analyzing data from the Staccato study, a randomized trial to compare two methods of treatment: the continuous anti-retroviral treatment and the CD4 guided interruption treatment (Ananworanich and the Staccato study group, 2006), for HIV infection of Thai patients on Highly Active Antiretroviral Therapy (HAART). Randomized treatment continued during an average of approximately three years. In order to analyze the impact of these two treatment methods on the quality of life in Thai patients, Nüesch et al. (2009) calculated summary scores of quality of life and transformed to standardized z-scores, with a mean of 50 and a standard deviation of 10, and compared the standardized scores between two treatment methods in the Staccato study. However, in this thesis we do not use the method of summary scores. We will use the raw scores from the quality of life questionnaire to analyze the data.

We are interested in the CDC stage at the end of the randomized treat-
3.1. INTRODUCTION

ment period as a survival endpoint of this study. Note that CDC is the classification system of HIV disease stage by CD4 cell counts and by the presence of specific HIV-related conditions defined by the U.S. Centers for Disease Control and Prevention (CDC) (Centers for Disease Control and Prevention, 1992). More information about the CDC stage will be given in Chapter 5 (data application).

Patients were also monitored longitudinally with respect to biological endpoints, such as the CD4 count, throughout the follow-up period. This was done by taking immunologic or virologic measures in the case of infectious diseases. A questionnaire was used to assess the quality of life every 24 weeks beginning at the baseline visit and at weeks 24, 48, 72, 96, 120 and 144 respectively, \((M\) protocol-specified time points for quality of life data collection, \(t_1, \ldots, t_M\)). There was no information about quality of life after 144 weeks, even though patients were followed up for survival. Often these longitudinal measures are incomplete but these measurements are still important because they may be predictive of survival. Furthermore, the optimal goals of treatment of HIV/AIDS patients are to prolong their survival and to help them maintain a good quality of life. Since the antiretroviral therapy is time consuming, the study of quality of life among HIV/AIDS patients receiving antiretroviral therapy would be necessary to assess the success of the treatment. However, giving different treatment strategies may affect the quality of life of treated patients. Therefore, joint model methods are appropriate when it is considered that
the evaluation of changes of quality of life over time and the survival endpoints may be dependent on each other. The joint model has an advantage in dealing with the problem of missing quality of life data resulting from dropout prior to death or withdrawal before the study has ended. This is because of the incomplete nature of the data actually observed. The quality of life data might not be completely collected at time \( t_m \). The complete information is very rarely available on all patients. In spite of this, the joint model has emerged as an effective way to gain information from both components. Since the model is so complex with a high dimensional data set, Bayesian methods have been considered and the parameters of the model are estimated using the Metropolis-Hastings Algorithm.

3.2 Model Structure

The Bayesian joint model consists of an item response model and a Cox proportional hazards model which are linked by the latent variables. To create the item response part, we consider the item responses from the MOS-HIV quality of life questionnaire at different time points during the follow-up. For example, the questionnaire assesses the general health of individuals and there is a question on health status that asks “In general would you say your health is Poor, Fair, Good, Very good, Excellent?”. Although we can assume the categories are ordered, we do not know the structure of this ordering with respect to a given explanatory variable. Thus, the stereotype model for item responses is developed to solve such
complicated measured scale problems. In the case of dichotomous outcome data, for example the questionnaire asks “Does your health keep you from working at a job, doing work around the house or going to school: Yes or No?”, the binary model is used for analysis.

If the quality of life information can only be measured indirectly by using observable indicators such as a questionnaire, a latent variable model is used for deriving the unobservable variable from the indicators. The latent variable is designed to collect all information in the set of questions and it is treated as a shared variable in our joint model. That is it appears as a covariate in both components. The joint likelihood of the ordinal longitudinal and survival data for a subject is specified by conditioning on the latent variable which is assumed to have a multivariate normal distribution. We generate the latent random variable for each individual to link the quality of life and survival time together.

In the Staccato study, there are two survival endpoints, CDC-B and CDC-C, but we consider only one clinical event, CDC-B, since there are very few patients who experience CDC-C at the end of the study (five out of 354). Then patients in category C are treated as censored in this approach. Thus, the clinical event CDC-B at the end of the randomized treatment period is used as the survival endpoint in the survival part. The Cox proportional hazards model is used to model time to diagnose the new CDC-B. We also consider treatment group assignments and initial CD4 counts as time-independent covariates and the quality of life latent vari-
ables that are created at failure time points as time-dependent covariates in this part.

3.2.1 Item Response Models

To contribute to the joint model, there are two models in the item response part: the stereotype model and the binary model since the MOS-HIV questionnaire has 35 items/questions with varying-dimensions. We will use the binary model to model the binary data and the stereotype model to model the ordinal data. Each question is in the same ordered general format from poor to excellent but there is a difference in numbers of ordered outcome categories. The stereotype model can be used to model all items simultaneously.

3.2.1.1 The Stereotype Model

We consider \( J \) item responses, \( Y_1, Y_2, \ldots, Y_J \), which are designed to measure health related quality of life at \( M \) protocol-specified time points for quality of life data collection, \( t_1, t_2, \ldots, t_M \), and each of the \( Y_i's \) can be written in the form of an ordered categorical set \( \{1, 2, \ldots, L_j\} \). The direction of ordering can be recorded from poor to excellent. That is, a lower level implies poor quality of life. The stereotype model is used to model the probabilities of responding on level \( \ell, \ell = 1, 2, \ldots, L_j \) for each of the quality of life items and in this study the first category, \( \ell = 1 \), is treated as the baseline category. Therefore, \( Y_{ijm} \) is used to denote the \( j^{th} \) item response
of the \(i^{th}\) subject at the \(m^{th}\) time point, \(Y_{ijm} = 1, 2, \ldots, L_j\) for \(i = 1, 2, \ldots, n\), \(j = 1, 2, \ldots, J\) and \(m = 1, 2, \ldots, M\). The stereotype model can be expressed as

\[
\log \left[ \frac{P(Y_{ijm} = \ell \mid \theta_i(t_m))}{P(Y_{ijm} = 1 \mid \theta_i(t_m))} \right] = a_\ell + \phi_\ell b_j \theta_i(t_m)
\] (3.1)

where

- \(a_\ell\) is an item intercept parameter and we set \(a_1 = 0\).
- \(b_j\) is an item slope coefficient and we assume \(b_j \geq 0\).
- \(\phi_\ell\) is a score for the outcome level and we consider the ordered stereotype model, \(0 = \phi_1 \leq \phi_2 \leq \ldots \leq \phi_L = 1\), throughout this thesis.

\(\theta_i(t_m)\) is defined to be a continuous latent random variable that represents a quality of life score for the \(i^{th}\) subject at the \(m^{th}\) time point.

Note that the assumption \(b_j \geq 0\) is a very strict assumption and may not always be necessary. This is reasonable when all the responses are ordered from low to high responses and when all items/questions show the increasing rates of higher responses as \(\theta_i\) increase. In the general case, we can allow to take \(b_j\) with any values from \((-\infty, +\infty)\). This would allow the questions/items with ordered responses that are negatively correlated to \(\theta_i\) to have negative coefficients. For example, a drug that improves quality of life overall, but which has specific negative side effects that influence, negatively, the response to just one or two questions.
Given the latent variable, $\theta_i(t_m)$, the responses $Y_{ijm}$, $i = 1, \ldots, n$, $j = 1, \ldots, J$, $m = 1, \ldots, M$ are treated as independent called the local independence assumption. The likelihood of the stereotype model has the form

$$L^J(a, b, \phi \theta) = \prod_{i=1}^{n} \prod_{m=1}^{M} \prod_{j=1}^{J} \prod_{\ell=1}^{L} \left( \frac{\exp(a_{\ell} + \phi_{\ell} b_j \theta_i(t_m))}{1 + \sum_{k=2}^{L} \exp(a_{k} + \phi_{k} b_j \theta_i(t_m))} \right)^{y_{ijm\ell}} \quad (3.2)$$

where $y_{ijm\ell} = I(Y_{ijm} = \ell)$ represents the binary outcome (“Yes” and “No”) on the response category $\ell$ for the $i^{th}$ subject, the $j^{th}$ item and the $m^{th}$ time point. For example, when $Y_{ijm} = 1$, this means that subject $i$ selects category $\ell = 1$ on question $j$ at time $t_m$. Then $y_{ijm1} = 1$ and the other categories $y_{ijm2} = 0, y_{ijm3} = 0, \ldots, y_{ijmL} = 0$.

According to $\{\phi_{\ell}\}$ with monotone increasing scores, the response $Y_{ijm}$ is treated as ordinal. The probabilities of $\{Y_{ijm} = \ell, \ell = 1, \ldots, L_j\}$ depend on the value of $b_j \theta_i(t_m)$. That is, when $b_j \theta_i(t_m)$ increases, then $Y_{ijm}$ tends to move toward the high end of the response scale. Therefore, to ensure the model has the usual interpretation by which a higher score of $\theta_i(t_m)$ means a better quality of life, $b_j$ is constrained to be a non-negative value. In other words, patients who have a greater value of $\theta_i(t_m)$ would have a high probability of responding to a high category. Since we ordered the quality of life responses from poor to excellent (from low to high) and we assume a positive correlation between quality of life and the probability of responding in a high category for all items/questions, then patients in a high category should have a better quality of life compared to the patients in a low category.
Although the numbers of response categories are not the same for all questions, we still can use the stereotype models to model all items/questions together. For example, one question in the MOS-HIV questionnaire is “How has the quality of your life been during the past four weeks: very bad, pretty bad, good and bad parts about equal, pretty good, very well?”. This question has five ordinal categories ($L = 5$). On the other hand, a question which assesses the physical domain and asks “Does your health now limit you in bending, lifting or stooping: limited a lot, limited a little, not limited” has just only three ordinal categories ($L = 3$). Thus, we can create two different stereotype models to incorporate into the joint model by allowing parameters $a_\ell$ and $\phi_\ell$ to be different but share the same $\theta_i(t_m)$. In general the stereotype model has the form as model (3.1) and this can be equivalent to

$$\log \left[ \frac{P(Y_{ijm} = \ell | \theta_i(t_m))}{P(Y_{ijm} = 1 | \theta_i(t_m))} \right] = a_\ell^L + \phi_\ell^L b_j \theta_i(t_m)$$

(3.3)

where $L$ refers to the number of ordered outcome categories for question/item $j$. The assumption in Chapter 2, section 2.5 on page 44 is used here.

### 3.2.1.2 The Binary Model

The binary model is appropriate when the response can be divided into two categories, for example agree or disagree. This is a special case of the stereotype model with $L = 2$ and the estimation of $(\phi_1, \phi_2) = (0, 1)$. In this
study, the participants were asked to answer some questions with “Yes” or “No” responses. Each patient must be in one or the other category and cannot be in both. For example, we could define $y_i = 0$ if the $i^{th}$ patient has been unable to do certain kinds or amounts of work because of his/her health, otherwise $y_i = 1$. When $y_i = 1$, it implies good quality of life.

Once we view $y_i$ as a realization of a random variable $Y_i$ that can take the values one and zero with probabilities $\pi_i$ and $1 - \pi_i$ respectively, the distribution of $Y_i$ can be written as

$$P(Y_i = y_i) = \pi_i^y (1 - \pi_i)^{1-y}$$

where $y_i = 0, 1$.

Subsequently, we can construct a logit model for the $q^{th}$ binary question at the $m^{th}$ time point as

$$\log \left[ \frac{P(Y_{iqm} = 1)}{1 - P(Y_{iqm} = 1)} \right] = a^* + b_q^* \theta_i(t_m)$$

where

- $a^*$ is an intercept parameter.
- $b_q^*$ is an item slope coefficient and $b_q^* \geq 0$.
- $\theta_i(t_m)$ is a continuous latent random variable for the $i^{th}$ subject at time $t_m$. 
3.2. MODEL STRUCTURE

Exponentiating equation (3.4) we find that the odds for the $i^{th}$ unit are given by

$$\frac{P(Y_{iqm} = 1)}{1 - P(Y_{iqm} = 1)} = \exp(a^* + b^*_q \theta_i(t_m))$$

Solving for the probability $P(Y_{iqm} = 1)$ gives

$$P(Y_{iqm} = 1) = \frac{\exp(a^* + b^*_q \theta_i(t_m))}{1 + \exp(a^* + b^*_q \theta_i(t_m))}$$

Thus, the likelihood of the binary model can be expressed as

$$L^Q(a^*, b^* | \theta) = \prod_{i=1}^n \prod_{m=1}^M \prod_{q=1}^Q \left( \frac{\exp(a^* + b^*_q \theta_i(t_m))}{1 + \exp(a^* + b^*_q \theta_i(t_m))} \right)^{y_{iqm}} \left( \frac{1}{1 + \exp(a^* + b^*_q \theta_i(t_m))} \right)^{1-y_{iqm}}$$

where $Q$ is the number of binary questions and $y_{iqm}$ denotes the item response for the $i^{th}$ subject with the $q^{th}$ item/question at the $m^{th}$ time point.

3.2.2 The Cox Proportional Hazards Model

Let $T$ be a nonnegative random variable associated with the survival time and assume that a probability distribution of $T$ is described by a probability density function $f(t)$. The cumulative distribution function of $T$ is then given by

$$F(t) = P(T < t) = \int_0^t f(u)du,$$
and $F(t)$ represents the probability that the survival time is less than some value $t$. The survival function, $S(t)$, is defined by

$$S(t) = P(T \geq t) = 1 - F(t) \quad (3.7)$$

This survival function, $S(t)$, denotes a probability that the random variable $T$ exceeds the specified time $t$. In other words, the survival function gives the probability that a person survives longer than some specified time $t$.

The hazard function, denoted by $h(t)$, specifies the instantaneous rate of failure at $T = t$ conditional on survival to time $t$ and is given by the formula

$$h(t) = \lim_{\Delta t \to 0} \left[ \frac{P(t \leq T < (t + \Delta t)|T \geq t)}{\Delta t} \right] \quad (3.8)$$

Using the probability theory $P(A|B) = \frac{P(AB)}{P(B)}$, the conditional probability of the hazard function $h(t)$ is

$$\frac{P(t \leq T < (t + \Delta t))}{P(T \geq t)}$$

and this equal to

$$\frac{F(t + \Delta t) - F(t)}{S(t)}$$
3.2. MODEL STRUCTURE

Then,

\[ h(t) = \lim_{\Delta t \to 0} \left[ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right] \frac{1}{S(t)} \]

and \( \lim_{\Delta t \to 0} \left[ \frac{F(t+\Delta t)-F(t)}{\Delta t} \right] \) is the definition of the derivative of \( F(t) \) with respect to \( t \), which is \( f(t) \) in equation (3.6). Thus, the relationship between the survival and hazard function can be shown as

\[ h(t) = \frac{f(t)}{S(t)}, \]

and

\[ h(t) = -\frac{d}{dt} (\log S(t)) \]

Therefore,

\[ S(t) = \exp(-H(t)) \]

where

\[ H(t) = \int_0^t h(u)du \]

\( H(t) \) is called the cumulative hazard function and \( H(t) \) can also be obtained from the survival function by \( H(t) = -\log S(t) \).
In this thesis, the Cox proportional hazards model (Cox, 1972) is used to model time to diagnose the new CDC-B and we also consider time-independent covariates such as treatment group assignments and the initial CD4 count in the survival process.

In a proportional hazard model, it is assumed that the hazard function for the failure time of the \( i \)th individual is of the form

\[
\lim_{\Delta t \to 0} \left[ \frac{P(t \leq T_i < (t + \Delta t) \mid T_i \geq t, X_i, \theta^*_i)}{\Delta t} \right] = h_i(t \mid X_i, \theta^*_i) = h_0(t) \exp(\theta^*_i \delta_0 + X_i \delta_1) \tag{3.9}
\]

where \( P(t \leq T_i < t + \Delta t \mid T_i \geq t) = P(\text{individual fails in the interval } [t, t + \Delta t] \mid \text{survival up to time } t) \). This is defined as the probability of failure during a very small time interval, assuming that the individual has survived to the beginning of the interval, or as the limit of the probability that an individual fails in a very short interval \( t + \Delta t \), given that the individual has survived to time \( t \).

The latent variable of quality of life, denoted by \( \theta^*_i \), is a latent random variable contributing to each individual’s underlying hazard. A time-independent covariate (e.g. treatment) is represented by \( X_i \) and \( \{\delta_0, \delta_1\} \) are regression coefficients. In general, the model can include more than one covariate.

The baseline hazard function, denoted by \( h_0(t) \), can be any shape as a function of \( t \) with the condition \( h_0(t) > 0 \). This is the nonparametric part of the model and a linear component, \( (\theta^*_i \delta_0 + X_i \delta_1) \), is the parametric part of
the model. Thus, the Cox proportional hazards model is a semiparametric model. However, in this chapter we consider a partial likelihood function that depends only on the parameter of interest. Methods for estimating the baseline hazard function and the semiparametric model will therefore be discussed in Chapter 4. The relevant likelihood function for the proportional hazards model in equation (3.9) is given by the expression

$$L(\delta_0, \delta_1 | \theta) = \prod_{i=1}^{n} \left( \frac{\exp(\theta_i^* \delta_0 + X_i \delta_1)}{\sum_{p \in R(t_i)} \exp(\theta_p^* \delta_0 + X_p \delta_1)} \right)^{d_i}$$ (3.10)

Cox (1972) called this equation (3.10) a partial likelihood function. This likelihood function depends on \(n\) observed failure or censored time points which are ordered and denoted by \(t_{(1)} \leq t_{(2)} \leq \ldots \leq t_{(n)}\). Then \(t_{(i)}\) is the \(i^{th}\) ordered survival time. The set of individuals who are alive and uncensored at time \(t_i\) is called the risk set expressed by \(R(t_{(i)})\) and \(d_i\) is an event indicator. If an individual failed, \(d_i = 1\) otherwise \(0\). In this study, the endpoint is not a fatal illness or death but refers to progression to CDC-B. Furthermore, in this survival part we also include a latent variable for quality of life, \(\theta_i^*\), at a failure time point which corresponds to time to CDC-B progression. There are \(K\) distinct times at which failures are observed and denoted by \(w_1 < \ldots < w_k < \ldots < w_K\) where \(w_k \in \{t_{(1)}, t_{(2)}, \ldots, t_{(n)}\}\). This means all observations have their own latent variables as quality of life scores \(\theta_i^*\) for each follow-up time point and we only calculate \(\theta_i^*\) for patients who present with a diagnosis of the CDC-B event. Therefore, the latent variables at this stage are treated as time-dependent covariates since
these values change over time. All these values are determined by linear interpolation between two known \( \theta \) points. As an example, if individual \( i \) failed at time \( w_k \), then we need to calculate \( \theta^*_i \) at time \( w_k \) denoted by \( \theta^*_i(w_k) \). Since we know that \( t_m < w_k < t_{m'} \) where \( t_m \) and \( t_{m'} \) correspond to \( \theta_i(t_m) \) and \( \theta_i(t_{m'}) \) respectively, then the estimate of \( \theta^*_i(w_k) \), the value corresponding to \( w_k \), is given by linear interpolation as

\[
\theta^*_i(w_k) = \theta_i(t_m) + \frac{w_k - t_m}{t_{m'} - t_m} (\theta_i(t_{m'}) - \theta_i(t_m))
\]

### 3.2.3 The Regression Model

Wang et al. (2002) demonstrated the regression process to incorporate the latent random variable into the joint model. We also used the same format as Wang et al. (2002) to generate a latent variable for each individual to link the quality of life and survival time together. The vector of latent variables is viewed as a random vector, denoted by \( \theta_i = (\theta_i(t_1), \ldots, \theta_i(t_M))^\prime \). We assume that \( \theta_i \sim MVN(\mu(\theta_i), \Sigma(\rho)) \) has a multivariate normal distribution, where \( \mu(\theta_i) \) and \( \Sigma(\rho) \) are referred to as mean vector and covariance matrix of \( \theta_i \), respectively and both of them are functions of time.

The regression part can be accommodated in the joint likelihood function by

\[
\Phi(\theta; \mu(\theta_i), \Sigma(\rho))
\]
where $\Phi$ is the multivariate normal density function. We consider two treatment methods (continuous and CD4-guided) as a covariate denoted by $X = 1$ for continuous treatment and $X = 0$ for CD4-guided treatment. The term $\mu(t_m) = \beta_0(t_m) + \beta_1(t_m)X_i$ is referred to as the mean of quality of life for each treatment group at time $t_m$ where $\beta_0(t_m)$ and $\beta_1(t_m)$ denote the intercept and the coefficient of $X_i$ at time $t_m$ respectively and $\mu(\theta_i) = (\mu(\theta_i(t_1)), \ldots, \mu(\theta_i(t_M)))$. The regression coefficients $\beta_1 = (\beta_1(t_1), \ldots, \beta_1(t_M))^\prime$ measure how treatment affects the quality of life. The elements of $\Sigma(\rho)$ follow a simple form of covariance parameterization as

$$
cov(\theta_{t_m}, \theta_{t_{m'}}) = \sigma^2 \rho^{|t_m - t_{m'}|}
$$

where in this study we set the variance of $\theta$, $\sigma^2$, at all time point to be 1 similar to Wang et al. (2002). Then the covariance structure becomes

$$
cov(\theta_{t_m}, \theta_{t_{m'}}) = \rho^{|t_m - t_{m'}|}
$$

where $0 \leq m \leq 6$ denotes the time point which are in units of weeks. This covariance structure is appropriate since the correlation between $\theta$ at different time points decreases as the time interval increases. As mentioned above, quality of life data have been collected every 24 weeks beginning at the baseline visit and at week 24, 48, 72, 96, 120, 144 respectively. Thus, \{0, 24, 48, 72, 96, 120, 144\} and we rescale by divid-
CHAPTER 3. A BAYESIAN JOINT MODEL

ing all time points by 24 so \( \{t_0, t_1, t_2, t_3, t_4, t_5, t_6\} \) becomes \( \{0, 1, 2, 3, 4, 5, 6\} \).

3.2.4 The Full Likelihood Function

The joint likelihood function is obtained from combining the item response part, the regression part and the survival part as follows:

\[
L(a, b, \phi, a^*, b^*, \beta, \rho, \delta, \theta) = L^J(a, b, \phi | \theta) L^Q(a^*, b^* | \theta) \Phi(\theta; \mu(\theta), \Sigma(\rho)) L(\delta | \theta) \quad (3.12)
\]

where the stereotype likelihood \( L^J(a, b, \phi | \theta) \), the binary likelihood \( L^Q(a^*, b^* | \theta) \) and the survival partial likelihood \( L(\delta | \theta) \) are defined in (3.2), (3.5), (3.10) respectively and \( \Phi(\theta; \mu(\theta), \Sigma(\rho)) \) is the multivariate normal density function of the latent variable \( \theta \).

When questions in a questionnaire contain a different number of ordinal responses, we can use separate parameters in the stereotype model. For example, suppose some questions have \( L_1 \) levels and the others have \( L_2 \) levels. All of them are included in the joint model and the full likelihood function from equation (3.12) can be written as

\[
L^J(a^{L_1}, b^{L_1}, b_{L_2}, \phi^{L_1}, \phi^{L_2} | \theta) L^Q(a^*, b^* | \theta) \Phi(\theta; \mu(\theta), \Sigma(\rho)) L(\delta | \theta) \quad (3.13)
\]

For simplicity, we only use notations \( a, b \) and \( \phi \) instead of \( a^{L_1}, a^{L_2}, b^{L_1}, b^{L_2}, \phi^{L_1}, \phi^{L_2} \) in the later sections.

The joint model is represented graphically in Figure 3.1.
3.2. MODEL STRUCTURE

Figure 3.1: The joint model. If \( Y_{ijm} \) is binary, then \( \phi = 1 \), \( a = a^* \) and \( b = b^* \). \( J \) is number of items, \( K \) is number of distinct failure time points, \( M \) is number of follow-up time points and \( n \) is number of individuals.
3.3 Parameter Estimation

We can estimate all parameters in the item response models and survival model simultaneously based on the joint likelihood function but this joint model is highly complicated due to unobservable latent variables, censoring and the nonlinear structures of the models. Therefore, we consider a Markov Chain Monte Carlo algorithm by drawing values of parameters from the joint posterior distribution. We use the Metropolis-Hastings Algorithm (Metropolis et al., 1953) to estimate model parameters simultaneously.

The idea of the Metropolis-Hastings Algorithm is that we generate values from a proposal distribution, denoted by $q(\cdot | \cdot)$, and we need to apply criteria to determine whether to accept or reject the simulated values. The criterion is calculated from ratios of the product of proposal and posterior densities. This criterion is commonly referred to as the acceptance probability. We used flat prior distributions on all parameters to reflect lack of prior information. The posterior density function is then proportional to the joint likelihood function (3.12). Using the flat prior can lead to an improper posterior distribution which can make the inference difficult. It is advisable to investigate the different prior.

Due to conditional independence relationships among the parameters, the joint posterior can be written as a product of several distributions which allows us to divide the parameter vector into blocks.

Suppose we have a parameter vector $\Theta$ that contains all parameters of
3.3. PARAMETER ESTIMATION

the likelihood function (3.12). We divided \( \Theta \) into \( K \) blocks, \( \Theta = \{ \Theta_1, \Theta_2, \ldots, \Theta_K \} \) with the posterior of interest \( P(\Theta_k|\cdot) \). We then update all blocks one by one when the other blocks remain fixed and are not affected.

The Metropolis-Hastings Algorithm for a single selected parameter vector iterates the following steps:

1. Select an arbitrary initial value for all \( \Theta^{(0)} \).

2. For each \( k \) in turn, sample \( \Theta_k^{\text{new}} \) from the proposal distribution, \( q(\Theta_k^{\text{new}}|\Theta_k^{\text{cur}}) \), which depends on the current state.

3. Calculate an acceptance probability as

\[
    r = \frac{q(\Theta_k^{\text{cur}}|\Theta_k^{\text{new}})P(\Theta_k^{\text{new}}|\cdot)}{q(\Theta_k^{\text{new}}|\Theta_k^{\text{cur}})P(\Theta_k^{\text{cur}}|\cdot)}
\]

where \( P(\Theta_k^{\text{new}}|\cdot) \) is part of the posterior density function depending on block \( \Theta_k \).

4. Sample a value \( u \) from the Uniform(0,1) distribution.

5. Accept the new value \( \Theta_k^{\text{new}} \) as \( \Theta_k^{(t)} \) if \( u \leq \min(1, r) \). Otherwise \( \Theta_k^{(t)} = \Theta_k^{\text{cur}} \) and the move is not allowed.

6. Change the iteration from \( t \) to \( t + 1 \) and return to step 2 until convergence to the posterior distribution is reached.
The Estimation Procedure

The set of parameters that we need to estimate is \( \{a, b, \phi, a^*, b^*, \beta_0, \beta_1, \rho, \delta, \theta\} \) and we set a uniform prior for all parameters, \( P(\cdot) \propto 1 \).

At the \( t^{th} \) iteration:

1. Updating parameters in the regression model.

1.1. Updating \( \rho \). First, we generate \( \rho \) for creating covariance matrix, \( \Sigma \). The posterior distribution of \( \rho \), \( P(\rho|\cdot) \), is

\[
P(\rho|\beta, X, \Sigma(\rho), \theta) \propto \prod_{i=1}^{n} \frac{1}{\sqrt{\det(\Sigma(\rho))}} \exp \left( -\frac{1}{2} (\theta_i - \mu(\theta_i))' \Sigma(\rho)^{-1} (\theta_i - \mu(\theta_i)) \right)
\]

Given the current value, \( \rho^{cur} \), propose a new value, \( \rho^{new} \), by taking a random draw from the truncated normal distribution with truncated to the interval \((0, 1)\), \( \rho^{new} \sim N(\rho^{cur}, \sigma^2) \) where \( \sigma^2 = 1 \). The proposed value, \( \rho^{new} \), is accepted with probability equal to the minimum of 1 and the ratio

\[
r = \frac{q(\rho^{cur}|\rho^{new})P(\rho^{new}|\cdot)}{q(\rho^{new}|\rho^{cur})P(\rho^{cur}|\cdot)}
\]

where the form of the ratio

\[
\frac{q(\rho^{cur}|\rho^{new})}{q(\rho^{new}|\rho^{cur})} = \frac{\Phi(1 - \rho^{cur}) - \Phi(-\rho^{cur})}{\Phi(1 - \rho^{new}) - \Phi(-\rho^{new})}
\]

where \( \Phi \) is the cumulative distribution function of the normal distribution. If the proposed value is not accepted, then \( \rho^{(t)} \) equal to \( \rho^{cur} \).
1.2. Creating covariance matrix $\Sigma$ by using $\rho$ from the previous step. We have seven time points: $0 \leq m \leq 6$, \{t_0, t_1, t_2, t_3, t_4, t_5, t_6\} = \{0, 1, 2, 3, 4, 5, 6\} which refer to quality of life data collection at baseline and at week 24, 48, 72, 96, 120, 144 respectively.

$$\Sigma = \begin{bmatrix}
1 & \text{cov}(\theta_{t_0}, \theta_{t_1}) & \cdots & \text{cov}(\theta_{t_0}, \theta_{t_6}) \\
\text{cov}(\theta_{t_1}, \theta_{t_0}) & 1 & \cdots & \text{cov}(\theta_{t_1}, \theta_{t_6}) \\
\vdots & \vdots & \ddots & \vdots \\
\text{cov}(\theta_{t_6}, \theta_{t_0}) & \text{cov}(\theta_{t_6}, \theta_{t_1}) & \cdots & 1
\end{bmatrix}$$

1.3. Updating $\beta$. There are two beta vectors to estimate and each beta vector has seven time points. Then, we estimate each beta one at a time.

$\beta_0 = \{\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta_{05}, \beta_{06}, \beta_{07}\}$ and $\beta_1 = \{\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{15}, \beta_{16}, \beta_{17}\}$

Thus, the posterior distribution of $\beta$, $P(\beta|\cdot)$, is

$$P(\beta|\mu(\theta), X, \Sigma(\rho), \theta) \propto \prod_{i=1}^{n} \exp \left( -\frac{1}{2} (\theta_i - \mu(\theta_i))'(\Sigma(\rho))^{-1} (\theta_i - \mu(\theta_i)) \right)$$

where $\mu(\theta_i) = \beta_0 + \beta_1 X$. For $m = 1, \ldots, M$, given the current value, $\beta_{0m}^{\text{cur}}$, propose a new value, $\beta_{0m}^{\text{new}}$, by taking a random draw from the univariate normal distribution, $\beta_{0m}^{\text{new}} \sim N(\beta_{0m}^{\text{cur}}, \sigma_{\beta_{0m}}^2)$ where $\sigma_{\beta_{0m}}^2 = 1$. The proposed value, $\beta_{0m}^{\text{new}}$, is accepted with probability equal to the minimum of 1 and the ratio

$$r = \frac{q(\beta_{0m}^{\text{cur}}|\beta_{0m}^{\text{new}})P(\beta_{0m}^{\text{new}}|\cdot)}{q(\beta_{0m}^{\text{new}}|\beta_{0m}^{\text{cur}})P(\beta_{0m}^{\text{cur}}|\cdot)}$$
where the form of the ratio
\[
\frac{q(\beta_{\text{cur}} | \beta_{\text{new}})}{q(\beta_{\text{new}} | \beta_{\text{cur}})} = \frac{1}{\sigma_{\beta_{0m}} \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_{\beta_{0m}}^2} (\beta_{\text{cur}} - \beta_{\text{new}})^2\right) = 1
\]

Since \(\frac{q(\beta_{\text{cur}} | \beta_{\text{new}})}{q(\beta_{\text{new}} | \beta_{\text{cur}})}\) is 1, \(r\) is effectively just the likelihood ratio. If the proposed value is not accepted, then \(\beta_{0m}\) equal to \(\beta_{\text{cur}}\). Updating \(\beta_{1m}\) can be done using the same procedure as updating \(\beta_{0m}\).

2. Generating missing values.

Any missing item responses for the \(i^{th}\) individual can be generated according to the response probabilities derived under the current parameters of the stereotype model and the binary model. If the response is ordinal, we generate such missing values from the stereotype model (3.1). Otherwise if the response is binary, we generate it from the binary model (3.4). The missing values are filled in with random draws from a multinomial distribution with the probability equal to \(P(Y_{ijm} = \ell | \text{current parameters})\). There are about 20 out of 354 patients with at least one question missing at the first time point. This number increases in the succeeding follow-up time points up to 152 patients at the last time point.

3. Updating parameters in the stereotype model.

At this stage, \(a, b\) and \(\phi\) are updated. The likelihood function of
3.3. PARAMETER ESTIMATION

\[ a, b \text{ and } \phi \text{ is} \]

\[ L(a, b, \phi|\theta) = \frac{\prod_{i=1}^{n} \prod_{m=1}^{M} \prod_{j=1}^{J} \prod_{\ell=1}^{L} \left( \exp(a_\ell + \phi_\ell b_j \theta_i(t_m)) \right)}{1 + \sum_{k=2}^{L} \exp(a_k + \phi_k b_j \theta_i(t_m))} \]

where \( a_1 = 0, b_j \geq 0 \) and we set \( \phi_1 = 0 \) and \( \phi_L = 1 \) to make the model identifiable.

3.1. For updating \( \phi \), we reparameterize the parameters in terms of the differences, \( \gamma_\ell = \phi_{\ell+1} - \phi_\ell \) for \( \ell = 1, \ldots, L - 1 \). Thus, \( \phi_\ell = \sum_{s=1}^{L-1} \gamma_s \) for \( \ell = 2, \ldots, L - 1 \). If \( 0 = \phi_0 < \phi_1 < \ldots < \phi_L = 1 \) are \( L \) order statistics from the uniform distribution \( U(0, 1) \), then \( \{\gamma_s\}, s = 1, \ldots, L - 1 \), follows a Dirichlet \( (\alpha_1, \ldots, \alpha_{L-1}) \) distribution (Ahn et al., 2009) and the posterior distribution for \( \{\gamma_s\} \) is expressed as

\[ P(\gamma_s|\cdot) \propto \frac{\prod_{i=1}^{n} \prod_{m=1}^{M} \prod_{j=1}^{J} \prod_{\ell=1}^{L} \left( \exp(a_\ell + \sum_{s=0}^{\ell-1} \gamma_s b_j \theta_i(t_m)) \right)}{1 + \sum_{k=2}^{L} \exp(a_k + \sum_{s=1}^{k-1} \gamma_s b_j \theta_i(t_m))} \times \gamma_s^{\alpha_s-1} \]

For the special case of Dirichlet \((1, \ldots, 1)\) prior, the posterior distribution reduced to

\[ P(\gamma_s|\cdot) \propto \frac{\prod_{i=1}^{n} \prod_{m=1}^{M} \prod_{j=1}^{J} \prod_{\ell=1}^{L} \left( \exp(a_\ell + \sum_{s=0}^{\ell-1} \gamma_s b_j \theta_i(t_m)) \right)}{1 + \sum_{k=2}^{L} \exp(a_k + \sum_{s=1}^{k-1} \gamma_s b_j \theta_i(t_m))} \times \gamma_s^{\alpha_s-1} \]
CHAPTER 3. A BAYESIAN JOINT MODEL

Given the current value, $\gamma^{\text{cur}}$, propose a new value, $\gamma^{\text{new}}$, by randomly selecting one of $\{\gamma_s\}$, $s \sim U(2, L - 1)$, and taking a random draw from the uniform distribution\(^1\), $\gamma^{\text{new}}_s \sim U(0, \gamma_{s-1} + \gamma_s)$. Then, reparameterization all $\{\gamma^{\text{new}}_s\}$ to add to 1 by

$$\gamma^{\text{new}}_{s-1} = (\gamma_{s-1} + \gamma_s) - \gamma^{\text{new}}_s$$

The proposed value, $\gamma^{\text{new}}_s$, is accepted with probability equal to the minimum of 1 and the ratio

$$r = \frac{q(\gamma^{\text{cur}}_s | \gamma^{\text{new}}_s) P(\gamma^{\text{new}}_s | \cdot)}{q(\gamma^{\text{new}}_s | \gamma^{\text{cur}}_s) P(\gamma^{\text{cur}}_s | \cdot)}$$

If the proposed value is not accepted, then $\gamma^{(t)}_s$ equal to $\gamma^{\text{cur}}_s$.

Next, convert $\{\gamma_s\}$ to $\{\phi_\ell\}$ for updating $a$ and $b$ respectively.

3.2. **Updating $a$.** The posterior distribution of $a$, $P(a | \cdot)$, is

$$P(a | b, \phi, \theta) \propto \prod_{i=1}^{n} \prod_{m=1}^{M} \prod_{j=1}^{J} \prod_{\ell=1}^{L} \left( \frac{\exp(a_\ell + \phi_\ell b_j \theta_i(t_m))}{1 + \sum_{k=2}^{L} \exp(a_k + \phi_k b_j \theta_i(t_m))} \right)^{y_{ijm\ell}}$$

where each $a_\ell$ is updated separately. Given the current value, $a^{\text{cur}}_\ell$, propose a new value, $a^{\text{new}}_\ell$, by taking a random draw from the univariate normal distribution, $a^{\text{new}}_\ell \sim N(a^{\text{cur}}_\ell, \sigma^2_a)$ where $\sigma^2_a = 1$. The proposed value, $a^{\text{new}}_\ell$, is accepted with probability equal to the minimum of 1 and the ratio

\(^1\)Originally we (incorrectly) drew $\gamma^{\text{new}}_s \sim U(\max(0, \gamma_s - a), \min(1, \gamma_s + a))$ and then reparameterized all $\{\gamma_s\}$ to sum up to 1. The corrected version we propose above properly accounts for the forward and reverse transition probability $q(\cdot | \cdot)$.\hfill
3.3. PARAMETER ESTIMATION

\[ r = \frac{q(a_{\text{cur}}^{\ell}|a_{\text{new}}^{\ell})P(a_{\text{new}}^{\ell}|\cdot)}{q(a_{\text{new}}^{\ell}|a_{\text{cur}}^{\ell})P(a_{\text{cur}}^{\ell}|\cdot)} \]

where the form of the ratio

\[ \frac{q(a_{\text{cur}}^{\ell}|a_{\text{new}}^{\ell})}{q(a_{\text{new}}^{\ell}|a_{\text{cur}}^{\ell})} = \frac{1}{\sigma_{\text{a}}\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_{\text{a}}^2}(a_{\text{cur}}^{\ell} - a_{\text{new}}^{\ell})^2\right) \]

If the proposed value is not accepted, then \( a_{\ell}^{(t)} \) equal to \( a_{\text{cur}}^{\ell} \).

3.3. Updating \( b \). The posterior distribution of \( b \), \( P(b|\cdot) \), is

\[ P(b|\alpha, \phi, \theta) \propto \prod_{i=1}^{n} \prod_{m=1}^{M} \prod_{j=1}^{J} \prod_{\ell=1}^{L} \left( \frac{\exp(a_{\ell} + \phi_{b}b_{j}t_{m})}{1 + \sum_{k=2}^{L} \exp(a_{k} + \phi_{b}b_{j}t_{m})} \right)^{y_{i,j,m,\ell}} \]

where each \( b_{j} \) is updated separately. Given the current value, \( b_{j}^{\text{cur}} \), propose a new value, \( b_{j}^{\text{new}} \), by taking a random draw from the normal distribution truncated at 0, \( b_{j}^{\text{new}} \sim N(b_{j}^{\text{cur}}, \sigma_{b}^2) \) where \( \sigma_{b}^2 = 1 \) and \( b_{j} \geq 0 \). The proposed value, \( b_{j}^{\text{new}} \), is accepted with probability equal to the minimum of 1 and the ratio

\[ r = \frac{q(b_{j}^{\text{cur}}|b_{j}^{\text{new}})P(b_{j}^{\text{new}}|\cdot)}{q(b_{j}^{\text{new}}|b_{j}^{\text{cur}})P(b_{j}^{\text{cur}}|\cdot)} \]

where the form of the ratio

\[ \frac{q(b_{j}^{\text{cur}}|b_{j}^{\text{new}})}{q(b_{j}^{\text{new}}|b_{j}^{\text{cur}})} = \frac{1 - \Phi(-b_{j}^{\text{cur}})}{1 - \Phi(-b_{j}^{\text{new}})} \]

If the proposed value is not accepted, then \( b_{j}^{(t)} \) equal to \( b_{j}^{\text{cur}} \).
4. Updating parameters in the binary model.

Updating $a^*$ and $b^*$ for the binary model. The likelihood function of $a^*$ and $b^*$ is expressed as model (3.5).

4.1. Updating $a^*$. The posterior distribution of $a^*$, $P(a^*|\cdot)$, is

$$P(a^*|\theta) \propto \prod_{i=1}^{n} \prod_{m=1}^{M} \prod_{q=1}^{Q} \left( \frac{\exp(a^* + b^*_q \theta_i(t_m))}{1 + \exp(a^* + b^*_q \theta_i(t_m))} \right)^{y_{iqm}} \left( \frac{1}{1 + \exp(a^* + b^*_q \theta_i(t_m))} \right)^{1-y_{iqm}}$$

Given the current value, $a^{*\text{cur}}$, propose a new value, $a^{*\text{new}}$, by taking a random draw from the univariate normal distribution,

$$a^{*\text{new}}|a^{*\text{cur}} \sim N(a^{*\text{cur}}, \sigma_{a^*}^2)$$

where $\sigma_{a^*} = 1$. The proposed value, $a^{*\text{new}}$, is accepted with probability equal to the minimum of 1 and the ratio

$$r = \frac{q(a^{*\text{cur}}|a^{*\text{new}})P(a^{*\text{new}}|\cdot)}{q(a^{*\text{new}}|a^{*\text{cur}})P(a^{*\text{cur}}|\cdot)}$$

where the form of the ratio

$$\frac{q(a^{*\text{cur}}|a^{*\text{new}})}{q(a^{*\text{new}}|a^{*\text{cur}})} = \frac{1}{\frac{1}{\sigma_{a^*} \sqrt{2\pi}} \exp\left( -\frac{1}{2\sigma_{a^*}^2} (a^{*\text{cur}} - a^{*\text{new}})^2 \right)} \frac{1}{\frac{1}{\sigma_{a^*} \sqrt{2\pi}} \exp\left( -\frac{1}{2\sigma_{a^*}^2} (a^{*\text{new}} - a^{*\text{cur}})^2 \right)} = 1$$

If the proposed value is not accepted, then $a^{*\text{cur}}$ equal to $a^{*\text{cur}}$.

4.2. Updating $b^*$. The posterior distribution of $b^*$, $P(b^*|\cdot)$, is

$$P(b^*|\theta) \propto \prod_{i=1}^{n} \prod_{m=1}^{M} \prod_{q=1}^{Q} \left( \frac{\exp(a^* + b^*_q \theta_i(t_m))}{1 + \exp(a^* + b^*_q \theta_i(t_m))} \right)^{y_{iqm}} \left( \frac{1}{1 + \exp(a^* + b^*_q \theta_i(t_m))} \right)^{1-y_{iqm}}$$

where each $b^*_q$ is updated separately. Given the current value, $b^{*\text{cur}}_q$, propose a new value, $b^{*\text{new}}_q$, by taking a random draw
3.3. PARAMETER ESTIMATION

from the normal distribution truncated at 0, \( \beta_{q}^{new} \sim N(\beta_{q}^{cur}, \sigma_{\beta}^{2}) \)
where \( \sigma_{\beta}^{2} = 1 \) and \( \beta_{q}^{*} \geq 0 \). The proposed value, \( \beta_{q}^{new} \), is accepted
with probability equal to the minimum of 1 and the ratio

\[
q = \frac{q(b_{q}^{cur}|b_{q}^{new})P(b_{q}^{new}|\cdot)}{q(b_{q}^{new}|b_{q}^{cur})P(b_{q}^{cur}|\cdot)}
\]

where the form of the ratio

\[
\frac{q(b_{q}^{cur}|b_{q}^{new})}{q(b_{q}^{new}|b_{q}^{cur})} = 1 - \Phi(-b_{q}^{cur})
\]

\[
= 1 - \Phi(-b_{q}^{new})
\]

If the proposed value is not accepted, then \( \beta_{q}^{(t)} \) equal to \( \beta_{q}^{cur} \).

5. Updating parameters in the Cox proportional hazards model.

5.1. Generating latent variables at failure time points . We con-
sider the event that occurs for some objects and denote the \( K \)
time points corresponding to time to CDC-B by \( w_{1} < \ldots < w_{K} \)
(K failure time points). In this study, there are 26 failure time
points: \( w_{1} < \ldots < w_{26} \). The values at the failure time points \( \theta^{*} \)
determined by linear interpolation between two known \( \theta \) points
as described in the previous section.

5.2. Generating \( \delta = \{\delta_0, \delta_1\} \) by using \( \theta^{*} \) from previous step and \( \delta \) is
the regression coefficient based on the partial likelihood func-
tion as shown in model (3.10).
5.2.1 Updating $\delta_0$. The posterior distribution of $\delta_0$, $P(\delta_0|\cdot)$, is

$$P(\delta_0|\delta_1, \theta^\ast) \propto \prod_{i=1}^{n} \left( \frac{\exp(\theta^*_i \delta_0 + X_i \delta_1)}{\sum_{p \in R(t(i))} \exp(\theta^*_p \delta_0 + X_p \delta_1)} \right)^{d_i}$$

Given the current value, $\delta^{\text{cur}}_0$, propose a new value, $\delta^{\text{new}}_0$, by taking a random draw from the univariate normal distribution, $\delta^{\text{new}}_0 \sim N(\delta^{\text{cur}}_0, \sigma^{2\delta_0})$ where $\sigma^{2\delta_0} = 1$. The proposed value, $\delta^{\text{new}}_0$, is accepted with probability equal to the minimum of 1 and the ratio

$$r = \frac{q(\delta^{\text{cur}}_0|\delta^{\text{new}}_0)P(\delta^{\text{new}}_0|\cdot)}{q(\delta^{\text{new}}_0|\delta^{\text{cur}}_0)P(\delta^{\text{cur}}_0|\cdot)}$$

where the form of the ratio

$$\frac{q(\delta^{\text{cur}}_0|\delta^{\text{new}}_0)}{q(\delta^{\text{new}}_0|\delta^{\text{cur}}_0)} = \frac{1}{\sigma_{\delta_0} \sqrt{2\pi}} \exp\left(\frac{-1}{2\sigma_{\delta_0}^2} (\delta^{\text{new}}_0 - \delta^{\text{cur}}_0)^2\right) = 1$$

If the proposed value is not accepted, then $\delta^{(t)}_0$ equal to $\delta^{\text{cur}}_0$.

5.2.2 Updating $\delta_1$. The posterior distribution of $\delta_1$, $P(\delta_1|\cdot)$, is

$$P(\delta_1|\delta_0, \theta^\ast) \propto \prod_{i=1}^{n} \left( \frac{\exp(\theta^*_i \delta_0 + X_i \delta_1)}{\sum_{p \in R(t(i))} \exp(\theta^*_p \delta_0 + X_p \delta_1)} \right)^{d_i}$$

Given the current value, $\delta^{\text{cur}}_1$, propose a new value, $\delta^{\text{new}}_1$, by taking a random draw from the univariate normal distribution, $\delta^{\text{new}}_1 \sim N(\delta^{\text{cur}}_1, \sigma^{2\delta_1})$ where $\sigma^{2\delta_1} = 1$. The proposed value, $\delta^{\text{new}}_1$, is accepted with probability equal to the minimum of 1 and the ratio
3.3. PARAMETER ESTIMATION

\[ r = \frac{q(\delta_{\text{cur}}|\delta_{\text{new}}) P(\delta_{\text{new}}|.)}{q(\delta_{\text{new}}|\delta_{\text{cur}}) P(\delta_{\text{cur}}|.)} \]

where the form of the ratio

\[ \frac{q(\delta_{\text{cur}}|\delta_{\text{new}})}{q(\delta_{\text{new}}|\delta_{\text{cur}})} = \frac{1}{\sigma_{\delta_1} \sqrt{2\pi}} \exp\left(\frac{-1}{2\sigma_{\delta_1}^2} (\delta_{\text{cur}} - \delta_{\text{new}})^2\right) = 1 \]

If the proposed value is not accepted, then \( \delta_{\text{cur}}^{(t)} \) equal to \( \delta_{\text{cur}}^{(t)} \).

6. Updating the latent variable.

Updating the latent variable, \( \theta \), for each individual. For the \( i^{th} \) individual from all parameters of the model, the posterior distribution of \( \theta_i \) is proportional to the joint likelihood function (3.12).

\[ P(\theta|a,b,\phi,a^*,b^*,\beta,\rho,\delta) \propto L(a,b,\phi,a^*,b^*,\beta,\rho,\delta) \]

The initial value of \( \theta^{(0)} \) is obtained as follows: the baseline values, \( \theta_{i1}^{(0)} \), are obtained by converting the rank of the sum of baseline item responses to quantiles of the standard normal distribution (Duncan and MacEachern, 2008) using the following equation,

\[ \theta_{i1}^{(0)} = \Phi^{-1}\left(\frac{\text{Rank}(\sum_j Y_{ij})}{n + 1}\right) \]

Then we generate the values of \( \theta^{(0)} \) for the other time points from a multivariate normal distribution with mean equal to baseline values, \( \theta_{i1}^{(0)} \), and covariance equal to \( \Sigma(\rho^{(0)}) \).
At the $t^{th}$ iteration, given the current value, $\theta^{cur}$, propose new values, $\theta^{new}$, by taking a random draw from the multivariate normal distribution, $q(\theta^{new} \sim MVN(\theta^{cur}, \Sigma(\rho))$. The proposed values, $\theta^{new}$, are accepted with probability equal to the minimum of 1 and the ratio

$$r = \frac{q(\theta^{cur}|\theta^{new})P(\theta^{new}|\cdot)}{q(\theta^{new}|\theta^{cur})P(\theta^{cur}|\cdot)}$$

where the form of the ratio

$$q(\theta^{cur}|\theta^{new}) = \frac{e^{-\frac{1}{2}(\theta^{cur}-\theta^{new})'(\Sigma(\rho))^{-1}(\theta^{cur}-\theta^{new})}}{e^{-\frac{1}{2}(\theta^{new}-\theta^{cur})'(\Sigma(\rho))^{-1}(\theta^{new}-\theta^{cur})}}$$

If the proposed values are not accepted, then $\theta^{(t)}$ equal to $\theta^{cur}$.

After updating all parameters from step 1 to step 6 has finished, change the iteration from $t$ to $t+1$ and repeat all six steps until convergence to the target distribution is reached.

### 3.4 Convergence Diagnostics

After the joint models have reached the stationary distribution, samples from the conditional distribution are used to determine the posterior distribution of parameters of interest. The questions are how long should the burn-in period be and how many iterations do we need to run the chains after burn-in. The iterations in the burn-in period should be discarded since they are not part of the posterior distribution. Thus, the inference is based on values that come from the posterior distribution. One way to
observe is to see how well the chain is mixing or moving around the parameter space. Since the stationary distribution will always be unknown in practice, convergence diagnostics of the chains are needed.

A convergence diagnostic procedure is a method for assessing how long a Markov chain needs to run in order to obtain good estimates of parameters. Although there are several approaches in the literature to diagnose the convergence, throughout this thesis we assessed the convergence by visual inspection using trace plot and density plot. Also, we used two statistic diagnostic tests, Gelman and Rubin’s multiple sequence diagnostic (Gelman and Rubin, 1992) and Heidelberger and Welch diagnostic (Heidelberger and Welch, 1983).

3.4.1 Visual Inspection

We used graphical diagnostics, trace and density plots, to assess convergence. These two approaches are included in CODA (Convergence Diagnosis and Output Analysis) software (Best et al., 1996; Plummer et al., 2006). CODA is available in R and also there are more formal diagnostic tests that can be implemented using this package. Gilks et al. (1996) suggested that visual inspection of the Monte-Carlo output-plots is the most obvious and commonly used method for determining burn-in.
CHAPTER 3. A BAYESIAN JOINT MODEL

3.4.1.1 Trace plot

A trace plot is a plot of the iteration number versus the value of the draw of the parameter at each iteration. This plot can be very useful in assessing convergence since it can show whether the chain has converged to its stationary distribution or the chain is mixing well. If the chain remains stable for a long period of time, this means there is a visual indication of convergence.

3.4.1.2 Density plot

A density plot shows a smoothed probability density curve of the draws or the distribution of the parameters. This plot gives an indication of the mean and spread of the values obtain from the sampler. A classic sign of non-convergence for most applications is multimodality of the density estimate.

3.4.2 Statistical Diagnostic Tests

Although visual inspection is a commonly used method, there is a general problem with the trace plot. If the chain remains attracted to a nonoptimal mode for a long period of time, there is no visual indication that this area is not the desired highest density region (Gill, 2008). Therefore, assessing convergence should be supported by more advanced techniques. In this section, we consider two statistical approaches to diagnose the convergence.
3.4. CONVERGENCE DIAGNOSTICS

3.4.2.1 Gelman and Rubin’s Multiple Sequence Diagnostic

The Gelman and Rubin approach (Gelman and Rubin, 1992) is based on comparing two or more parallel chains drawn from different starting points and checking that they are indistinguishable. Their method was illustrated by a comparison of the within and between chain variances for each parameter. Approximate convergence is diagnosed when the variance within each chain should be the same or no less than the variance between the different chains.

Suppose we have $m$ parallel chains each of length $n$ of each parameter $\Theta$. The values are denoted by $\Theta_{ij}$ where $i = 1, \ldots, m$ and $j = 1, \ldots, n$. The statistic for each parameter of interest ($\Theta$) is based on the following procedure:

1. Run $m \geq 2$ chains of length $2n$ each from overdispersed starting values.

\[
\Theta^{[0]}_{(1)}, \Theta^{[1]}_{(1)}, \ldots, \Theta^{[n]}_{(1)}, \ldots, \Theta^{[2n-1]}_{(1)}, \Theta^{[2n]}_{(1)} \\
\Theta^{[0]}_{(2)}, \Theta^{[1]}_{(2)}, \ldots, \Theta^{[n]}_{(2)}, \ldots, \Theta^{[2n-1]}_{(2)}, \Theta^{[2n]}_{(2)} \\
\vdots \\
\Theta^{[0]}_{(m)}, \Theta^{[1]}_{(m)}, \ldots, \Theta^{[n]}_{(m)}, \ldots, \Theta^{[2n-1]}_{(m)}, \Theta^{[2n]}_{(1)}
\]

2. Discard the first $n$ draws in each chain.

3. Calculate the within-chain and between-chain variance.

\[
\text{The variance within-chain: } W = \frac{1}{m} \sum_{i=1}^{m} s_i^2
\]
where \( s_i^2 = \frac{1}{n-1} \sum_{j=1}^{n} (\Theta_{ij} - \bar{\Theta}_i)^2 \)

The variance between chain: \( B = \frac{n}{m-1} \sum_{i=1}^{m} (\bar{\Theta}_i - \bar{\bar{\Theta}})^2 \)

where \( \bar{\bar{\Theta}} = \frac{1}{m} \sum_{i=1}^{m} \bar{\Theta}_i \), \( \bar{\bar{\Theta}} \) is the grand mean (the mean of means since each subchain is of equal length)

4. Calculate the estimated variance of the parameter as a weighted sum of the within-chain and between-chain variance.

\[
\text{Estimated variance: } \hat{\text{Var}}(\Theta) = \frac{n-1}{n} W + \frac{1}{n} B
\]

This estimate of the variance is unbiased for stationary distributions if the starting points of the chain were actually drawn from the target distribution.

5. Calculate the potential scale reduction factor.

\[ \hat{R} = \sqrt{\frac{\hat{\text{Var}}(\Theta)}{W}} \]

The values of \( \hat{R} \) near 1 mean the \( m \) chains are all operating on the same distribution but in practice values less than roughly 1.1 or 1.2 are acceptable (Gelman, 1996). If \( \hat{R} \) is high, then longer chains are needed in order to achieve convergence.

The idea of this test is that before reaching convergence, the variance within-chain (\( W \)) underestimates total posterior variation in \( \Theta \) because the chains have not fully explored the target distribution. On the other hand, \( \hat{\text{Var}}(\Theta) \) overestimates total posterior variance because the starting
points are overdispersed relative to the target. Once convergence has been reached, this means that the parallel Markov chains are essentially overlapping.

### 3.4.2.2 Heidelberger and Welch Diagnostic

The Heidelberger and Welch method (Heidelberger and Welch, 1983) calculates a test statistic based on Brownian bridge theory and uses the Cramervon-Mises test statistic to accept and reject the null hypothesis that the Markov chain is from a stationary distribution. The diagnostic consists of two parts: a stationarity test and a halfwidth test. The general form for steps of the stationary test are then as follows:

1. Generating a chain of $N$ iterations and specify an accuracy ($\epsilon$) and an alpha level of the test ($\alpha$).

2. Using the first 10% of chain values calculate the test statistic on these samples and observe whether it indicates a tail value.

3. If the null hypothesis is rejected, the first 10% of the iterations are discarded and the test is run again.

4. If the null hypothesis is again rejected, a further 10% of iterations are discarded.

5. This process continues until either 50% of the data have been dismissed or the null hypothesis is accepted with the remaining iterations. Otherwise, a longer run length is needed.
If the stationary test is passed, then the halfwidth analysis part of the
diagnostic is performed. The proportion of the data which passed the sta-
tionary test is used to calculate the halfwidth of the \((1 - \alpha)\%\) credible in-
terval around the sample mean. If the halfwidth is less than the rough
estimate of the accuracy, \(\epsilon\), then the chain is considered to be converged.
Otherwise, a longer chain is required.

3.5  Bayesian Model Comparison

In this section, we consider the Bayesian Information Criterion (BIC) and
the Deviance Information Criterion (DIC) in order to make comparisons
among different models to find the best fitting model to the data.

3.5.1  The Bayesian Information Criterion (BIC)

Kass and Raftery (1995) suggest using the Bayesian Information Criterion
(BIC) as an estimate of the Bayes Factor (Kass and Raftery, 1995) for two
competing models. The BIC is defined as

\[
BIC = -2\ell(\hat{\Theta}|x) + p\log(N)
\]

where \(\ell(\hat{\Theta}|x)\) is the maximized log likelihood value, \(p\) is the number of pa-
rameters in the model and \(N\) is the number of observation. This criterion
consists of two terms, the first one is twice negative value of the maxi-
mized log likelihood function and the second one depends on the number
of parameters and the sample size \( N \). The idea of BIC is to select one model from all of the competing models that have the minimum BIC value. This means the model with the lowest value of BIC is chosen as the best model.

### 3.5.2 The Deviance Information Criterion (DIC)

The Deviance Information Criterion (DIC) was described by Spiegelhalter et al. (2002). There are two objectives for DIC: describing model fit and model complexity as in the BIC.

The first quantity to define before considering the DIC is the Bayesian deviance (Spiegelhalter et al., 2002). Suppose \( p(y|\Theta) \) is the likelihood of a model that we consider, \( y \) is the observed data and \( \Theta \) is a vector of parameters. The Bayesian deviance is given by

\[
D(\Theta) = -2\log[p(y|\Theta)] + 2\log[f(y)]
\]  

where \( f(y) \) is a function of the data and \( f(y) \) also serves as a standardizing factor. Spiegelhalter et al. (2002) suggested using \( f(y) = 1 \) (giving zero contribution in equation (3.14)) since this term must be identical to calculate both models and therefore it can be canceled out and does not need to be known. As we mentioned above, in order to evaluate the Bayesian model fit, we can insert a condition on the data \( y \) and taking an expectation
over $\Theta$ in equation (3.14) as

$$
\overline{D}(\Theta) = E_\Theta[-2\log[p(y|\Theta)|y] + 2\log[f(y)]]
$$

(3.15)

where $\overline{D}(\Theta)$ is called the posterior mean of Bayesian deviance. Subsequently, we define $\tilde{\Theta}$ as a posterior estimate of $\Theta$ and insert into equation (3.14) shown as

$$
D(\tilde{\Theta}) = -2\log[p(y|\tilde{\Theta})] + 2\log[f(y)]
$$

(3.16)

where $D(\tilde{\Theta})$ is called the Bayesian deviance of the posterior mean. Then, the effective dimension of the model is now defined by:

$$
p_D = \overline{D}(\Theta) - D(\tilde{\Theta})
$$

(3.17)

This $p_D$ is used to measure the model complexity for the effective number of parameters in the model. Applying Bayes’ theorem with equation (3.17), we obtain

$$
p_D = \{ E_\Theta[-2\log[p(y|\Theta)|y] + 2\log[f(y)]]\} - \{ -2\log[p(y|\tilde{\Theta})] + 2\log[f(y)]\}
$$

$$
= E_{\Theta|y}[ -2\log(p(y|\Theta))] + 2\log(p(y|\tilde{\Theta}))
$$

$$
= E_{\Theta|y}[ -2\log\left(\frac{p(\Theta|y)p(y)}{p(\Theta)}\right)] + 2\log\left(\frac{p(\tilde{\Theta}|y)p(y)}{p(\Theta)}\right)
$$

$$
= E_{\Theta|y}[-2\log(p(\Theta|y)/p(\Theta))] + 2\log(p(\tilde{\Theta}|y)/p(\tilde{\Theta}))
$$
The DIC is then defined by combining two terms (3.15) and (3.17) together

\[ DIC = \overline{D}(\Theta) + p_D = 2\overline{D}(\Theta) - D(\tilde{\Theta}) \]  

(3.18)

Thus, the model that has the lowest DIC will have the highest posterior probability and such a model should be selected.
Chapter 4

A Semiparametric Joint Model

4.1 Introduction

The Cox proportional hazards model that we used in the survival part of the joint model is known as a semiparametric model. Because the baseline hazard, \( h_0(t) \), is an unspecified function, this property makes the Cox model a semiparametric model. The specific form of the model gives the hazard function as a product of an unspecified baseline hazard, \( h_0(t) \), and an exponential expression involving the covariates. Although the baseline hazard is not specified, it is still possible to estimate the covariate coefficients in the exponential part of the model. As we did in the previous chapter, we evaluated the effect of the covariates: treatment and quality of life. The effect estimates were calculated without having to estimate the baseline hazard function. The Cox model likelihood function without the baseline hazard is called the partial likelihood function. This characterizes
how the hazard function changes as a function of the covariates. If we also need to describe how the hazard function changes as a function of survival time, the baseline hazard function should be included in the model. The likelihood function is then called the full likelihood function.

However, Cox (1972) stated that the estimated parameters from the partial likelihood function would be asymptotically equivalent to those from the full likelihood function. In the previous chapter, it was sufficient to use the partial survival likelihood function since the Bayesian approach is based on the distribution of parameters. Thus, the partial likelihood function will give reliable enough results and we do not need to worry about whether the wrong parametric form is chosen for the hazard function (Kleinbaum and Klein, 2005). There are some applied settings in literature that assumed the baseline hazard, $h_0(t)$, as a nuisance parameter since it typically has a very high dimension and is difficult to calculate. For example, Ye et al. (2008) evaluated the effect of the longitudinal biomarker prostate specific antigen on the recurrence of prostate cancer in a prostate cancer study. They proposed a penalized joint likelihood approach to avoid estimation of the baseline hazard function.

The full likelihood function is also widely used in literature such as Tseng et al. (2005), Hsieh et al. (2006) and Dang et al. (2007). It is based on all the observed data and it allows the estimation of the baseline hazard function and the exponential expression simultaneously. Therefore, this chapter demonstrates a semiparametric joint model where the base-
line hazard function is included in the survival component, by specifying the baseline hazard as a step function. The stereotype model is represented in the item response part, while the semiparametric model is used in the survival part. We present a shared latent variable in the case that it is assumed to be a discrete variable to classify cases to their most likely group. Each group may differ in patient characteristic or attitudes from the survey responses. Furthermore, an advantage of using this methodology is that it can be extended to cluster analysis (Pledger and Arnold, 2011). Since the latent variable is treated as a discrete variable, the distribution of the observed item responses is a finite mixture. This idea has been proposed by Titterington et al. (1985) and a similar idea has also been expressed by Woodruff and Hanson (1996). The EM algorithm for the finite mixture can be used to estimate the parameters in such a complicated structure. That is, the latent variables are estimated via the EM algorithm iteratively together with other parameters, including the baseline hazard function. Thus, the joint model in this chapter is a combination of the finite mixture models and the semiparametric model. The finite mixture models and the semiparametric model will be discussed in further sections.

4.2 Finite Mixture Models

Finite mixture models have been used in many fields, for instance biology, genetics, medicine, economics and engineering. The finite mixture approach was used to estimate the probability of developing AIDS after
HIV infection and the distribution of the time interval from infection of HIV to the date of diagnosis with AIDS (Lui et al., 1988; Struthers and Farewell, 1989). Also, many statistical models involve finite mixture distribution in some way or other. The basic idea represents the subgroups within an overall population and we need to allocate observations into groups of homogeneous characteristic and similar observations. The parameters of the specified model are different for each subgroup. Finite mixture models are often used for the purpose of clustering in order to find a homogeneous group among data (Frühwirth-Schnatter, 2006).

Consider \( Y \) as a random vector corresponding to the observed data and assume that the distribution of \( Y \) should be heterogeneous across and homogeneous within subgroups. As a result, \( Y \) has a different probability distribution in each group. Let \( Y = (Y_1, Y_2, \ldots, Y_n) \) be independent and assumed to be from the same parametric family \( P(Y|\theta) \). In this case, parameters \( \theta \) are different across the groups. A discrete variable \( r \) is defined to be a group indicator which takes values in the set \( \{1, 2, \ldots, R\} \). The \( Y \) then follows the distribution \( P(Y|\theta_r) \) where \( \theta_r \) is the parameter in group \( r \). The joint density of \( P(Y, \theta_r) \) is given by

\[
P(Y, \theta_r) = P(Y|\theta_r)P(\theta_r) = P(Y|\theta_r)\pi_r
\]

where \( \pi_r \) represents the proportion or the prior probability that the observation \( Y_i \) belongs to group \( r \) with corresponding density \( P(Y|\theta_r) \). The proportion \( \pi_r \) for each distribution must be non-negative (\( \pi_r \geq 0 \) for \( r = \))
The finite mixture distribution happens when we can observe only the random variable $Y$ but it is not possible to record the group indicator $r$ for each individual with unknown parameters $\theta$. We then immediately have the marginal density $P(Y)$ as shown below

$$P(Y) = \sum_{r=1}^{R} P(Y|\theta_r)\pi_r$$

(4.1)

$$= P(Y|\theta_1)\pi_1 + P(Y|\theta_2)\pi_2 + \ldots + P(Y|\theta_R)\pi_R$$

The likelihood function of the finite mixture distribution can be expressed as

$$L(\theta) = \prod_{i=1}^{n} \sum_{r=1}^{R} P(Y_i|\theta_r)\pi_r$$

(4.2)

**Estimation in the finite mixture models**

Models with too many parameters are easy to lose identifiability. Since the likelihood function form (4.2) is typically complicated, traditional optimization methods do not work. Therefore, the primary approach to estimate the parameters in finite mixture models is the likelihood maximization via the EM algorithm (Dempster et al., 1977; McLachlan and Krishnan, 1997).

The EM algorithm is an efficient iterative procedure to compute the
maximum likelihood estimate in the presence of missing or incomplete data. Each iteration of the EM Algorithm consists of two processes: an expectation step (E-step) and a maximization step (M-step). This process can be applied to the wide range of examples where there is incomplete information such as truncated distributions, latent variable models and mixture models.

- The E-step: the missing data are estimated given the observed data and current estimate of the model parameters, using their expected values.

- The M-step: the likelihood function is maximized under the assumption that the missing data are known. Estimates of the missing data from the E-step are used instead of the actual missing data.

Suppose $\mathbf{Y}$ is a random vector corresponding to the observed data, $\mathbf{Z}$ is a random vector corresponding to the missing information and $\Theta$ is the set of parameters. Then a complete data set exists $(\mathbf{Y}, \mathbf{Z})$ and the actual observed data $\mathbf{Y}$ refers to the incomplete data. In this chapter, we assume a latent variable, $\theta$, to be a discrete variable. The latent variable $\theta$ can take on $R$ known discrete values $\theta_r$ with associated unknown probabilities $\pi_r, r = 1, \ldots, R$. Thus, the product of the likelihoods for the individuals can be expressed as

$$L(\Theta | \mathbf{Y}) = \prod_{i=1}^{n} \left( \sum_{r=1}^{R} P(Y_i | \theta_r, \beta) \pi_r \right) \quad (4.3)$$
Equation (4.3) is called the incomplete data likelihood function and the incomplete data log likelihood function is then

\[ \log(L(\Theta|Y)) = \sum_{i=1}^{n} \log \left( \sum_{r=1}^{R} P(Y_i | \theta_r, \beta) \pi_r \right) \] (4.4)

where the parameters in the model are \( \Theta = (\theta, \beta) \) and \( \pi_r, \quad r = 1, \ldots, R \) with \( \sum_{r=1}^{R} \pi_r = 1 \) (0 < \( \pi_r < 1 \)). This incomplete data log likelihood is the likelihood that we need to maximize. However, it does not have a simple form and it involves the log of a sum, which is inconvenient. Therefore, we need the EM Algorithm. If the incomplete data likelihood is difficult to maximize, it will sometimes be the case that the complete data likelihood is easier to work with.

The complete data likelihood function can be defined as

\[ L(\Theta|Y, Z) = P(Y, Z|\Theta) \]

Let \( Z_{ir} \) be the missing information, \( Z_{ir} = 1 \) if the \( i^{th} \) individual is from the \( r^{th} \) group and 0 otherwise. Thus, \( \sum_r Z_{ir} = 1 \) for all \( i \). The complete data log likelihood becomes
\[
\log L(\Theta | Y, Z) = \log \left[ P(Y, Z | \Theta) \right] \\
= \log \left[ P(Y | Z, \Theta) P(Z | \Theta) \right] \\
= \log \left[ \prod_{r=1}^{R} \prod_{i=1}^{n} \left( P(Y_i | \theta_r, \beta) \times \pi_r \right)^{Z_{ir}} \right] \\
= \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \log \left( P(Y_i | \theta_r, \beta) \right) + \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \log(\pi_r) \quad (4.5)
\]

Then, the expectation of the complete data log likelihood over \(Z\), given the observed data \(Y\) and the current estimates \(\Theta^{(t)}\), becomes

\[
Q(\Theta, \Theta^{(t)}) = E_Z \left[ \log P(Y, Z | \Theta) \ \bigg| \ Y, \Theta^{(t)} \right] \\
= E_Z \left[ \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \log(\pi_r) + \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \log \left( P(Y_i | \theta_r, \beta) \right) \ \bigg| \ Y, \Theta^{(t)} \right] \\
= \sum_{i=1}^{n} \sum_{r=1}^{R} \left[ \log(\pi_r) + \log \left( P(Y_i | \theta_r, \beta) \right) \right] P(Z_{ir} = 1 | Y_i, \Theta^{(t)}) \\
= \sum_{i=1}^{n} \sum_{r=1}^{R} P(Z_{ir} = 1 | Y_i, \Theta^{(t)}) \log(\pi_r) \\
+ \sum_{i=1}^{n} \sum_{r=1}^{R} P(Z_{ir} = 1 | Y_i, \Theta^{(t)}) \log \left( P(Y_i | \theta_r, \beta) \right) \\
\quad (4.6)
\]

The Expectation Step (E-step)

In the E-step we use the current parameter estimates \(\Theta^{(t)}\) to find the expected values of \(Z_{ir}\). The expected value of \(Z_{ir}\), a Bernoulli variable, is the posterior probability of success for individual \(i\) being in group \(r\) given the
observed data. Therefore, using Bayes’ rule, we can compute

\[ \hat{Z}_{ir} = E[Z_{ir}|Y_i, \Theta^{(t)}] = P(Z_{ir} = 1|Y_i, \Theta^{(t)}) \]

\[ = \frac{P(Y_i|Z_{ir} = 1, \Theta^{(t)})P(Z_{ir} = 1|\Theta^{(t)})}{\sum_{g=1}^R P(Y_i|Z_{ig} = 1, \Theta^{(t)})P(Z_{ig} = 1|\Theta^{(t)})} \]

\[ = \frac{\pi_r^{(t)} P(Y_i|\theta_r^{(t)}, \beta^{(t)})}{\sum_{g=1}^R \pi_g^{(t)} P(Y_i|\theta_g^{(t)}, \beta^{(t)})} \] (4.7)

At the E-step of the \( t^{th} \) iteration, the posterior probabilities \( (\hat{Z}_{ir}) \) are calculated.

**The Maximization Step (M-step)**

We maximize equation (4.6) with respect to \( \pi_r \) and \( \Theta = (\theta, \beta) \). Since they are not related, the M-step estimates can be calculated in two parts. The estimates of \( \pi_r \) are easy to compute but the computation of the estimates of \( \Theta \) will typically require iterative numerical methods.

1. To find the estimates of \( \pi_r \), we use the Langrange multiplier \( \psi \) with the constraint that \( \sum_{r=1}^R \pi_r = 1 \) and solve the following equation

\[ \frac{\partial}{\partial \pi_r} \left[ \sum_{i=1}^n \sum_{r=1}^R \log(\pi_r)P(Z_{ir} = 1|Y_i, \Theta^{(t)}) + \psi \left( \sum_{r=1}^R \pi_r - 1 \right) \right] = 0 \]

or

\[ \sum_{i=1}^n \frac{1}{\pi_r} P(Z_{ir} = 1|Y_i, \Theta^{(t)}) + \psi = 0 \] (4.8)
Let
\[ \sum_{i=1}^{n} \frac{P(Z_{ir} = 1| Y_i, \Theta^{(t)})}{n} = \frac{\psi}{n} \]
and summing both sides over \( r \), we obtain
\[ \frac{1}{\pi_r} \left( \pi^*_r \right) = \frac{-\psi}{n} \]
\[ \sum_{r=1}^{R} \pi^*_r = \frac{-\psi}{n} \sum_{r=1}^{R} \pi_r \]
\[ \therefore \psi = -n \]

replacing \( \psi \) into equation (4.8) then we get
\[ \sum_{i=1}^{n} \frac{1}{\pi_r} P(Z_{ir} = 1| Y_i, \Theta^{(t)}) = n \]
\[ \hat{\pi}_r = \frac{1}{n} \sum_{i=1}^{n} P(Z_{ir} = 1| Y_i, \Theta^{(t)}) \]
\[ \hat{\pi}_r = \frac{1}{n} \sum_{i=1}^{n} \hat{Z}_{ir} \]

2. Similarly, to find the estimate of \( \Theta \) in the second part of equation (4.6), the derivative of the second term can be taken with respect to \( \Theta = (\theta, \beta) \). The result is set to zero and solved for \( \Theta \).
4.3. SEMIPARAMETRIC MODEL

Since this part may not necessarily be of closed form, the computation of
the estimate of $\Theta$ means the complete data log likelihood of equation (4.5)
needs numerical maximization. Iteration between two steps (the E-step
and the M-step) is required until we reach convergence. This means the E-
and M-steps are alternated repeatedly until the absolute relative difference
between two consecutive iterations is close to zero. That is,

$$\left| \frac{L(\Theta^{(t+1)}|Y, Z) - L(\Theta^{(t)}|Y, Z)}{L(\Theta^{(t)}|Y, Z)} \right| \approx 0$$

Finally, each observation is assigned to the group having the highest pos-
terior probability that the observation originated from this group.

4.3 Semiparametric Model

In the survival part, we consider the Cox proportional hazards model.
This model is widely used, even though the baseline hazard function is
not specified. The parametric survival model is also available when a spe-
cific probability distribution is assumed for the survival times. We would
prefer to use the parametric model if we were sure of the correct model,
otherwise the Cox model will give reliable enough results (Kleinbaum and
Klein, 2005). Comparison of the parametric survival models and the Cox
model has been investigated in this thesis by using Wald test and stan-
dardized measures of variability (Nardi and Schemper, 2003) for the sur-
vival data only ignoring the ordinal responses. We found that the para-
metric survival models perform more poorly than the Cox model. Therefore, we prefer to use the Cox model to model the survival data in the joint model.

The Cox model is referred to as the semiparametric model because, while the baseline hazard function can take any form of probability distribution, the covariates enter the model linearly. In other words, the hazard function is the product of an unspecified baseline hazard and a parametric function of the covariates. The Cox proportional hazards model is given by

\[ h(t|X, \delta) = h_0(t) \exp(X'\delta) \]  

(4.9)

where \( h_0(t) \) is a baseline hazard function, \( \delta \) is an unknown parameter vector and \( X \) is a covariate vector. The baseline hazard function \( h_0(t) \) can be interpreted as the hazard when all covariate values are zero. According to its semiparametric structure, the baseline hazard function is not specified and the likelihood function of the Cox proportional hazards model does not depend only on \( \delta \) but also on \( h_0(t) \). However, the baseline hazard \( h_0(t) \) is not as easily calculated because it depends on time. We now consider how to estimate the baseline hazard, \( h_0(t) \), which drops out of the partial likelihood. Under the Cox proportional hazards model, the baseline hazard function takes the form of a nonparametric maximum likelihood estimate which is a discrete function. This finding has been shown by Kalbfleisch and Prentice (2002). It has been supported by the work of Nel-
4.3. SEMIPARAMETRIC MODEL

son (1972) and Kaplan and Meier (1958). The nonparametric maximum likelihood estimate of the baseline hazard $h_0(t)$ is also defined in Kiefer and Wolfowitz (1956) which has discrete masses at each failure time.

The survival function for a model with hazard function (4.9) can be defined as

$$S(t|X, \delta) = \exp(-H(t|X, \delta)) \quad (4.10)$$

where $H(t|X, \delta)$ is the cumulative hazard function at time $t$ for an individual with covariate $X$. The cumulative hazard function may be expressed as

$$H(t|X, \delta) = \int_0^t h(t|X, \delta) dt$$

$$= \int_0^t h_0(t) \exp(X'\delta) dt$$

$$= \exp(X'\delta) \int_0^t h_0(t) dt$$

$$= \exp(X'\delta) H_0(t) \text{ or}$$

$$= H_0(t)e^{X'\delta} \quad (4.11)$$

where $H_0(t)$ is the cumulative baseline hazard function. Substituting equation (4.11) into (4.10), we then have the survival function for the general
semiparametric hazard function as

\[ S(t|X, \delta) = \exp(-H_0(t)e^{X\delta}) \]  

(4.12)

The density function is obtained by multiplying the survivor function and the hazard function. We denote the density function corresponding to \( f(t|X, \delta) \) which gives the probability of an individual failed at time \( t \).

\[ f(t|X, \delta) = h(t|X, \delta) \times S(t|X, \delta) \]  

(4.13)

The value of \( f(t|X, \delta) \) is for a non-censored observation. If an individual is still alive at time \( t \), the probability of this event is \( S(t|X, \delta) \). The value of \( S(t|X, \delta) \) is then for a censored observation. Consequently, these two functions become the contribution of each individual to the likelihood function defined by

\[ L(\cdot) = \prod_{i=1}^{n} \left[ f(t_i|X_i, \delta) \right]^{d_i} \times \left[ S(t_i|X_i, \delta) \right]^{1-d_i} \]  

(4.14)

where \( d_i = 0 \) (censored) or 1 (failed). Since \( f(t|X, \delta) = h(t|X, \delta) \times S(t|X, \delta) \) as shown in (4.13), the likelihood function (4.14) becomes

\[ L(\cdot) = \prod_{i=1}^{n} \left[ h(t_i|X_i, \delta) \right]^{d_i} \times \left[ S(t_i|X_i, \delta) \right] \]  

(4.15)

Substituting (4.9) and (4.12) into (4.15), the full likelihood function includ-
4.3. SEMIPARAMETRIC MODEL

ing both $\delta$ and the baseline hazard function $h_0(t)$ can be written as follows

$$L(\cdot) = \prod_{i=1}^{n} \left\{ [h_0(t_i) \exp(X'_i\delta)]^{d_i} \left[ \exp(-H_0(t_i)e^{X'_i\delta}) \right] \right\} \quad (4.16)$$

where

$h_0(t)$ is a positive-valued baseline hazard function.

$H_0(t)$ is a baseline cumulative hazard function.

$d_i$ is an indicator of an individual $i$, $d_i = 1$ if we observe failure time, otherwise $d_i = 0$.

Here we present the method of nonparametric maximum likelihood (NPMLE) described in (Kalbfleisch and Prentice, 2002, section 4.3) and also in (Vaart, 1998, p.403). In this method, we assume the survival time is discrete, let

$h_0(t_i) = \sum_{i=1}^{n} \lambda_i \delta(t - t_i)$ where $\delta(\cdot)$ is the Dirac delta function (see, for example Arfken, 1985, p.481). Suppose $t_i$ are arranged in increasing order $t_1 < t_2 < \ldots < t_n$ and the hazard is zero between adjacent times. The corresponding cumulative hazard function $H_0(t_i) = \sum_{p \leq i} \lambda_p$ will be a step function with jumps at the failure time $t_i$.

Therefore, the full likelihood function (4.16) can be written as

$$L(\lambda, \delta) = \prod_{i=1}^{n} \left\{ [\lambda_i \exp(X'_i\delta)]^{d_i} \left[ \exp(-\sum_{p \leq i} \lambda_p e^{X'_i\delta}) \right] \right\} \quad (4.17)$$
The log likelihood has the following expression

$$
\ell(\lambda, \delta) = \sum_{i=1}^{n} \left[ d_i (\log \lambda_i + X'_i \delta) - \left( \sum_{p \leq i} \lambda_p \right) \exp(X'_p \delta) \right]
$$

(4.18)

Hence, to estimate the baseline hazard \( \lambda_i \), the log likelihood function (4.18) has to be maximized with respect to \( \lambda_i \) and holding \( \delta \) fixed. By differentiating the log likelihood above with respect to \( \lambda_i \) we obtain

$$
\frac{\partial}{\partial \lambda_i} \ell(\lambda, \delta) = \frac{d_i}{\lambda_i} - \sum_{p: p \geq i} \exp(X'_p \delta) = 0,
$$

and

$$
\frac{d_i}{\lambda_i} = \sum_{p: p \geq i} \exp(X'_p \delta)
$$

Thus, the baseline hazard function at time \( t_i \) is estimated by

$$
\hat{\lambda}_i = \frac{d_i}{\sum_{p: p \geq i} \exp(X'_p \delta)}
$$

(4.19)

We will use this idea to calculate the baseline hazard function in the joint model in the next section.

### 4.4 Model Structure

The semiparametric joint model is composed of the item response model and the survival model which is connected by a discrete latent random
variable. The stereotype model is used to model the measure of quality of life while the survival model is used to model time to diagnose the new CDC-B. Additionally, in this thesis the idea of finite mixture models has been used to create the semiparametric joint model. The literature on joint models (for example, Song et al. (2002)) proposed a semiparametric likelihood approach to joint between longitudinal and time-to-event data using the EM algorithm for implementation. Another perspective from McLachlan and McGiffin (1994) illustrated the use of finite mixture models in survival analysis to investigate the distribution of time to re-operation for the degeneration of xenograft values implanted in the aortic position of patients. In this study, the EM algorithm was used to estimate the baseline survival function.

4.4.1 The Stereotype Model

Since the response structure of the questionnaire is ordered, the stereotype model is created to take into account the semiparametric joint model with the discrete latent variable using the structure of finite mixture models when we construct the likelihood function.

We consider $J$ item responses, $Y_1, Y_2, \ldots, Y_J$, which are designed to measure the quality of life with $M$ protocol-specified time points for quality of life data collection, $t_1, t_2, \ldots, t_M$, and each of $Y_i$'s can be written in the form of ordered categorical set $\{1, 2, \ldots, L\}$. The direction of ordering can be recorded from poor to excellent. Subsequently, the stereotype model is
used to model the probabilities of responding on level $\ell$, $\ell = 1, 2, \ldots, L$ for each of the quality of life items and in this study the first category, $\ell = 1$, is treated as the baseline category. Therefore, $Y_{ijm}$ is used to denote the $j^{th}$ item response of the $i^{th}$ subject at the $m^{th}$ time point, $Y_{ijm} = 1, 2, \ldots, L$ for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, J$ and $m = 1, 2, \ldots, M$. The stereotype model with the discrete latent variable can be illustrated as

$$
\log \left[ \frac{P(Y_{ijm} = \ell \mid \theta_r)}{P(Y_{ijm} = 1 \mid \theta_r)} \right] = a_\ell + \phi_\ell (b_j + \theta_r) \quad (4.20)
$$

where

- $a_\ell$ is an item intercept parameter and we set $a_1 = 0$.
- $b_j$ is an item slope coefficient and set $b_1 = 0$.
- $\phi_\ell$ is a score for the outcome level and we consider the ordered stereotype model, $0 = \phi_1 \leq \phi_2 \leq \ldots \leq \phi_L = 1$.
- $\theta_r$ is defined to be a discrete latent random variable or it can be referred to as an average of quality of life score of patients in group $r$ and set $\theta_1 = 0$.

Next, we will construct the likelihood function from the model (4.20) by using the finite mixture idea to further contribute into the joint likelihood function.

Let $Y_i = \{(y_{i111}, \ldots, y_{i11L}), \ldots, (y_{ijm1}, \ldots, y_{ijmL}), \ldots, (y_{iJM1}, \ldots, y_{iJM1L})\}$, then $Y = (Y'_1, \ldots, Y'_n)'$ be a random vector corresponding to the ob-
4.4. MODEL STRUCTURE

served data of $L$ ordered response categories for $i = 1, 2, \ldots, n$ individuals, $j = 1, 2, \ldots, J$ item responses and $m = 1, 2, \ldots, M_i$ time points. Note that $M_i$ is the number of observed time points for individual $i$.

$\Theta$ be all the parameters of the model.

The incomplete data likelihood function (4.3) becomes

$$L(\Theta|Y) = \prod_{i=1}^{n} \left( \sum_{r=1}^{R} \pi_r P(Y_i|\theta_r, \alpha) \right)$$  \hspace{1cm} (4.21)

where

$$P(Y_i|\theta_r, \alpha) = \prod_{m=1}^{M_i} \prod_{j=1}^{J} \prod_{\ell=1}^{L} \left( \frac{\exp(a_{\ell} + \phi_{\ell}(b_j + \theta_r))}{1 + \sum_{k=2}^{L} \exp(a_k + \phi_k(b_j + \theta_r))} \right)^{y_{ijm\ell}}$$  \hspace{1cm} (4.22)

and $\alpha = (a, b, \phi)$. The $y_{ijm\ell}$ represents the binary outcome (“Yes” and “No”) on the response category $\ell$ for the $i^{th}$ subject, $j^{th}$ item and $m^{th}$ time point. Each follow-up time point may have a different number of observations because some patient responses are missing. The latent variable $\theta_r$ in this stereotype model contains all information of quality of life measurement averaged over the visit time.

Thus, the incomplete data likelihood function becomes

$$L(\Theta|Y) = \prod_{i=1}^{n} \left( \sum_{r=1}^{R} \pi_r \prod_{m=1}^{M_i} \prod_{j=1}^{J} \prod_{\ell=1}^{L} \left( \frac{\exp(a_{\ell} + \phi_{\ell}(b_j + \theta_r))}{1 + \sum_{k=2}^{L} \exp(a_k + \phi_k(b_j + \theta_r))} \right)^{y_{ijm\ell}} \right)$$  \hspace{1cm} (4.23)
with the parameters $\Theta = (\alpha, \theta)$ and $\pi$ where $\sum_{r=1}^{R} \pi_r = 1$. Then the incomplete data log likelihood can be shown as

$$
\log(L(\Theta|Y)) = \sum_{i=1}^{n} \log \left( \sum_{r=1}^{R} \pi_r P(Y_i | \theta_r, \alpha) \right)
$$

(4.24)

This incomplete data log likelihood is the likelihood that we need to maximize.

### 4.4.2 The Semiparametric Proportional Hazards Model

We consider the semiparametric proportional hazards model for the survival part in the joint model. In this approach the baseline hazard function, $h_0(t)$, needs to be estimated.

Let $X$ be the time-independent covariate which is the treatment in this study.

$\theta_r$ be a latent discrete random variable which contains the effects of group $r$ on the hazard.

$\delta$ be all the parameters of the survival model.

In a proportional hazards model, it is assumed that the hazard function for the failure time of the $i^{th}$ individual is of the form

$$
\lim_{\Delta t \to 0} \left[ \frac{P(t \leq T_i < (t + \Delta t) | T_i \geq t, X_i, \theta_r)}{\Delta t} \right] = h(t | X_i, \theta_r) = h_0(t) \exp(\theta_r \delta_0 + X_i \delta_1)
$$

(4.25)
where \( P(t \leq T_i \leq t + \Delta t | T_i \geq t) = P(\text{individual fails in the interval } [t, t + \Delta t] | \text{survival up to time } t) \). This is defined as the probability of failure during a very small time interval, assuming that the individual has survived to the beginning of the interval, or as the limit of probability that an individual fails in a very short interval \( t + \Delta t \), given that the individual has survived to time \( t \).

Then, the full likelihood function (4.14) becomes

\[
L(\cdot) = \prod_{i=1}^{n} \left\{ \left( h(t_i | X_i, \theta_r) S(t_i | X_i, \theta_r) \right)^{d_i} \times \left( S(t_i | X_i, \theta_r) \right)^{1-d_i} \right\}
\]

\[
= \prod_{i=1}^{n} \left\{ \left( h(t_i | X_i, \theta_r) \right)^{d_i} \times \left( S(t_i | X_i, \theta_r) \right) \right\}
\]

\[
= \prod_{i=1}^{n} \left\{ \left( h_0(t) \exp(\theta_r \delta_0 + X_i \delta_1) \right)^{d_i} \times \exp \left( -H_0(t) \exp(\theta_r \delta_0 + X_i \delta_1) \right) \right\}
\]

(4.26)

where \( h_0(t) \) is a positive-valued baseline hazard function.

\( H_0(t) \) is a baseline cumulative hazard function.

Let \( h_0(t_i) = \sum_{i=1}^{n} \lambda_i \delta(t - t_i) \) where \( \delta(\cdot) \) is the Dirac delta function and \( H_0(t_i) = \sum_{p \leq i} \lambda_p \). The probability distribution function (4.26) becomes

\[
L(\cdot) = \prod_{i=1}^{n} \left\{ \left( \lambda_i \exp(\theta_r \delta_0 + X_i \delta_1) \right)^{d_i} \times \exp \left( -\sum_{p \leq i} \lambda_p \exp(\theta_r \delta_0 + X_i \delta_1) \right) \right\}
\]

(4.27)
This probability distribution is composed of parametric and nonparametric components. In our example, the $X_i$ is the treatment (Continuation anti-retroviral treatment, or CD4-guided interruption treatment). In general, the model can have more than one covariate. The parameters in the model include $\lambda_i$, $\delta_0$ and $\delta_1$. The $\lambda_i$ is a nuisance parameter which is treated nonparametrically (Begun et al., 1983). The parametric part is the exponential function of the unknown regression coefficient ($\delta_0$ and $\delta_1$).

### 4.4.3 The Full Likelihood Function

The joint likelihood function is obtained from combining item response part and survival part together as follows:

Let $P(Y_i | \theta_r, \alpha)$ be the probability distribution for the stereotype model specified by

$$
P(Y_i | \theta_r, \alpha) = \prod_{m=1}^{M_i} \prod_{j=1}^{J} \prod_{\ell=1}^{L} \left( \frac{\exp(\alpha_{\ell} + \phi_{\ell}(b_j + \theta_r))}{1 + \sum_{k=2}^{L} \exp(\alpha_k + \phi_k(b_j + \theta_r))} \right)^{y_{ijm\ell}} \tag{4.28}
$$

$P(T_i, D_i | \lambda_i, \theta_r, \delta)$ be the probability distribution for the survival model defined as

$$
P(T_i, D_i | \lambda_i, \theta_r, \delta) = (\lambda_i \exp(\theta_r \delta_0 + X_i \delta_1))^{d_i} \times \exp \left( - \sum_{p \leq i} \lambda_p \exp(\theta_r \delta_0 + X_i \delta_1) \right) \tag{4.29}
$$

$\pi_r$ be the unknown probabilities, $r = 1, \ldots, R$

$\Theta$ be all the unknown parameters of the model.
4.4. MODEL STRUCTURE

The incomplete data likelihood function is the product of the likelihoods for the individuals

\[ L(\Theta | Y, T, D) = \prod_{i=1}^{n} \left( \sum_{r=1}^{R} P(Y_i | \theta_r, \alpha) P(T_i, D_i | \lambda_i, \theta_r, \delta) \pi_r \right) \]

and the incomplete data log likelihood function is then

\[
\log(L(\Theta | Y, T, D)) = \sum_{i=1}^{n} \log \left( \sum_{r=1}^{R} P(Y_i | \theta_r, \alpha) P(T_i, D_i | \lambda_i, \theta_r, \delta) \pi_r \right)
\]

(4.30)

Let \( Z_{ir} \) be the missing information, \( Z_{ir} = 1 \) if the \( i^{th} \) individual was from the \( r^{th} \) group and 0 otherwise. The complete data log likelihood becomes

\[
\log(L(\Theta | Y, T, D, Z)) = \log \left[ P(Y, T, D, Z | \Theta) \right]
\]

\[
= \log \left[ P(Y, T, D | Z, \Theta) P(Z | \Theta) \right]
\]

\[
= \log \left[ \prod_{i=1}^{n} \prod_{r=1}^{R} \left( P(Y_i | \theta_r, \alpha) P(T_i, D_i | \lambda_i, \theta_r, \delta) \pi_r \right)^{Z_{ir}} \right]
\]

\[
= \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \log \left( P(Y_i | \theta_r, \alpha) P(T_i, D_i | \lambda_i, \theta_r, \delta) \pi_r \right)
\]

\[
= \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \log(\pi_r) + \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \log \left( P(Y_i | \theta_r, \alpha) \right)
\]

\[
+ \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \log \left( P(T_i, D_i | \lambda_i, \theta_r, \delta) \right) \quad (4.31)
\]

where \( P(Y_i | \theta_r, \alpha) \) and \( P(T_i, D_i | \lambda_i, \theta_r, \delta) \) are defined in equation (4.28) and (4.29) respectively.
4.5 Parameter Estimation

The incomplete data log likelihood (4.30) is the likelihood that we need to maximize. However, it does not have a simple form and it involves the log of summation, which is inconvenient. It may be also impossible to directly maximize this log likelihood in the semiparametric setting due to the nonparametric component. If the incomplete data likelihood is difficult to work with, it will sometimes be the case that the complete data likelihood (4.31) can be easier. The EM Algorithm can be employed readily to estimate the parameters in the joint model. Similar to equation (4.6), given the observed data \( Y, T, D \) and the current estimates \( \Theta^{(t)} \), the expectation of the complete data log likelihood is

\[
Q(\Theta, \Theta^{(t)}) = E[Z \log P(Y, T, D, Z|\Theta) | Y, T, D, \Theta^{(t)}]
\]

\[
= E[Z \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \log(\pi_r) + \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \log \left( P(Y_i|\theta_r, \alpha) \right) + \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \log \left( P(T_i, D_i|\lambda_i, \theta_r, \delta) \right) | Y, T, D, \Theta^{(t)}]
\]

\[
= \sum_{i=1}^{n} \sum_{r=1}^{R} \left[ \log(\pi_r) + \log \left( P(Y_i|\theta_r, \alpha) \right) + \log \left( P(T_i, D_i|\lambda_i, \theta_r, \delta) \right) \right]
\times P(Z_{ir} = 1 | Y_i, T_i, D_i, \Theta^{(t)})
\]
4.5. PARAMETER ESTIMATION

\[ = \sum_{i=1}^{n} \sum_{r=1}^{R} P(Z_{ir} = 1|Y_i, T_i, D_i, \Theta^{(t)}) \log(\pi_r) \]
\[ + \sum_{i=1}^{n} \sum_{r=1}^{R} P(Z_{ir} = 1|Y_i, T_i, D_i, \Theta^{(t)}) \log(P(Y_i | \theta_r, \alpha)) \]
\[ + \sum_{i=1}^{n} \sum_{r=1}^{R} P(Z_{ir} = 1|Y_i, T_i, D_i, \Theta^{(t)}) \log(P(T_i, D_i | \lambda_i, \theta_r, \delta)) \tag{4.32} \]

4.5.1 Baseline Hazard Estimation

Before starting the EM-step, we first consider the third part of the equation (4.32) which is the survival part. This part is composed of the baseline hazard function and can be separately written in terms of log likelihood function as

\[
\ell(\lambda, \theta, \delta) = \sum_{i=1}^{n} \sum_{r=1}^{R} P(Z_{ir} = 1|Y_i, T_i, D_i, \Theta^{(t)}) \left\{ d_i \left( \log \lambda_i + \theta_r \delta_0 + X_i \delta_1 \right) \right. \\
- \left. \sum_{p \leq i} \lambda_p \exp(\theta_r \delta_0 + X_i \delta_1) \right\} \tag{4.33}
\]

As can be seen, this component has the baseline hazard \( \lambda_i \) and the update of the hazard at the failure time for the \( i^{th} \) individual depends on \((\theta, \delta)\). Thus, we need to determine the profile likelihood function using the same idea that we used to obtain the result in (4.19) in the previous section. That is, the log likelihood (4.33) has to be maximized with respect to \( \lambda_i \).
and holding \((\theta, \delta)\) fixed. We obtain

\[
\frac{\partial}{\partial \lambda_i} \ell(\lambda, \theta, \delta) = \sum_{r=1}^{R} P(Z_{ir} = 1|Y_i, T_i, D_i, \Theta^{(t)}) \frac{d_i}{\lambda_i} - \sum_{p \geq i} \sum_{r=1}^{R} P(Z_{pr} = 1|Y_p, T_p, D_p, \Theta^{(t)}) \exp(\theta_r \delta_0 + X_p \delta_1) = 0
\]

then

\[
\sum_{r=1}^{R} P(Z_{ir} = 1|Y_i, T_i, D_i, \Theta^{(t)}) \frac{d_i}{\lambda_i} = \sum_{p \geq i} \sum_{r=1}^{R} P(Z_{pr} = 1|Y_p, T_p, D_p, \Theta^{(t)}) \exp(\theta_r \delta_0 + X_p \delta_1).
\]

Using \(\sum_{r=1}^{R} P(Z_{ir} = 1|Y_i, T_i, D_i, \Theta^{(t)}) = 1\) we obtain

\[
\frac{d_i}{\lambda_i} = \sum_{p \geq i} \sum_{r=1}^{R} P(Z_{pr} = 1|Y_p, T_p, D_p, \Theta^{(t)}) \exp(\theta_r \delta_0 + X_p \delta_1)
\]

\[
\hat{\lambda}_i = \frac{d_i}{\sum_{p \geq i} \sum_{r=1}^{R} P(Z_{pr} = 1|Y_p, T_p, D_p, \Theta^{(t)}) \exp(\theta_r \delta_0 + X_p \delta_1)} \quad (4.34)
\]

This \(\hat{\lambda}_i\) will be used in the E-step and the M-step.

The expectation of the complete data log likelihood will be calculated in the E-step. The M-step consists of two parts since we maximize equation (4.32) with respect to \(\pi_r\) and \(\Theta = (\theta, \alpha, \delta)\). Due to the fact that there is no relationship between \(\pi_r\) and \(\Theta\), the M-step estimates can be calculated separately. The estimates of \(\pi_r\) are easy to compute but the computation of estimates of parameters \(\Theta\) will typically require iterative numerical methods. The E and M steps will be iteratively estimated together with
the baseline hazard function which is replaced by its nonparametric maximum likelihood.

### 4.5.2 The E-step

In the E-step, we use the current parameter estimates $\Theta^{(t)}$ to find the expected values of $Z_{ir}$ of the complete data log likelihood. The expected values of a Bernoulli distribution are the probability of success for individual $i$ being in group $r$ given the observed data. Therefore, using Bayes’ rule, we can compute

$$
E[Z_{ir} | Y_i, T_i, D_i, \Theta^{(t)}] = P(Z_{ir} = 1 | Y_i, T_i, D_i, \Theta^{(t)})
$$

$$
= \frac{P(Y_i | Z_{ir} = 1, \Theta^{(t)}) P(T_i, D_i | Z_{ir} = 1, \Theta^{(t)}) P(Z_{ir} = 1 | \Theta^{(t)})}{\sum_{g=1}^{R} P(Y_i | Z_{ig} = 1, \Theta^{(t)}) P(T_i, D_i | Z_{ig} = 1, \Theta^{(t)}) P(Z_{ig} = 1 | \Theta^{(t)})}
$$

$$
= \frac{\pi_r^{(t)} P(Y_i | \theta_r^{(t)}, \alpha^{(t)}) P(T_i, D_i | \lambda_i, \theta_r^{(t)}, \delta^{(t)})}{\sum_{g=1}^{R} \pi_g^{(t)} P(Y_i | \theta_g^{(t)}, \alpha^{(t)}) P(T_i, D_i | \lambda_i, \theta_g^{(t)}, \delta^{(t)})}
$$

Thus, the posterior class membership probabilities for the $i^{th}$ individual are defined by

$$
P(Z_{ir} = 1 | Y_i, T_i, D_i, \Theta^{(t)}) = \frac{\pi_r^{(t)} P(Y_i | \theta_r^{(t)}, \alpha^{(t)}) P(T_i, D_i | \lambda_i, \theta_r^{(t)}, \delta^{(t)})}{\sum_{g=1}^{R} \pi_g^{(t)} P(Y_i | \theta_g^{(t)}, \alpha^{(t)}) P(T_i, D_i | \lambda_i, \theta_g^{(t)}, \delta^{(t)})}
$$

(4.35)
4.5.3 The M-step

In the M-step, we maximize equation (4.32) with respect to $\pi_r$ and $\Theta = (\theta, \alpha, \delta)$

1. We follow the argument on pages 99-100 to find the estimates of $\pi_r$, then we obtain

$$\sum_{i=1}^{n} \frac{1}{\pi_r} P(Z_{ir} = 1 | Y_i, T_i, D_i, \Theta^{(t)}) = \frac{n}{\hat{\pi}_r} = \frac{1}{n} \sum_{i=1}^{n} P(Z_{ir} = 1 | Y_i, T_i, D_i, \Theta^{(t)})$$

2. The second and third parts of equation (4.32) are then numerically maximized with respect to $\Theta = (\theta, \alpha, \delta)$. The former is the item response part and the latter is the survival part. Throughout this process, the stereotype likelihood function needs to be optimized iteratively. Because of the complicating factor of being multiplicative in parameters and the difficulty in fitting ordering constraints on the scores $\{\phi_k\}$, an alternating algorithm (Greenland, 1994) is used to estimate the parameters in the model. That is, $\{\phi_k\}$ and $(b_j + \theta_r)$ are alternately held fixed while the other is estimated. The update of parameters in the full likelihood survival function is also difficult because $\lambda_i$ is non parametric. However, we use the two-step iterative method to estimate all parameters in both models simultaneously.

The proposed iterative methods to maximize the likelihood function
are described as follows:

2.1. Updating all parameters in $\Theta$ simultaneously with fixed $\phi_\ell$ and treating $\lambda_i$ as a nuisance parameter using ordinary maximum likelihood approach. In this step, $\phi_\ell$ is treated as a known predictor.

2.2. Treating the estimated $\Theta$ from the previous step as fixed ($(b_j + \theta_r)$ being predictor) and $\phi_\ell$ is treated as unknown parameter. Updating $\phi_\ell$ also uses the maximum likelihood estimation.

2.3. Calculating $\lambda_i$ by plug-in the estimated parameters from the previous steps (2.1. and 2.2.) into equation (4.34).

2.4. Iterating the alternative fixing process until convergence.

The estimated parameters from the M-step are returned into the E-step until convergence.

The standard errors of the solution $\hat{\Theta}$ obtained from the EM algorithm, can be estimated based on the observed information matrix. This requires the Hessian matrix, which can be achieved by taking the second-order partial derivatives of the log likelihood function. For example, with three parameters $\theta_1, \theta_2, \theta_3$, the Hessian is

$$
H = \begin{pmatrix}
\frac{\partial^2 \ell}{\partial \theta_1^2} & \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_3} \\
\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 \ell}{\partial \theta_2^2} & \frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_3} \\
\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_3} & \frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_3} & \frac{\partial^2 \ell}{\partial \theta_3^2}
\end{pmatrix}
$$
Taking the negative of the Hessian matrix yields the observed information matrix and then inverting this matrix produces a matrix containing the variances of the parameters on its diagonal and the asymptotic covariances of the parameters in the off-diagonal position. The square root of the diagonal elements is then the standard errors.

However, when estimating the standard error of $\hat{\phi}_\ell$, we encountered the difficulty that the standard error of $\hat{\phi}_\ell$ was not straightforward to calculate. This is because we treated $\phi_\ell$ as the ordered constraint and we used the two-step fitting approach to evaluate the models. This means while $\phi_\ell$ and $b_j$ were alternately fixed, in fact $\phi_\ell$ was also estimated in one step of each iteration. Greenland (1994) indicated that this estimation can cause incorrect estimates of the standard errors as it ignores the sampling variability due to the estimation of the other group of parameters. For this reason, he recommended using the bootstrap to estimate valid standard errors. Holtbrugge and Schumacher (1991) demonstrated an iterative reweighted least square (IRLS) for actual estimation. Direct maximization of the likelihood function with a quasi-Newton method was presented by Kuss (2006) to solve this problem. This approach depends on the first and second derivatives of the likelihood function that uses finite difference methods for the first derivatives and the approximated Hessian matrix is gradually built up as the iteration proceed. Although the three different methods above give standard errors that are valid in principle, the inferences from the standard errors are still not recommended when
the scores, $\phi_t$, are estimated (Greenland, 1994).

4.6 Model Comparison

Akaike’s information criterion (AIC) is the most widespread information criterion which compares all candidate models at once and does not require nested alternative models (Akaike, 1973). AIC is defined as

$$AIC = -2 \log L(\cdot) + 2k$$

where $L(\cdot)$ is the likelihood and $k$ is the number of parameters in the model. The small values represent a better overall fit with all information criteria. However, Xu et al. (2006) proposed the associated criterion of AIC that is the profile Akaike information (pAIC) for models with nuisance parameters. They developed an AIC using the profile likelihood function. Only the parameters of interest are selected for pAIC calculation. In this study, we use the pAIC for comparing models since we have nuisance parameter, $\lambda$, in the joint model. We will discuss the idea of pAIC below.

Let our semiparametric joint model be parameterized by $\Theta$ and $\lambda$, where $\Theta$ is the set of parameters of the stereotype model and the Cox proportional hazards model but the baseline hazard $\lambda$ is not included. Then $\Theta = (a, b, \phi, \theta, \delta)$ is the parameter of interest and the baseline hazard $\lambda$ is the nuisance parameter. The pAIC approach is interested in selecting
the parameter of interest $\Theta$, while leaving the nuisance parameter $\lambda$ the same across all competing models. This idea is similar to Claeskens et al. (2003). As a result, the semiparametric joint model now is indexed by $\Theta$ alone. Obviously, if $\Theta$ and $\lambda$ have a finite dimension $p$ and $q$ respectively, the AIC for $(\Theta, \lambda)$ is $-2 \log L(\cdot) + 2(p + q)$. On the contrary, the pAIC for $\Theta$ is $-2 \log L(\cdot) + 2p$.

Therefore, the profile AIC can be expressed as

$$pAIC = -2 \log L(\cdot) + 2p$$

where $L(\cdot)$ is the likelihood and $p$ is the number of parameters of interest in the model. The smaller pAIC is preferred as with the standard of AIC. (see Xu et al. (2009) for more details about a theorem and proof).
Chapter 5

Data Application

5.1 Introduction

In this study we analyze data from the Staccato study (Anaworaphich and the Staccato study group, 2006) which is a randomized trial of two treatment methods, the continuous anti-retroviral treatment compared to the CD4-guided interruption treatment.

Method 1: Continuation anti-retroviral treatment (control): Drugs are continued or changed according to current guidelines and good clinical practice.

Method 2: CD4-guided interruption treatment: Drugs are discontinued and reintroduced according to CD4 count, with Highly Active Antiretroviral Therapy (HAART) being administered only if CD4 count is $< 350 \text{ cells/µl}$. Patients on HAART, with viremia below 50 copies/ml, CD4 count above
350 cells/µl and no evidence for pre-existing drug resistance were randomized to one of two methods. Note that the most common measure used to assess immunological health of a HIV patient is the CD4 cell count. Higher CD4 cell counts indicate a strong immune system that is more prepared to resist infection. Lower CD4 cell counts indicate a higher risk of CDC-C class (details of CDC-C will be given below). Viral load is a measure of the amount of virus in the blood plasma. A lower viral load is preferable and may indicate successful treatment of the disease. These two markers are often used to evaluate a patient’s success on treatment. When a patient begins a successful treatment regimen, the viral load may drop drastically and fall below a detectable level. The CD4 cell count may take longer to respond or may not respond at all. As viral load decreases, we may expect the CD4 cell count to increase as the immune system has time to recover.

**Clinical Status**

CDC is the classification system of HIV disease stage by CD4 count and by the presence of specific HIV-related conditions. The definition of AIDS includes all HIV-infected individuals with CD4 count of less than 200 cells/µL (or CD4 percentage < 14%) as well as those with certain HIV-related conditions and symptoms. Although the fine points of the classification system are rarely used in the routine clinical management of HIV-infected patients, a working knowledge of the staging criteria (in particular the definition of AIDS) is useful in patient care. In addition, the
CDC system is used in clinical and epidemiological research.

The clinical categories of HIV infection (Centers for Disease Control and Prevention, 1992) are defined as follows:

**Category A** (Asymptomatic, Acute HIV, or Persistent generalized lymphadenopathy): consists of one or more of the conditions listed below in an adolescent or adult (greater than or equal to 13 years) with documented HIV infection. Conditions listed in Categories B and C must not have occurred. The conditions are:

- Asymptomatic HIV infection
- Persistent generalized lymphadenopathy
- Acute (primary) HIV infection with accompanying illness or history of acute HIV infection.

**Category B** (Symptomatic conditions): consists of symptomatic conditions in an HIV-infected adolescent or adult that are not included among conditions listed in clinical Category C and that meet at least one of the following criteria:

- the conditions are attributed to HIV infection or are indicative of a defect in cell-mediated immunity.
- the conditions are considered by physicians to have a clinical course or to require management that is complicated by HIV infection.
**Category C** (AIDS-indicator conditions): includes the clinical conditions listed in the AIDS surveillance case definition (Centers for Disease Control and Prevention, 1987).

For classification purposes, Category B conditions take precedence over those in Category A. For example, someone who previously had a condition that once met the criteria for Category B but now is asymptomatic should be classified in clinical Category B. For classification purposes, once a Category C condition has occurred, the person will remain in Category C.

We are interested in only the clinical events CDC-B at the end of the randomized treatment period as the survival endpoint in the survival model in our study. This is because there are very few patients who experience CDC-C at the end of the study (five out of 354). The patients in category-C are then treated as censored observations.

**Medical Outcome Study HIV Health Survey (MOS-HIV)**

Throughout this study, a randomized trial of two treatment methods continued for an average of approximately three years, and this was followed by a period of 12 to 24 weeks continuous treatment, for patients in both arms. The patient’s quality of life was measured at various points during this period of study by using the Medical Outcome Study HIV Health Survey (MOS-HIV) questionnaire (Wu et al., 1991, 1997). Patients were asked to complete the validated Thai version of MOS-HIV questionnaire every 24 weeks beginning at the baseline visit, and at weeks 24, 48,
72, 96, 120 and 144 respectively. There is no information about quality of life after 144 weeks, even though patients were followed up for survival. Patients were also monitored longitudinally with respect to the biologic endpoints throughout the follow-up period, such as CD4 count and HIV viral load.

The MOS-HIV questionnaire was developed to assess aspects of functional status and well-being in HIV-infected patients in clinical trials. It has been extensively used in clinical trials and is available in 20 languages. Among other languages, the MOS-HIV questionnaire has been translated into Thai and used in several studies (Ichikawa and Natpratan, 2004; Ananworanich et al., 2005; Cardiello et al., 2005; Nüesch et al., 2009; Charilyalert-sak et al., 2011). In the Staccato study (Nüesch et al., 2009), quality of life scores were calculated by dividing them into 10 subscores ranging from 0 to 100 (greater value indicates better quality of life). Then the summary scores were transformed to standardized scores, with a mean of 50 and a standard deviation of 10. A similar approach, which treated the scores as continuous values vary from 0 to 100, is used by Wachtel et al. (1992) and Ciconelli et al. (1999). However, since the quality of life scales tend to generate discrete, asymmetrical and limited distributions, the traditional methods such as t-test and linear regression that assume the standard normal assumption may not be appropriate (Abreu et al., 2008). By contrast, in this study we use the raw scores which are derived from the questionnaire since we consider the ordinal nature of the data. The ordinary
rank, for example, corresponds to the following options: poor(1), fair(2), good(3), very good(4) and excellent(5). Some questions have three outcome categories; the ordinal ranking is defined as: poor(1), good(2), and excellent(3). We use all questions to contribute to the stereotype models as described in Chapter 2 and Chapter 3. Although this causes a multidimensional variable in the model, the more questions we include, the more accurate results we obtain. However, we use a latent variable to represent the true scores of the quality of life. The latent variable is not directly observable but we can estimate it from the observed variables through their characteristics. In other words, the latent variable contains all the information from the observations of the quality of life. Moreover, the principle of the latent variable is to express the distribution of the observed variables in terms of a small number of latent variables. This is supported by Moustaki (2000), who reported that the number of latent variables typically is less than the number of observed variables.

In this chapter, we apply the methodologies to data from the Staccato study by using both a Bayesian joint model (Chapter 3) and a semiparametric joint model (Chapter 4). According to the model selection and MCMC diagnostics, we construct a number of models for the Bayesian joint model approach with different covariates in the survival part and different initial values for each model. We also construct a number of models for the semiparametric joint model approach with different covariates in the survival part and a different number of groups for the purpose of
5.2. DATA ANALYSIS: BAYESIAN JOINT MODEL

model comparison. However, because each model contains many parameters, we do not have space to report all models. In this chapter we will illustrate only one Bayesian joint model for an interpretation and illustrate some parameters for the MCMC diagnostics. Similarly, for the semiparametric joint model we will present the pAIC of the candidate models and the estimated values of the selected model that has the smallest pAIC. The summary of all fitted models for the Bayesian joint model approach is shown in Appendix A and the semiparametric joint model approach is shown in Appendix B.

5.2 Data Analysis: Bayesian Joint Model Approach

We analyze three different models. The item response part and the regression part are the same for all models but the survival part is different in terms of the covariates. The first model has quality of life and treatment as covariates, the second model has quality of life, treatment and initial CD4 count as covariates and the third model has quality of life, treatment, initial CD4 count and an interaction between treatment and initial CD4 count as covariates. We report here only two models since the interaction term in the third model is not significant.

1. Model 1: This model has two covariates which are quality of life and treatment in the survival part. The item response part consists
of two stereotype models and a binary model. The stereotype model can be expressed as

\[
\log \left[ \frac{P(Y_{ijm} = \ell | \theta_i(t_m))}{P(Y_{ijm} = 1 | \theta_i(t_m))} \right] = a^L \ell + \phi^L j i(t_m)
\]

Because the numbers of response categories are not the same for all questions, we created two stereotype models to model all questions together by allowing parameters \(a_\ell\) and \(\phi_\ell\) to be different but share the same \(\theta_i(t_m)\). The first stereotype model is used to model the set of questions that has five ordinal categories \((L = 5)\). The other one is used to model the set of questions that has three ordinal categories \((L = 3)\). The binary model is used to model the set of questions that has two categories and such a model can be shown as

\[
\log \left[ \frac{P(Y_{iqm} = 1)}{1 - P(Y_{iqm} = 1)} \right] = a^* + b^* q i(t_m)
\]

The Cox proportional hazards model is used in the survival process. There are two covariates in Model 1, so the model has the form

\[
h(t|\theta_i^*, X_i) = h_0(t) \exp(\theta_i^* \delta_0 + X_i \delta_1)
\]

The baseline hazard \(h_0(t)\) is excluded from the Bayesian joint model. In this model, \(X\) refers to treatment, \(X = 1\) for the continuation anti-retroviral treatment and \(X = 0\) for the CD4-guided interruption treatment.
We generate the latent variable $\theta_i = (\theta_i(t_1), \ldots, \theta_i(t_M))'$ to link the quality of life and survival time together. We assume that $\theta_i \sim MVN(\mu(\theta_i), \Sigma(\rho))$ with $\mu(\theta_i(t_m)) = \beta_0(t_m) + \beta_1(t_m)X_i$ and the elements of $\Sigma(\rho)$ follow a simple form of covariance parameterization as $\text{cov}(\theta_{tm}, \theta_{tm'}) = \rho^{|t_m - t_{m'}|}$.

There are 68 parameters in Model 1, which consist of 51 parameters in the item response part, 15 parameters in the regression part and 2 parameters in the survival part.

2. **Model 2**: This model has three covariates which are quality of life, treatment and initial CD4 count in the survival part. As mentioned above, the difference between Model 1 and Model 2 is the covariates in the survival part. Thus, the Cox proportional hazards model can be written as

$$h(t|\theta_i^*, X_{1i}, X_{2i}) = h_0(t) \exp(\theta_i^* \delta_0 + X_{1i} \delta_1 + X_{2i} \delta_2)$$

In this model, $X_1$ refers to treatment, $X_1 = 1$ for the continuation anti-retroviral treatment and $X_1 = 0$ for the CD4-guided interruption treatment. The covariate $X_2$ refers to initial CD4 count, $X_2 = 1$ for the initial CD4 count $\geq 350$ and $X_2 = 0$ for the initial CD4 count $< 350$. There are 69 parameters in Model 2, which consist of 51 parameters in the item response part, 15 parameters in the regression part and 3 parameters in the survival part.
5.2.1 Convergence Diagnostics

5.2.1.1 Graphical Diagnostics: Trace and Density Plots

Because there are many parameters in each model, we illustrate the trace and density plots for some parameters only. The other parameters are shown in Appendix A.

![Trace and Density Plots of $b_1$ and $b_2$ of model 1](image)

Figure 5.1: Trace and Density Plots of $b_1$ and $b_2$ of model 1

Figure 5.1 shows two common visual diagnostics: trace plots and density plots of parameters $b_1$ and $b_2$ of the stereotype model. From the trace plots, the chains start with an initial value which is a considerable distance from the target distribution. As can be seen from the figure, the first few thousand iterations should be discarded. After that the chains appear to
be mixing very well and have settled into a stable running mean.

Figure 5.2 provides the trace and density plots of parameters $\delta_0$ and $\delta_1$ of the Cox proportional hazards model. Both chains are mixing well with very small fluctuations and retain the same pattern along a period of time. This indicates that the chains could have reached the target distribution quickly.

![Trace and Density Plots](image)

Figure 5.2: Trace and Density Plots of $\delta_0$ and $\delta_1$ of model 1

The density plots show an indication of the mean and spread of the values obtained from the sampler. As can be seen from the second column in figure 5.1 and 5.2, the mean appears to be around the value 3, 2.2, -0.2 and -1 for parameters $b_1$, $b_2$, $\delta_0$ and $\delta_1$ respectively. These are similar to
the center of the chains from the trace plots that appear to be close to such values as well. A multimodality of the density estimate is a classic sign of nonconvergence for most applications (Gill, 2008), but this is not apparent here.

5.2.1.2 Statistical Diagnostics

We illustrate the statistical diagnostics of Model 1 only in this section. The other models diagnostics are shown in Appendix A.

Gelman and Rubin’s Multiple Sequence Diagnostic

We illustrate the Gelman and Rubin diagnostic (Gelman and Rubin, 1992) with four separate chains from several different starting points. This is because we need to show that the Markov chains converge to the same target distribution regardless of the starting point. Thus, it would be best to select different initial values for all four chains to avoid a local maximum problem. Each chain is run for 100,000 iterations with the first 40,000 iterations discarded. The coda package in R is used to produce the following results (Table 5.1), which show no evidence of nonconvergence.
### Table 5.1: Gelman and Rubin Diagnostic, Model 1

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Point est. ((\hat{R}))</th>
<th>97.5% quantile</th>
<th>Multivariate psrf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereotype Model (5 levels)</td>
<td>(a_2)</td>
<td>1.19</td>
<td>1.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(a_3)</td>
<td>1.19</td>
<td>1.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(a_4)</td>
<td>1.21</td>
<td>1.71</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(a_5)</td>
<td>1.15</td>
<td>1.52</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>(b_1)</td>
<td>1.01</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_2)</td>
<td>1.01</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_3)</td>
<td>1.00</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_4)</td>
<td>1.01</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_5)</td>
<td>1.01</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_6)</td>
<td>1.02</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_7)</td>
<td>1.01</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_8)</td>
<td>1.03</td>
<td>1.12</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>(b_9)</td>
<td>1.01</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_{10})</td>
<td>1.03</td>
<td>1.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_{11})</td>
<td>1.01</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_{12})</td>
<td>1.02</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_{13})</td>
<td>1.01</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_{14})</td>
<td>1.01</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_{15})</td>
<td>1.01</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_{16})</td>
<td>1.01</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b_{17})</td>
<td>1.03</td>
<td>1.12</td>
<td></td>
</tr>
</tbody>
</table>

*continued on next page*
<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Point est. $(\hat{R})$</th>
<th>97.5% quantile</th>
<th>Multivariate psrf</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b_{18}$</td>
<td>1.01</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{19}$</td>
<td>1.01</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{20}$</td>
<td>1.01</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{21}$</td>
<td>1.02</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{22}$</td>
<td>1.02</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{23}$</td>
<td>1.01</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{24}$</td>
<td>1.02</td>
<td>1.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{25}$</td>
<td>1.01</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{26}$</td>
<td>1.01</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{27}$</td>
<td>1.01</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{28}$</td>
<td>1.00</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{29}$</td>
<td>1.01</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{30}$</td>
<td>1.00</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{31}$</td>
<td>1.01</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{32}$</td>
<td>1.01</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\phi_2$</td>
<td>1.12</td>
<td>1.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\phi_3$</td>
<td>1.13</td>
<td>1.46</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>$\phi_4$</td>
<td>1.27</td>
<td>1.94</td>
<td></td>
</tr>
<tr>
<td>Stereotype Model</td>
<td>$a_2$</td>
<td>1.03</td>
<td>1.08</td>
<td>1.01</td>
</tr>
<tr>
<td>(3 levels)</td>
<td>$a_3$</td>
<td>1.02</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>$L = 3$ levels</td>
<td>$b_1$</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$J = 6$ questions</td>
<td>$b_2$</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_3$</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_4$</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_5$</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

*continued on next page*
5.2. *DATA ANALYSIS: BAYESIAN JOINT MODEL*

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Point est. $\hat{\beta}$</th>
<th>97.5% quantile</th>
<th>Multivariate psrf</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b_6$</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\phi_2$</td>
<td>1.01</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>Binary Model</td>
<td>$a$</td>
<td>1</td>
<td>1.02</td>
<td>-</td>
</tr>
<tr>
<td>$Q = 2$ questions</td>
<td>$b_1$</td>
<td>1</td>
<td>1.01</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$b_2$</td>
<td>1</td>
<td>1.00</td>
<td>1</td>
</tr>
<tr>
<td>Regression Model</td>
<td>$\rho$</td>
<td>1</td>
<td>1.02</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_0)$</td>
<td>1.00</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_1)$</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_2)$</td>
<td>1.00</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_3)$</td>
<td>1.02</td>
<td>1.07</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_4)$</td>
<td>1.01</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_5)$</td>
<td>1.07</td>
<td>1.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_6)$</td>
<td>1.02</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_0)$</td>
<td>1.00</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_1)$</td>
<td>1.00</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_2)$</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_3)$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_4)$</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_5)$</td>
<td>1.02</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_6)$</td>
<td>1.01</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td>Cox’s Model</td>
<td>$\delta_0$</td>
<td>1.02</td>
<td>1.08</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$\delta_1$</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>
The idea of Gelman and Rubin’s diagnostic is to compare the variances within chain and the variance between chains (see Chapter 3). The large deviation between these two variances indicates nonconvergence. From Table 5.1, the potential scale reduction factor (psrf) or $\hat{R}$ in the third column is the square root of the ratio of the estimated variance and the variance within the chain. If the $\hat{R}$ statistic is large, this means the variance between chains is substantially greater than the variances within the chain. Thus, longer chains are required to improve convergence to the target distribution. If the $\hat{R}$ statistic is close to 1, we can conclude that each of the four chains are all operating on the same distribution and they are likely to reach the target distribution. However, in practice $\hat{R}$ is acceptable if the value is less than roughly 1.1 or 1.2 (Gelman, 1996).

The Gelman and Rubin approach also produces an upper 97.5% confidence limit of $\hat{R}$ as in the fourth column of Table 5.1. At this point, $\hat{R}$ has an F-distribution and because we are only concerned with a large $\hat{R}$ value then only the upper 97.5% confidence limit is reported.

**Heidelberger and Welch Diagnostic**

The Heidelberger and Welch test (Heidelberger and Welch, 1983) is composed of two parts: a stationary test and a halfwidth test as shown in the third column and the forth column respectively in Table 5.2.
Table 5.2: Heidelberger and Welch Diagnostic, Model 1
Running the Heidelberger and Welch Diagnostic at $\epsilon = 0.1$ and $\alpha = 0.05$

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Stationary Test</th>
<th>Halfwidth Test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereotype Model (5 levels)</td>
<td>$a_2$</td>
<td>passed</td>
<td>passed</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>$a_3$</td>
<td>passed</td>
<td>passed</td>
<td>0.116</td>
</tr>
<tr>
<td>L = 5 levels</td>
<td>$a_4$</td>
<td>passed</td>
<td>passed</td>
<td>0.096</td>
</tr>
<tr>
<td>J = 32 questions</td>
<td>$a_5$</td>
<td>passed</td>
<td>passed</td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td>$b_1$</td>
<td>passed</td>
<td>passed</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>$b_2$</td>
<td>passed</td>
<td>passed</td>
<td>0.265</td>
</tr>
<tr>
<td></td>
<td>$b_3$</td>
<td>passed</td>
<td>passed</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td>$b_4$</td>
<td>passed</td>
<td>passed</td>
<td>0.171</td>
</tr>
<tr>
<td></td>
<td>$b_5$</td>
<td>passed</td>
<td>passed</td>
<td>0.214</td>
</tr>
<tr>
<td></td>
<td>$b_6$</td>
<td>passed</td>
<td>passed</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>$b_7$</td>
<td>passed</td>
<td>passed</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>$b_8$</td>
<td>passed</td>
<td>passed</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td>$b_9$</td>
<td>passed</td>
<td>passed</td>
<td>0.103</td>
</tr>
<tr>
<td></td>
<td>$b_{10}$</td>
<td>passed</td>
<td>passed</td>
<td>0.154</td>
</tr>
<tr>
<td></td>
<td>$b_{11}$</td>
<td>passed</td>
<td>passed</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>$b_{12}$</td>
<td>passed</td>
<td>passed</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td>$b_{13}$</td>
<td>passed</td>
<td>passed</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td>$b_{14}$</td>
<td>passed</td>
<td>passed</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>$b_{15}$</td>
<td>passed</td>
<td>passed</td>
<td>0.354</td>
</tr>
<tr>
<td></td>
<td>$b_{16}$</td>
<td>passed</td>
<td>passed</td>
<td>0.345</td>
</tr>
<tr>
<td></td>
<td>$b_{17}$</td>
<td>passed</td>
<td>passed</td>
<td>0.073</td>
</tr>
<tr>
<td></td>
<td>$b_{18}$</td>
<td>passed</td>
<td>passed</td>
<td>0.077</td>
</tr>
</tbody>
</table>

*continued on next page*
<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Stationary Test</th>
<th>Halfwidth Test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b_{19}$</td>
<td>passed</td>
<td>passed</td>
<td>0.156</td>
</tr>
<tr>
<td></td>
<td>$b_{20}$</td>
<td>passed</td>
<td>passed</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>$b_{21}$</td>
<td>passed</td>
<td>passed</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>$b_{22}$</td>
<td>passed</td>
<td>passed</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>$b_{23}$</td>
<td>passed</td>
<td>passed</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>$b_{24}$</td>
<td>passed</td>
<td>passed</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>$b_{25}$</td>
<td>passed</td>
<td>passed</td>
<td>0.175</td>
</tr>
<tr>
<td></td>
<td>$b_{26}$</td>
<td>passed</td>
<td>passed</td>
<td>0.226</td>
</tr>
<tr>
<td></td>
<td>$b_{27}$</td>
<td>passed</td>
<td>passed</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>$b_{28}$</td>
<td>passed</td>
<td>passed</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td>$b_{29}$</td>
<td>passed</td>
<td>passed</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>$b_{30}$</td>
<td>passed</td>
<td>passed</td>
<td>0.186</td>
</tr>
<tr>
<td></td>
<td>$b_{31}$</td>
<td>passed</td>
<td>passed</td>
<td>0.198</td>
</tr>
<tr>
<td></td>
<td>$b_{32}$</td>
<td>passed</td>
<td>passed</td>
<td>0.123</td>
</tr>
<tr>
<td>Stereotype Model</td>
<td>$\phi_2$</td>
<td>passed</td>
<td>passed</td>
<td>0.215</td>
</tr>
<tr>
<td>(3 levels)</td>
<td>$\phi_3$</td>
<td>passed</td>
<td>passed</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>$\phi_4$</td>
<td>passed</td>
<td>passed</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>$a_{2}$</td>
<td>passed</td>
<td>passed</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>$a_{3}$</td>
<td>passed</td>
<td>passed</td>
<td>0.175</td>
</tr>
<tr>
<td>$L = 3$ levels</td>
<td>$b_{1}$</td>
<td>passed</td>
<td>passed</td>
<td>0.368</td>
</tr>
<tr>
<td>$J = 6$ questions</td>
<td>$b_{2}$</td>
<td>passed</td>
<td>passed</td>
<td>0.417</td>
</tr>
<tr>
<td></td>
<td>$b_{3}$</td>
<td>passed</td>
<td>passed</td>
<td>0.297</td>
</tr>
<tr>
<td></td>
<td>$b_{4}$</td>
<td>passed</td>
<td>passed</td>
<td>0.175</td>
</tr>
<tr>
<td></td>
<td>$b_{5}$</td>
<td>passed</td>
<td>passed</td>
<td>0.259</td>
</tr>
<tr>
<td></td>
<td>$b_{6}$</td>
<td>passed</td>
<td>passed</td>
<td>0.435</td>
</tr>
</tbody>
</table>

*continued on next page*
### 5.2. DATA ANALYSIS: BAYESIAN JOINT MODEL

#### Table (5.2) continued from previous page

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Stationary Test</th>
<th>Halfwidth Test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>φ_2</td>
<td>passed</td>
<td>passed</td>
<td>0.075</td>
<td></td>
</tr>
<tr>
<td>Binary Model</td>
<td>α</td>
<td>passed</td>
<td>passed</td>
<td>0.874</td>
</tr>
<tr>
<td>Q = 2 questions</td>
<td>b_1</td>
<td>passed</td>
<td>passed</td>
<td>0.680</td>
</tr>
<tr>
<td></td>
<td>b_2</td>
<td>passed</td>
<td>passed</td>
<td>0.256</td>
</tr>
<tr>
<td>Regression Model</td>
<td>ρ</td>
<td>passed</td>
<td>passed</td>
<td>0.757</td>
</tr>
<tr>
<td></td>
<td>β_0(t_0)</td>
<td>passed</td>
<td>passed</td>
<td>0.665</td>
</tr>
<tr>
<td></td>
<td>β_0(t_1)</td>
<td>passed</td>
<td>passed</td>
<td>0.203</td>
</tr>
<tr>
<td></td>
<td>β_0(t_2)</td>
<td>passed</td>
<td>passed</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>β_0(t_3)</td>
<td>passed</td>
<td>passed</td>
<td>0.409</td>
</tr>
<tr>
<td></td>
<td>β_0(t_4)</td>
<td>passed</td>
<td>passed</td>
<td>0.393</td>
</tr>
<tr>
<td></td>
<td>β_0(t_5)</td>
<td>passed</td>
<td>passed</td>
<td>0.312</td>
</tr>
<tr>
<td></td>
<td>β_0(t_6)</td>
<td>passed</td>
<td>passed</td>
<td>0.225</td>
</tr>
<tr>
<td></td>
<td>β_1(t_0)</td>
<td>passed</td>
<td>passed</td>
<td>0.454</td>
</tr>
<tr>
<td></td>
<td>β_1(t_1)</td>
<td>passed</td>
<td>passed</td>
<td>0.420</td>
</tr>
<tr>
<td></td>
<td>β_1(t_2)</td>
<td>passed</td>
<td>passed</td>
<td>0.463</td>
</tr>
<tr>
<td></td>
<td>β_1(t_3)</td>
<td>passed</td>
<td>passed</td>
<td>0.936</td>
</tr>
<tr>
<td></td>
<td>β_1(t_4)</td>
<td>passed</td>
<td>passed</td>
<td>0.592</td>
</tr>
<tr>
<td></td>
<td>β_1(t_5)</td>
<td>passed</td>
<td>passed</td>
<td>0.261</td>
</tr>
<tr>
<td></td>
<td>β_1(t_6)</td>
<td>passed</td>
<td>passed</td>
<td>0.235</td>
</tr>
<tr>
<td>Cox’s Model</td>
<td>δ_0</td>
<td>passed</td>
<td>passed</td>
<td>0.576</td>
</tr>
<tr>
<td></td>
<td>δ_1</td>
<td>passed</td>
<td>passed</td>
<td>0.306</td>
</tr>
</tbody>
</table>

The null hypothesis for the stationary test is that the chain is currently in a stationary distribution. The halfwidth test is performed when some proportion of the data are found to be consistent with stationarity in order to estimate the mean values. The test statistic is the Cramer-von-Mises as
mentioned in Chapter 3. We first assess the whole chain which is 100,000 iterations at $\varepsilon = 0.1$ (an accuracy) and $\alpha = 0.05$ (an alpha level of the test). Then we evaluate the first 10% of the chain iterations and calculate the test statistic to see whether to accept or reject the null hypothesis. If the null hypothesis is rejected, the first 10% of the chain iterations are discarded and the test is run again. We repeat the process until the null hypothesis is accepted with the remaining 60,000 iterations. It can be seen from the $p$-value in Table 5.2 that all of the chain values are greater than the alpha level of the test ($\alpha = 0.05$). This means the chain passes the stationary test and subsequently the chain that is not discarded from the first part is taken to test the halfwidth. To perform the halfwidth test, the `coda` package calculates half the width of the $(1-\alpha) \%$ credible interval around the mean. If the ratio of the halfwidth and the mean is lower than a rough estimate of the variance of the sample mean or some accuracy ($\varepsilon = 0.1$), then the chain passes the test. From Table 5.2, the chain is considered to be converged since the halfwidth is less than the rough estimate of the variance.

After we discard a burn-in from the chains, the next step is to use the remaining iterations to calculate the model checking criteria. In this thesis we use the Bayesian information criterion (BIC) and the Deviance Information Criterion (DIC).
5.2. DATA ANALYSIS: BAYESIAN JOINT MODEL

5.2.2 Bayesian Model Selection

The BIC and DIC are based on the posterior distribution of the log likelihood or the deviance. The BIC also depends on the sample size as well as the number of parameters. For the calculation procedures, we apply a multiple imputation approach proposed by Rubin (1987, 1996) to the criterion calculation since missing data have been involved in our joint model. Multiple imputation is done by repeating the imputation process several times in order to capture some of the uncertainty associated with the estimation of missing data. Missing values are generated from an appropriate imputation model (the posterior distribution of missing data); typically between 3 and 10 sets are sufficient (Rubin, 1987). The posterior estimates are obtained by averaging the results over the imputed samples.

Since our joint model contains many parameters and some of them are the latent variables, the DIC is used to overcome the problem of needing to identify the number of parameters in the model which is required for calculating the BIC. For both methods, smaller values represent a better overall fit.
Table 5.3: **Model Comparison, BIC and DIC**

<table>
<thead>
<tr>
<th>Model</th>
<th>BIC</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>1605155</td>
<td>1602788</td>
</tr>
<tr>
<td>Model 2</td>
<td>1653515</td>
<td>1603696</td>
</tr>
</tbody>
</table>

In Table 5.3 we report BIC together with DIC. Although we used different initial values in several runs, the BIC and DIC estimates obtained never varied by more than 0.5. The comparison of the BIC and DIC between Model 1 and Model 2 suggests that Model 1 is the best fit to the data since it has the lowest BIC and DIC values. Model 1 consists of the item response part, the regression part and the survival part with two covariates: quality of life and treatment. The next section shows the posterior mean and the 95% credible interval for each parameter of Model 1.

### 5.2.3 Posterior Summary

For Model 1, we ran the chain for a total of 100,000 iterations with a burn-in period of 40,000 iterations. We used the last 60,000 iterations for parameter estimation. After the burn-in iterations were discarded, the posterior results are summarized as follows:
Table 5.4: The Posterior Summary, Model 1

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Mean</th>
<th>Standard Error</th>
<th>95% Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereotype Model</td>
<td>$a_2$</td>
<td>0.948</td>
<td>0.036</td>
<td>(0.873,1.012)</td>
</tr>
<tr>
<td>(5 levels)</td>
<td>$a_3$</td>
<td>2.855</td>
<td>0.031</td>
<td>(2.799,2.921)</td>
</tr>
<tr>
<td>$L = 5$ levels</td>
<td>$a_4$</td>
<td>3.441</td>
<td>0.036</td>
<td>(3.361,3.520)</td>
</tr>
<tr>
<td>$J = 32$ questions</td>
<td>$a_5$</td>
<td>3.656</td>
<td>0.038</td>
<td>(3.573,3.732)</td>
</tr>
<tr>
<td></td>
<td>$b_1$</td>
<td>3.241</td>
<td>0.112</td>
<td>(3.021,3.459)</td>
</tr>
<tr>
<td></td>
<td>$b_2$</td>
<td>2.287</td>
<td>0.093</td>
<td>(2.104,2.471)</td>
</tr>
<tr>
<td></td>
<td>$b_3$</td>
<td>0.914</td>
<td>0.081</td>
<td>(0.753,1.071)</td>
</tr>
<tr>
<td></td>
<td>$b_4$</td>
<td>0.926</td>
<td>0.080</td>
<td>(0.765,1.085)</td>
</tr>
<tr>
<td></td>
<td>$b_5$</td>
<td>2.292</td>
<td>0.095</td>
<td>(2.109,2.482)</td>
</tr>
<tr>
<td></td>
<td>$b_6$</td>
<td>3.999</td>
<td>0.132</td>
<td>(3.746,4.262)</td>
</tr>
<tr>
<td></td>
<td>$b_7$</td>
<td>2.603</td>
<td>0.102</td>
<td>(2.401,2.813)</td>
</tr>
<tr>
<td></td>
<td>$b_8$</td>
<td>4.355</td>
<td>0.138</td>
<td>(4.092,4.632)</td>
</tr>
<tr>
<td></td>
<td>$b_9$</td>
<td>1.372</td>
<td>0.085</td>
<td>(1.027,1.541)</td>
</tr>
<tr>
<td></td>
<td>$b_{10}$</td>
<td>5.028</td>
<td>0.155</td>
<td>(4.725,5.336)</td>
</tr>
<tr>
<td></td>
<td>$b_{11}$</td>
<td>1.945</td>
<td>0.088</td>
<td>(1.778,2.123)</td>
</tr>
<tr>
<td></td>
<td>$b_{12}$</td>
<td>2.664</td>
<td>0.103</td>
<td>(2.464,2.866)</td>
</tr>
<tr>
<td></td>
<td>$b_{13}$</td>
<td>3.859</td>
<td>0.127</td>
<td>(3.617,4.123)</td>
</tr>
<tr>
<td></td>
<td>$b_{14}$</td>
<td>3.305</td>
<td>0.117</td>
<td>(3.075,3.534)</td>
</tr>
<tr>
<td></td>
<td>$b_{15}$</td>
<td>2.842</td>
<td>0.107</td>
<td>(2.637,3.500)</td>
</tr>
<tr>
<td></td>
<td>$b_{16}$</td>
<td>2.428</td>
<td>0.100</td>
<td>(2.232,2.627)</td>
</tr>
<tr>
<td></td>
<td>$b_{17}$</td>
<td>3.234</td>
<td>0.115</td>
<td>(3.012,3.458)</td>
</tr>
<tr>
<td></td>
<td>$b_{18}$</td>
<td>2.437</td>
<td>0.099</td>
<td>(2.242,2.636)</td>
</tr>
</tbody>
</table>

continued on next page
### Table (5.4) continued from previous page

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Mean</th>
<th>Standard Error</th>
<th>95% Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b_{19}$</td>
<td>2.294</td>
<td>0.095</td>
<td>(2.110, 2.480)</td>
</tr>
<tr>
<td></td>
<td>$b_{20}$</td>
<td>2.033</td>
<td>0.092</td>
<td>(1.860, 2.211)</td>
</tr>
<tr>
<td></td>
<td>$b_{21}$</td>
<td>2.422</td>
<td>0.098</td>
<td>(2.232, 2.614)</td>
</tr>
<tr>
<td></td>
<td>$b_{22}$</td>
<td>2.038</td>
<td>0.095</td>
<td>(1.853, 2.229)</td>
</tr>
<tr>
<td></td>
<td>$b_{23}$</td>
<td>2.392</td>
<td>0.982</td>
<td>(2.204, 2.587)</td>
</tr>
<tr>
<td></td>
<td>$b_{24}$</td>
<td>2.848</td>
<td>0.106</td>
<td>(2.645, 3.058)</td>
</tr>
<tr>
<td></td>
<td>$b_{25}$</td>
<td>2.197</td>
<td>0.098</td>
<td>(2.008, 2.390)</td>
</tr>
<tr>
<td></td>
<td>$b_{26}$</td>
<td>1.425</td>
<td>0.082</td>
<td>(1.265, 1.591)</td>
</tr>
<tr>
<td></td>
<td>$b_{27}$</td>
<td>1.091</td>
<td>0.092</td>
<td>(1.727, 2.088)</td>
</tr>
<tr>
<td></td>
<td>$b_{28}$</td>
<td>1.020</td>
<td>0.082</td>
<td>(0.857, 1.183)</td>
</tr>
<tr>
<td></td>
<td>$b_{29}$</td>
<td>0.395</td>
<td>0.083</td>
<td>(0.231, 0.551)</td>
</tr>
<tr>
<td></td>
<td>$b_{30}$</td>
<td>0.018</td>
<td>0.017</td>
<td>(0.004, 0.065)</td>
</tr>
<tr>
<td></td>
<td>$b_{31}$</td>
<td>0.967</td>
<td>0.084</td>
<td>(0.806, 1.134)</td>
</tr>
<tr>
<td></td>
<td>$b_{32}$</td>
<td>2.117</td>
<td>0.095</td>
<td>(1.936, 2.304)</td>
</tr>
<tr>
<td></td>
<td>$\phi_2$</td>
<td>0.024</td>
<td>0.008</td>
<td>(0.007, 0.039)</td>
</tr>
<tr>
<td></td>
<td>$\phi_3$</td>
<td>0.351</td>
<td>0.006</td>
<td>(0.338, 0.364)</td>
</tr>
<tr>
<td></td>
<td>$\phi_4$</td>
<td>0.707</td>
<td>0.007</td>
<td>(0.692, 0.720)</td>
</tr>
<tr>
<td>Stereotype Model</td>
<td>$a_2$</td>
<td>1.846</td>
<td>0.097</td>
<td>(1.668, 2.032)</td>
</tr>
<tr>
<td></td>
<td>$a_3$</td>
<td>3.642</td>
<td>0.096</td>
<td>(3.461, 3.829)</td>
</tr>
<tr>
<td></td>
<td>$b_1$</td>
<td>1.996</td>
<td>0.106</td>
<td>(1.787, 2.206)</td>
</tr>
<tr>
<td></td>
<td>$b_2$</td>
<td>1.374</td>
<td>0.087</td>
<td>(1.206, 1.552)</td>
</tr>
<tr>
<td></td>
<td>$b_3$</td>
<td>1.560</td>
<td>0.095</td>
<td>(1.378, 1.748)</td>
</tr>
<tr>
<td></td>
<td>$b_4$</td>
<td>1.324</td>
<td>0.086</td>
<td>(1.159, 1.495)</td>
</tr>
<tr>
<td></td>
<td>$b_5$</td>
<td>1.075</td>
<td>0.078</td>
<td>(0.929, 1.236)</td>
</tr>
<tr>
<td></td>
<td>$b_6$</td>
<td>0.643</td>
<td>0.077</td>
<td>(0.492, 0.792)</td>
</tr>
</tbody>
</table>

*continued on next page*
Table 5.4 shows the estimate of all parameters in the Bayesian joint model and their standard errors with a 95% credible interval except the latent variables $\theta$. Beginning with the item response models, two stereotype models are included in the Bayesian joint model since the questions in
the questionnaire contain a different number of ordinal categories. One has five ordinal categories \((L = 5)\) and the other has three ordinal categories \((L = 3)\). These two stereotype models were fitted simultaneously. The score parameters \(\phi_\ell\) are monotonically increasing in both models as expected as follows: \(\phi^{L=5} = \{0, 0.024, 0.351, 0.707, 1\}\) and \(\phi^{L=3} = \{0, 0.196, 1\}\). We can see how close adjacent response categories \(\ell\) and \(\ell'\) are, based on how close \(\phi_\ell\) and \(\phi_{\ell'}\) are. Also, the distinguishability and order of outcome categories can be addressed through these constraints. Relative to the ordering for the former stereotype model, the estimated \(\phi_2\) is not far from \(\phi_1\). We can group the score parameters into four values, \(\phi_1 = \phi_2, \phi_3, \phi_4\) and \(\phi_5\) for distinguishability of categories. These correspond to four groups of distinguishable response categories in MOS-HIV questionnaire (poor and fair, good, very good, excellent). Likewise, the response categories for the latter stereotype model are distinguishable from one another relative to their standard errors.

The item slope coefficients \(b\) indicate the influence of the quality of life scores \(\theta\) on how patients select the response category. All of the coefficients \(b\) are constrained to be positive which means patients who have greater quality of life scores would have a high probability of responding to a high category. From the MOS-HIV questionnaire we ordered the response categories from low to high (poor to excellent); then patients in a high category should have a better quality of life compared to the patients in a low category. This suggests the higher the value of \(b_j \theta_i(t_m)\), the
greater probability of choosing the excellent categories. Furthermore, the stereotype model allows each individual to have a different probability of responding depending on the strength of the question. For example, for a fixed value of $\theta_i(t_m)$, say $\theta_i(t_m) = \theta$ and $\theta > 0$, the odds of responding category $\ell$ instead of 1 for question 10 against question 30 equals

$$\frac{P(Y = \ell | \theta, b_{10})}{P(Y = 1 | \theta, b_{30})} / \frac{P(Y = 1 | \theta, b_{10})}{P(Y = 1 | \theta, b_{30})} = \exp \left( \phi_\ell (b_{10} - b_{30}) \theta \right)$$ (5.1)

The estimate of the odds ratio (5.1) equals $\exp \left( \phi_\ell (5.028 - 0.018) \theta \right)$. This means the probability of responding category $\ell$ in question 10 is higher than question 30. Likewise, the odds of responding $Y = 1$ (patients are able to do certain kinds or amounts of work by themselves without health problems) for question 1 against question 2 of the binary model equals

$$\frac{P(Y = 1 | \theta, b_1)}{P(Y = 0 | \theta, b_2)} / \frac{P(Y = 0 | \theta, b_1)}{P(Y = 0 | \theta, b_2)} = \exp \left( (b_1 - b_2) \theta \right)$$ (5.2)

The estimate of the odds ratio (5.2) equals $\exp \left( (1.039 - 0.931) \theta \right)$. This means the probability of responding $Y = 1$ in question 1 is higher than question 2.

The quality of life scores $\theta$ and the treatment effect are the covariates of the Cox proportional hazards model. The estimates of $\delta_0$ and $\delta_1$ shown in Table 5.4 represent the regression coefficients of the quality of life and the treatment respectively. Since $\delta_0$ is a negative value, this implies that a higher quality of life score or a better quality of life is associated with
a reduced hazard for having CDC-B. In a similar manner, a negative regression coefficient $\delta_1$ means that the hazard of progression to CDC-B of patients on a continuous treatment is lower than patients on CD4-guided treatment.

Figure 5.3: Mean quality of life estimates

Figure 5.3 illustrates the estimated mean quality of life using the form of $\mu(\theta(t_m)) = \beta_0(t_m) + \beta_1(t_m)X$ in the regression model. The mean quality of life here is a function of time since the data were collected every
24 weeks. The regression coefficients $\{\beta_i(t_m), m = 0 \ldots, 6\}$ describe how treatment affects the quality of life. In this figure we see that the mean quality of life of the patients on a continuous treatment (dashed line) is higher than the mean quality of life of the patients on CD4-guided treatment (solid line). This shows a connection to the results we obtained from the survival model and the stereotype model. However, the mean quality of life of both treatment methods are slightly similar at the first week visit. After that they all markedly decrease within 24 weeks and then the mean quality of life improves with continuous treatment but not with CD4-guided treatment.

The Bayesian joint model approach generated the latent variables for all individuals ($n = 354$) with every follow-up time point ($m = 7$). In total, there are $n \times m = 354 \times 7$ elements in $\theta$. The posterior quality of life scores are randomly selected from some patients (a patient number 28 and a patient number 39) to present here as shown in Figure 5.4. Patient number 28 (dashed line) is on continuous treatment and the other patient is on CD4-guided treatment. The quality of life scores increase and decrease during the follow-up period. This implies that not only the treatment but also other aspects of functioning have an impact on the quality of life. The study about HAART (antiretroviral drugs) and its effect on quality of life in the literature has been controversial. Some studies found overall quality of life improved with the use of HAART (Nieuwkerk et al., 2002; Carrieri et al., 2003) but Gill et al. (2002) among others, reported that physical qual-
Figure 5.4: Posterior quality of life scores

Quality of life decreased under HAART.
5.3 Data Analysis: Semiparametric Joint Model

Approach

To illustrate the semiparametric joint model, we also used data from the Staccato study for our analysis and we used 32 questions from the MOS-HIV questionnaire where each question has five response categories. Here, the baseline hazard function is calculated whereas it is not included in the Bayesian joint model. The latent variable is treated as a discrete variable while it is assumed to be a continuous variable in the Bayesian joint model. The semiparametric joint model consists of the stereotype model and the Cox proportional hazards model. The two models are linked through a discrete latent variable that is assumed to underlie the quality of life and the hazard. In this section, although our work has focused on the joint model in another aspect, we also consider the model comparison.

5.3.1 Model Comparison

We constructed various semiparametric joint models that have the same structure of the stereotype model but each model has different covariates included for the survival part of the model such as treatment, sex and age. The number of discrete latent variables or the number of groups which are accommodated in the stereotype model and the Cox proportional hazards model are also different across the models. The profile Akaike information criterion (pAIC) is used for model selection since we have nuisance
parameters in the models. The summary of the profile Akaike information criterion is shown in Table 5.5 below. The number of groups for the discrete latent variables in the model is represented by $N_{gr}$ and the number of parameters is $N_{par}$.

Table 5.5: The summary of profile Akaike information criterion (pAIC)

<table>
<thead>
<tr>
<th>Covariates in the survival part</th>
<th>$N_{gr}$</th>
<th>$N_{par}$</th>
<th>pAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) QOL and treatment</td>
<td>2</td>
<td>38</td>
<td>126764.2</td>
</tr>
<tr>
<td>(b) QOL, treatment and initial CD4 count</td>
<td>2</td>
<td>39</td>
<td>119005.51</td>
</tr>
<tr>
<td>(c) QOL and treatment</td>
<td>3</td>
<td>39</td>
<td>116466.3</td>
</tr>
<tr>
<td>(d) QOL, treatment and initial CD4 count</td>
<td>3</td>
<td>40</td>
<td>116468.3</td>
</tr>
<tr>
<td>(e) QOL, treatment, initial CD4 count and age</td>
<td>3</td>
<td>41</td>
<td>116594.8</td>
</tr>
<tr>
<td>(f) QOL, treatment, initial CD4 count, age and sex</td>
<td>3</td>
<td>42</td>
<td>116611.0</td>
</tr>
<tr>
<td>(g) QOL, treatment, initial CD4 count, age and weight</td>
<td>3</td>
<td>42</td>
<td>116841.1</td>
</tr>
<tr>
<td>(h) QOL, treatment, age and sex</td>
<td>3</td>
<td>41</td>
<td>116480.8</td>
</tr>
<tr>
<td>(i) QOL, treatment, initial CD4 count and sex</td>
<td>3</td>
<td>41</td>
<td>116483.8</td>
</tr>
<tr>
<td>(j) QOL, treatment and age</td>
<td>3</td>
<td>40</td>
<td>116465.3</td>
</tr>
<tr>
<td>(k) QOL, treatment and sex</td>
<td>3</td>
<td>40</td>
<td>116464.0</td>
</tr>
<tr>
<td>(l) QOL and treatment</td>
<td>4</td>
<td>40</td>
<td>115603.4</td>
</tr>
<tr>
<td>(m) QOL, treatment and CD4 count</td>
<td>4</td>
<td>41</td>
<td>115605.4</td>
</tr>
<tr>
<td>(n) QOL and treatment</td>
<td>5</td>
<td>41</td>
<td>115205.2</td>
</tr>
</tbody>
</table>

*continued on next page*
Table 5.5 presents the profile Akaike information criterion (pAIC) for each model and the model with the smaller pAIC is preferred. It can be seen from Table 5.5 that the range of pAIC among the models is very small. However, the pAIC values suggest that Model (n) is the most suitable model to this data set. Model (n) includes effects of the treatment, and quality of life as defined by the latent discrete variable with five groups in the survival part. In the following the estimated parameters and the standard errors of selected model are given in Table 5.6.

### 5.3.2 Estimated parameters

**Table 5.6: EM estimate from the selected model**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimated value</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_2$</td>
<td>0.845</td>
<td>0.027</td>
</tr>
<tr>
<td>$a_3$</td>
<td>0.052</td>
<td>0.045</td>
</tr>
<tr>
<td>$a_4$</td>
<td>-1.080</td>
<td>0.065</td>
</tr>
<tr>
<td>$a_5$</td>
<td>-4.174</td>
<td>0.115</td>
</tr>
<tr>
<td>$b_2$</td>
<td>2.606</td>
<td>0.133</td>
</tr>
</tbody>
</table>

continued on next page
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimated value</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>b₃</td>
<td>4.759</td>
<td>0.139</td>
</tr>
<tr>
<td>b₄</td>
<td>5.653</td>
<td>0.144</td>
</tr>
<tr>
<td>b₅</td>
<td>3.152</td>
<td>0.135</td>
</tr>
<tr>
<td>b₆</td>
<td>0.186</td>
<td>0.137</td>
</tr>
<tr>
<td>b₇</td>
<td>2.951</td>
<td>0.135</td>
</tr>
<tr>
<td>b₈</td>
<td>0.648</td>
<td>0.136</td>
</tr>
<tr>
<td>b₉</td>
<td>5.085</td>
<td>0.142</td>
</tr>
<tr>
<td>b₁₀</td>
<td>0.456</td>
<td>0.136</td>
</tr>
<tr>
<td>b₁₁</td>
<td>3.934</td>
<td>0.137</td>
</tr>
<tr>
<td>b₁₂</td>
<td>1.786</td>
<td>0.134</td>
</tr>
<tr>
<td>b₁₃</td>
<td>1.473</td>
<td>0.134</td>
</tr>
<tr>
<td>b₁₄</td>
<td>1.793</td>
<td>0.134</td>
</tr>
<tr>
<td>b₁₅</td>
<td>2.997</td>
<td>0.135</td>
</tr>
<tr>
<td>b₁₆</td>
<td>3.861</td>
<td>0.137</td>
</tr>
<tr>
<td>b₁₇</td>
<td>2.217</td>
<td>0.134</td>
</tr>
<tr>
<td>b₁₈</td>
<td>2.416</td>
<td>0.134</td>
</tr>
<tr>
<td>b₁₉</td>
<td>1.938</td>
<td>0.134</td>
</tr>
<tr>
<td>b₂₀</td>
<td>2.727</td>
<td>0.134</td>
</tr>
<tr>
<td>b₂₁</td>
<td>2.570</td>
<td>0.134</td>
</tr>
<tr>
<td>b₂₂</td>
<td>3.372</td>
<td>0.135</td>
</tr>
</tbody>
</table>

continued on next page
5.3. DATA ANALYSIS: SEMIPARAMETRIC JOINT MODEL

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimated value</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_{23}$</td>
<td>2.329</td>
<td>0.134</td>
</tr>
<tr>
<td>$b_{24}$</td>
<td>1.496</td>
<td>0.134</td>
</tr>
<tr>
<td>$b_{25}$</td>
<td>3.379</td>
<td>0.135</td>
</tr>
<tr>
<td>$b_{26}$</td>
<td>2.650</td>
<td>0.133</td>
</tr>
<tr>
<td>$b_{27}$</td>
<td>2.536</td>
<td>0.133</td>
</tr>
<tr>
<td>$b_{28}$</td>
<td>4.880</td>
<td>0.141</td>
</tr>
<tr>
<td>$b_{29}$</td>
<td>6.289</td>
<td>0.151</td>
</tr>
<tr>
<td>$b_{30}$</td>
<td>8.496</td>
<td>0.182</td>
</tr>
<tr>
<td>$b_{31}$</td>
<td>4.696</td>
<td>0.141</td>
</tr>
<tr>
<td>$b_{32}$</td>
<td>3.006</td>
<td>0.136</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.364</td>
<td>-</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>0.569</td>
<td>-</td>
</tr>
<tr>
<td>$\delta_0$</td>
<td>-0.172</td>
<td>0.064</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>-0.943</td>
<td>0.088</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>3.790</td>
<td>0.073</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>7.251</td>
<td>0.062</td>
</tr>
<tr>
<td>$\theta_4$</td>
<td>5.427</td>
<td>0.050</td>
</tr>
<tr>
<td>$\theta_5$</td>
<td>2.233</td>
<td>0.436</td>
</tr>
</tbody>
</table>
Table 5.6 shows the estimated parameters and their standard errors from the EM algorithm approach. The results suggest that for this data set patients belong to five different groups classified by their quality of life score \( \theta_r \). Treatment is included as a predictor together with the quality of life in the survival part. All of the coefficients \( b \) are positive values which mean positive effects. Since \( b_j \) expresses the strength of the \( j^{th} \) question, the higher value of \( b \), the higher the probability that patients respond to a high category in such questions. For example, if \( b_1 > b_2 \), the probability that patients will respond in a high category for question one is higher than question two. Similarly, \( \theta_r \) serves as a quality of life score of patients in group \( r \); the greater the value of \( \theta \) the higher the probability of the patient responding to a high category. This also indicates that the patients who are in a group with the high value of \( \theta \), have a better quality of life.

As mentioned above, there are two predictors in the survival part. The regression coefficient \( \delta_0 \) is the coefficient of quality of life and it has a negative value. This result suggests that better quality of life is associated with a reduced hazard of having CDC-B. Likewise, \( \delta_1 \) corresponds to the effect of treatment and it shows a negative value. This means the patients on continuous treatment have a lower hazard of CDC-B progression than the patients on CD4-guided treatment.
The cumulative hazard function as shown in Figure 5.5 is obtained from the calculation of (4.34) with the EM algorithm in Chapter 4. The graph shows the increasing step function or monotonic increasing in time by definition. However, in most applications, we typically describe how long the study subjects live rather than how quickly they die. Thus, the survival function has received more attention than the baseline cumulative hazard function in terms of interpretation. The survival function, $S(t)$, can be derived from the cumulative hazard function $H(t)$ by $S(t) = \exp(-H(t))$. 
The estimated survival function is illustrated in Figure 5.6. It decreases sharply for approximately the first 24 weeks. The graph significantly declines again from week 40 to week 72. There are larger drops in the steps in the graph from week 72 to week 144 due to the small number of patients still at risk but many of these are censored during the follow-up time. The initial steep decrease is because there is a high rate of progression to CDC-B in the first 24 weeks. The progression rate then decreases and remains at about the same level for the remainder of the follow-up period.
Chapter 6

The Extension of the Joint Model

6.1 Introduction

In Chapter 3 we consider a single outcome CDC-B as the survival endpoint. In this chapter we focus on the multiple survival endpoints since sometimes patients may fail from different causes or patients may face different risks. If this is the case, it may not be appropriate to use the classical survival techniques. Thus, in this chapter we extend the Bayesian joint model from Chapter 3 by presenting the specific methods to deal with the case of multiple failure types. The situation where several causes of failure are possible and all of them are of equal interest is known as the competing risks (Putter et al., 2007). If failures are different causes of death, only the occurrence of the first of them can be observed. The term competing risks also refers to data when subjects under study are at risk of more than one mutually exclusive event and the interest is in the first failure; the remain-
CHAPTER 6. THE EXTENSION OF THE JOINT MODEL

ing failures are competing events. Gooley et al. (1999) also defined competing risks as events whose occurrence either precludes the occurrence of another event under investigation or fundamentally alters the probability of occurrence of this other event. Examples of competing risks are often found in medical studies such as cancer and heart disease where the competing risks could be competing causes of death. In clinical cancer research, individuals undergoing treatment for surgical resectable disease may experience recurrence near the removed tumor, metastatic recurrence at other sites, occurrence of a second cancer, or death resulting from non-cancerous causes before any of these events (Dignam and Kocherginsky, 2008). In another example of competing risks, the primary outcome was the onset of dementia and other types of events were stroke and death in a study on a community-based cohort of patients diagnosed with atrial fibrillation between 1986 to 2000 in Olmsted Country, Minnesota United States (Miyasaka et al., 2007).

The competing risks model as a survival component in the joint model has been increasingly considered in medical studies. Elashoff et al. (2007) considered a more general joint model, which incorporates the competing risks model for the survival endpoint, using a likelihood approach. However, their EM algorithm involved intractable computation for high dimension integrals. Furthermore, Williamson et al. (2008) demonstrated a cause-specific hazards sub-model to allow for competing risks, with a separate latent association between longitudinal measurements and each
cause of failure. They also used the EM algorithm to estimate the parameters of interest where the random effects were treated as missing data. Hu et al. (2009) developed a Bayesian method for joint modeling of a linear mixed effects model and a competing risks model. They indicated that the Bayesian approach can avoid the difficulty of implementation for high-dimensional random effects, unlike the likelihood approach. Deslandes and Chevret (2010) also proposed a joint model of competing risks with longitudinal responses using the Bayesian method. Although various studies on the joint model have been proposed, no one has applied the stereotype model to longitudinal ordinal data. Recently, Li et al. (2010) presented a joint model for longitudinal ordinal measurements and competing risks failure time data, but they used a partial proportional odds model to model the longitudinal ordinal data. Therefore, this thesis is the first to apply the stereotype model to the joint model of competing risks and longitudinal ordinal data.

6.2 Model Structure

In this chapter we present the joint model which is composed of an item response model to model quality of life data and the competing risks model is used to model survival times with multiple endpoints. These two models are linked by a latent variable process. Similarly, we consider the Bayesian method to estimate the parameters as we did in Chapter 3 since the model is very complicated with a high-dimensional data set.
The stereotype model is used to create the item response term and for demonstration we consider only the questions that have five response categories from the measure of quality of life MOS-HIV questionnaire. The fitting procedure is the same as in Chapter 3, except the survival process part; we illustrate with two event types in the competing risks model. In general, one may extend to more event types. Next, we discuss the competing risks model.

6.2.1 The Competing Risks Model

In a simple survival model, the hazard function for the failure time is given by

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

where \( t \) is the time to failure and \( X \) is a covariate vector.

In a competing risks model, the observed data are represented by the time of failure \( t \), covariate \( X \) and the cause of failure \( k \). Suppose there is a risk of experiencing \( K(k = 1, \ldots, K) \) event types. The contribution of each individual to the likelihood is specified through a joint distribution of \((t, k)\). For cause \( k \), we can write the cause-specific hazard as

$$h_k(t|X) = \lim_{\Delta t \to 0} \left[ \frac{P(t \leq T_k < (t + \Delta t), \text{event} = k|T_k \geq t, X)}{\Delta t} \right]$$

The numerator is defined as the conditional probability that an individual
with covariate $X$ fails in an interval $t + \Delta t$ and the cause of failure is the $k^{th}$ cause, given that the individual has survived to time $t$ or no failure of any kind has occurred thus far. We turn the probability into a rate by dividing by $\Delta t$ and then taking the limit as $\Delta t \to 0$. This further represents the rate of occurrence of the $k^{th}$ failure. The overall hazard $h(t|X)$ for any type of failure at time $t$ is the sum of the cause-specific hazards $h_k(t|X)$. Thus, the overall hazard can be expressed as

$$h(t|X) = \sum_{k=1}^{K} h_k(t|X)$$

This shows each individual must fail from only one cause out of $K$ causes. The overall survival function can be defined as

$$S(t|X) = e^{-H(t|X)}$$

where $H(t|X)$ is the cumulative risk and can be factorized into the $K$ functions as follows

$$S(t|X) = e^{-\sum_{k=1}^{K} H_k(t|X)}$$

$$= \prod_{k=1}^{K} e^{-H_k(t|X)}$$

$$= \prod_{k=1}^{K} S_k(t|X)$$

We can also define a cause-specific density function of failures at time
CHAPTER 6. THE EXTENSION OF THE JOINT MODEL

$t$ as follows

\[ f_k(t|X) = h_k(t|X)S(t|X) \]

By the law of the total probability, the overall density of failures at time $t$ is

\[ f(t|X) = \sum_{k=1}^{K} f_k(t|X) \]

Suppose we have $n$ individuals ($i = 1, 2, \ldots, n$) and $K$ event types ($k = 1, 2, \ldots, K$). Under the assumption of independent observations, the contribution of a given observation with failure due to risk $k$ to the likelihood function can be written as

\[ L(\cdot) = \prod_{i=1}^{n} \left\{ \left( h_{k_i}(t_i|X_i) \right)^{d_i} \times S(t_i|X_i) \right\} \quad (6.1) \]

where $d_i = 1$ if we observe the failure time and it is 0 if we observe a censoring. The cause of failure index $k$ takes a value from 1 to $K$ and is undefined for censored cases. The indicator $d_i$ allows us to write the two types of terms as follows:

For the censored cases, an individual censored at time $t_i$ contributes to the probability of being alive at that time presented by $S(t_i|X_i)$. On the other hand, if the individual fails at time $t$ from cause $k$, we can write the density in terms of the hazard and survival functions as $h_{k_i}(t_i|X_i)S(t_i|X_i)$. 
6.2. MODEL STRUCTURE

Since \( S(t|X) = \prod_{k=1}^{K} S_k(t|X) \), we can write the likelihood function from equation (6.1) as

\[
L(\cdot) = \prod_{i=1}^{n} \left\{ \left( h_{k_i}(t_i|X_i) \right)^{d_{ik}} \right\} \prod_{k=1}^{K} \left\{ e^{-H_k(t_i|X_i)} \right\} \tag{6.2}
\]

Let \( d_{ik} \) indicate whether an individual \( i \) fails from cause \( k \) and \( d_i = \sum_k d_{ik} \) because each individual can fail from one cause only. We can then write the likelihood function as

\[
L(\cdot) = \prod_{k=1}^{K} \prod_{i=1}^{n_k} \left\{ \left( h_{k_i}(t_i|X_i) \right)^{d_{ik}} e^{-H_k(t_i|X_i)} \right\} \tag{6.3}
\]

or the full likelihood function above is proportional to

\[
L(\cdot) = \prod_{k=1}^{K} \prod_{i=1}^{n_k} \left\{ \left[ h_{0k}(t_{ki}) \exp(X'_{ki}\delta_k) \right]^{d_{ik}} \left[ \exp(-H_{0k}(t_{ki})e^{X'_{ki}\delta_k}) \right] \right\} \tag{6.4}
\]

where

- \( h_{0k}(t_{ki}) \) is a baseline hazard function of an event type \( k \).
- \( H_{0k}(t_{ki}) \) is a baseline cumulative hazard function of an event type \( k \).
- \( d_{ik} \) is an indicator of an individual \( i \) fails from event type \( k \).

Let \( h_0(t_i) = \sum_{i=1}^{n} \lambda_i \delta(t - t_i) \) where \( \delta(\cdot) \) is the Dirac delta function. We now make the assumption that the survival time is discrete and the baseline hazard is zero between adjacent failure times. The baseline cumulative hazard \( H_{0k}(t_{ki}) = \sum_{j:j \leq i} \lambda_k \) will be a step function with jumps at \( t_{ki} \).
whenever $d_{ik} = 1$.

Therefore, the likelihood can be written as:

$$ L(h, \delta h, \delta h, \delta) = \prod_{k=1}^{K} \prod_{i=1}^{n_k} \left\{ \lambda_{ki} \exp(X'_{ki} \delta_k) \right\}^{d_{ik}} \left[ \exp \left( - \sum_{j:j \leq i} \lambda_{kj} \exp(X'_{kj} \delta_k) \right) \right] $$

(6.5)

The log likelihood has the following expression:

$$ \ell(h, \delta) = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left\{ d_{ik} (\log \lambda_{ki} + X'_{ki} \delta_k) - \sum_{j:j \leq i} \lambda_{kj} \exp(X'_{kj} \delta_k) \right\} $$

(6.6)

To determine the partial likelihood function, the log likelihood function (6.6) has to be maximized with respect to $\lambda_{ki}$ holding $\delta_k$ fixed. We then obtain

$$ \frac{\partial}{\partial \lambda_{ki}} \ell(h, \delta) = \frac{d_{ik}}{\lambda_{ki}} - \sum_{l;l \geq i} \exp(X'_{kl} \delta_k) = 0 $$

(6.7)

$$ \frac{d_{ik}}{\lambda_{ki}} = \sum_{l;l \geq i} \exp(X'_{kl} \delta_k) $$

$$ \lambda_{ki} = \frac{d_{ik}}{\sum_{l;l \geq i} \exp(X'_{kl} \delta_k)} $$

Plugging (6.7) into (6.4), we will obtain the corresponding partial like-
6.2. MODEL STRUCTURE

The overall partial likelihood is a product of \( K \) partial likelihoods which is one for each type of failure. Therefore, at time \( t_{ki} \), the contribution is the conditional probability that an individual \( i \) fails of cause \( k \) given the risk set at \( t_{ki} \), where all other causes of failure are treated as censored observations. In standard survival analysis, the risk set is defined as the group of individuals who are alive and uncensored at each failure time (Collett,
In a similar manner, the risk sets in the competing risks analysis are defined in standard survival analysis but are modified to allow for competing events. That is, individuals who have a competing event can be removed from all later risk sets for the event of interest (Pintilie, 2006).

In addition, the partial likelihood in (6.8) can be maximized by separately maximizing (Hosmer et al., 2008).

$$L(\delta_k) = \prod_{i=1}^{n_k} \left( \frac{\exp(X'_{ki} \delta_k)}{\sum_{p \in R(t_{ki})} \exp(X'_{kp} \delta_k)} \right)^{d_{ik}} \quad (6.9)$$

The expression above implies that we can simply demonstrate separate model fits for each event type where the remaining event types and the censoring are now censored. For example, for two possible failure types (event 1 and event 2), we could fit two separate proportional hazards models. The first one is modeled by using event type 1 and censoring is indicated by event type 2 and censored observations. The second one is modeled by using event type 2 and censoring is indicated by event type 1 and censored observations.

In the survival part of the joint model, since we consider multiple causes of failure, the competing risks model should be employed to handle this issue. We assume that all event types are of equal interest. The likelihood function is modified from the previous expression (6.8) shown as

$$L(\delta_{0k}, \delta_{1k}|\theta) = \prod_{k=1}^{K} \prod_{i=1}^{n_k} \left( \frac{\exp(\theta'_{ki} \delta_{0k} + X_{ki} \delta_{1k})}{\sum_{p \in R(t_{ki})} \exp(\theta'_{kp} \delta_{0k} + X_{kp} \delta_{1k})} \right)^{d_{ik}} \quad (6.10)$$
This partial likelihood function includes survival times and censored observations which are ordered and denoted by $t_{(k_1)} \leq t_{(k_2)} \leq \ldots \leq t_{(k_n)}$ for all $K$ causes. Then $t_{(k_i)}$ is the $i^{th}$ ordered survival time for cause $k$ at time $t_{(k_i)}$. The set of individuals who do not fail from cause $k$ and are uncensored at time $t_{(k_i)}$ is called the risk set and expressed by $R(t_{(k_i)})$. Moreover, in this survival part we include treatment, $X$, as a covariate. We also include a latent variable for quality of life, $\theta_i^*$, at a failure time point for all event types. This means all observations have their own latent variables as quality of life scores ($\theta_i$) for each follow-up time point and we only calculate $\theta_i^*$ for patients who failed from cause $k$ only. Therefore, the latent variables at this stage are treated as a time-dependent covariate since these values change over time. All these values are determined by linear interpolation between two known $\theta$ points as described in Chapter 3.

### 6.2.2 The Full Likelihood Function

The stereotype model, the regression model and the competing risks model are then combined together to obtain the joint likelihood function. The full likelihood function can be expressed as

$$L(a, b, \phi, \beta, \rho, \delta, \theta) = L(a, b, \phi|\theta)\Phi \left( \theta; \mu(\theta), \sum (\rho) \right) L(\delta|\theta)$$

(6.11)

where the stereotype likelihood $L(a, b, \phi|\theta)$ and the competing risks partial likelihood $L(\delta|\theta)$ are defined in (3.2) and (6.10) respectively. The multivariate normal density function of the latent variable, $\Phi \left( \theta; \mu(\theta), \sum (\rho) \right)$, is
described in Chapter 3 on pages 64-65.

6.3 Parameter estimation

In this chapter, we still consider the Metropolis-Hastings Algorithm (Metropolis et al., 1953) to estimate model parameters simultaneously. The proposal distribution is denoted by \( q(\cdot | \cdot) \) and the criteria need to be applied whether to accept or reject the simulated values. Additionally, the criterion is calculated from ratios of the product of proposal and posterior densities. This criterion is commonly referred to as the acceptance probability and the proposal distribution gives the probability of the new values given the current values. We used the flat prior distributions on parameters so that the posterior density function is proportional to the joint likelihood function (6.11). The Metropolis-Hastings Algorithm steps for updating parameters have been described in Chapter 3.

The Estimation Procedures

The set of parameters that we need to estimate is \( \{a, b, \phi, \beta_0, \beta_1, \rho, \delta, \theta\} \) and we set a uniform prior for all parameters, \( P(\cdot) \propto 1 \).

At the \( t^{th} \) iteration:

1. **Updating parameters in the regression model.**
   
   This step is the same procedure as the estimation procedure step 1 in Chapter 3, section 3.3 on pages 70-71.
2. **Generating missing values.**
   
   This step is the same procedure as the estimation procedure step 2 in Chapter 3, section 3.3 on page 72.

3. **Updating parameters in the stereotype model.**
   
   This step is the same procedure as the estimation procedure step 3 in Chapter 3, section 3.3 on pages 72-75.

4. **Updating parameters in the competing risks model.**

   4.1. Generating latent variables at failure time points. The values at failure time points, $\theta^*$, are determined by linear interpolation between two known $\theta$ points as described in Chapter 3.

   4.2. Generating $\delta_k = \{\delta_{0k}, \delta_{1k}\}$, $k = 1, \ldots, K$, by using $\theta^*$ from previous step and $\delta_k$ is the regression coefficient for event type $k$ based on the partial likelihood function as shown in model (6.10).

   4.2.1 **Updating $\delta_{0k}$.** The posterior distribution of $\delta_{0k}$, $P(\delta_{0k} | \cdot)$, is

   $$
   P(\delta_{0k} | \delta_{1k}, \theta^*) \propto \prod_{k=1}^{K} \prod_{i=1}^{n_k} \left( \frac{\exp(\theta^*_{0k} \delta_{0k} + X_{ki} \delta_{1k})}{\sum_{p \in R(t_{ki})} \exp(\theta^*_{p} \delta_{0k} + X_{kp} \delta_{1k})} \right)^{d_{ik}}
   $$

   Given the current value, $\delta_{0k}^{cur}$, propose a new value, $\delta_{0k}^{new}$, by taking a random draw from the univariate normal distribution, $\delta_{0k}^{new} \sim N(\delta_{0k}^{cur}, \sigma_{\delta_{0k}}^2)$ where $\sigma_{\delta_{0k}}^2 = 1$. The proposed value, $\delta_{0k}^{new}$, is accepted with probability equal to the mini-
mum of 1 and the ratio
\[ r = \frac{q(\delta_{\text{cur}} | \delta_{\text{new}})}{q(\delta_{\text{new}} | \delta_{\text{cur}})} \frac{P(\delta_{\text{new}} | \cdot)}{P(\delta_{\text{cur}} | \cdot)} \]

where the form of the ratio
\[ \frac{q(\delta_{\text{cur}} | \delta_{\text{new}})}{q(\delta_{\text{new}} | \delta_{\text{cur}})} = \frac{1}{\sigma_{\delta_{1k}} \sqrt{2\pi}} \exp\left( -\frac{1}{2\sigma_{\delta_{1k}}^2}(\delta_{\text{cur}} - \delta_{\text{new}})^2 \right) \]
\[ \frac{1}{\sigma_{\delta_{1k}} \sqrt{2\pi}} \exp\left( -\frac{1}{2\sigma_{\delta_{1k}}^2}(\delta_{\text{new}} - \delta_{\text{cur}})^2 \right) = 1 \]

If the proposed value is not accepted, then \( \delta_{1k} \) equal to \( \delta_{\text{cur}}_{1k} \).

4.2.2 Updating \( \delta_{1k} \). The posterior distribution of \( \delta_{1k} \), \( P(\delta_{1k} | \cdot) \), is
\[ P(\delta_{1k} | \delta_{0k}, \theta^*) \propto \prod_{k=1}^{K} \prod_{i=1}^{n_k} \left( \frac{\exp\left( \theta^*_{k_i} \delta_{0k} + X_{k_i} \delta_{1k} \right)}{\sum_{p \in R(t_{ki})} \exp\left( \theta^*_{k_p} \delta_{0k} + X_{k_p} \delta_{1k} \right)} \right)^{d_{ik}} \]

Given the current value, \( \delta_{\text{cur}}_{1k} \), propose a new value, \( \delta_{\text{new}}_{1k} \), by taking a random draw from the univariate normal distribution, \( \delta_{\text{new}}_{1k} \sim N(\delta_{\text{cur}}_{1k}, \sigma_{\delta_{1k}}^2) \) where \( \sigma_{\delta_{1k}}^2 = 1 \). The proposed value, \( \delta_{\text{new}}_{1k} \), is accepted with probability equal to the minimum of 1 and the ratio
\[ r = \frac{q(\delta_{\text{cur}}_{1k} | \delta_{\text{new}}_{1k})P(\delta_{\text{new}}_{1k} | \cdot)}{q(\delta_{\text{new}}_{1k} | \delta_{\text{cur}}_{1k})P(\delta_{\text{cur}}_{1k} | \cdot)} \]

where the form of the ratio
\[ \frac{q(\delta_{\text{cur}}_{1k} | \delta_{\text{new}}_{1k})}{q(\delta_{\text{new}}_{1k} | \delta_{\text{cur}}_{1k})} = \frac{1}{\sigma_{\delta_{1k}} \sqrt{2\pi}} \exp\left( -\frac{1}{2\sigma_{\delta_{1k}}^2}(\delta_{\text{cur}}_{1k} - \delta_{\text{new}}_{1k})^2 \right) \]
\[ \frac{1}{\sigma_{\delta_{1k}} \sqrt{2\pi}} \exp\left( -\frac{1}{2\sigma_{\delta_{1k}}^2}(\delta_{\text{new}}_{1k} - \delta_{\text{cur}}_{1k})^2 \right) = 1 \]

If the proposed value is not accepted, then \( \delta_{1k} \) equal to \( \delta_{\text{cur}}_{1k} \).
5. **Updating the latent variable.**

The posterior distribution of $\theta_i$ is proportional to the joint likelihood function (6.11).

$$P(\theta | a, b, \phi, \beta, \rho, \delta) \propto L(a, b, \phi, \beta, \rho, \delta, \theta)$$

This step is the same procedure as the estimation procedure step 6 in Chapter 3, section 3.3 on pages 79-80.

After we have finished updating all parameters from step 1 to step 5, we changed the iteration from $t$ to $t + 1$ and repeated all five steps until convergence is reached.

### 6.4 Application to Pseudo Data

Since this chapter concerns the competing risks model for the survival part of the joint model, we have to consider the case of multiple failure types such as CDC-B and CDC-C. We use data from the Staccato study but we still consider only CDC-B as a survival endpoint. This is because there were very few patients, five out of 354, who experienced CDC-C at the end of the study. We do not have sufficient information to treat CDC-C as an additional failure type. For that reason the patients in category-C are treated as censored observations. Thus, the CDC-B events (Ananworanich and the Staccato study group, 2006) are classified into two categories for the purpose of the study. Some CDC-B events are grouped
together, without any conditions on the medical point of view, in order to demonstrate the methodologies. The first category is composed of the following symptomatic conditions: candidiasis oropharyngeal, candidiasis vulvovaginal (persistent, recurrent, or unresponsive to therapy) and herpes zoster (shingles), involving at least two distinct episodes or more than one dermatome. The second category consists of the following symptomatic conditions: oral hairy leukoplakia, peripheral neuropathy and pruritic papular eruptions. Henceforth, suppose that patients are at risk with two possible failure types, CDC-B(1)(22 patients) and CDC-B(2)(14 patients), and all other events are treated as censored. We are interested in estimating the risk or incidence of each particular event type over time from an initial stating point \( t_0 \). Also, we assume that both of the failure types are of equal interest and that each subject can experience only one type of event at any particular time. Treatment group assignments have been considered as a covariate in this study. Because the majority of classification is symptomatic conditions, this causes patients in the same treatment to be in the same risk type. We need to generate patients for another treatment to get enough data to demonstrate the method. Therefore, the following results are not based on real data.

Since we used the Bayesian approach to estimate the parameters, we do need to check the convergence of the model. The details of convergence diagnostics have already been given in Chapter 3. We will not show the results of the convergence diagnostics here but we will illustrate the
posterior summary after the chain iterations are discarded. We ran the chain for 100,000 iterations and a burn-in period is 30,000 iterations. The posterior results are summarized as follows:

Table 6.1: The Posterior Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Mean</th>
<th>Standard Error</th>
<th>95% Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereotype Model</td>
<td>$a_2$</td>
<td>0.844</td>
<td>0.025</td>
<td>(0.796, 0.893)</td>
</tr>
<tr>
<td>(5 levels)</td>
<td>$a_3$</td>
<td>1.686</td>
<td>0.025</td>
<td>(1.636, 1.739)</td>
</tr>
<tr>
<td>$L = 5$ levels</td>
<td>$a_4$</td>
<td>1.556</td>
<td>0.027</td>
<td>(1.504, 1.609)</td>
</tr>
<tr>
<td>$J = 32$ questions</td>
<td>$a_5$</td>
<td>0.994</td>
<td>0.033</td>
<td>(0.925, 1.067)</td>
</tr>
<tr>
<td></td>
<td>$b_1$</td>
<td>0.018</td>
<td>0.016</td>
<td>(0.0004, 0.060)</td>
</tr>
<tr>
<td></td>
<td>$b_2$</td>
<td>0.778</td>
<td>0.047</td>
<td>(0.687, 0.871)</td>
</tr>
<tr>
<td></td>
<td>$b_3$</td>
<td>1.435</td>
<td>0.062</td>
<td>(1.317, 1.559)</td>
</tr>
<tr>
<td></td>
<td>$b_4$</td>
<td>1.886</td>
<td>0.077</td>
<td>(1.742, 2.042)</td>
</tr>
<tr>
<td></td>
<td>$b_5$</td>
<td>1.037</td>
<td>0.053</td>
<td>(0.937, 1.148)</td>
</tr>
<tr>
<td></td>
<td>$b_6$</td>
<td>0.174</td>
<td>0.043</td>
<td>(0.090, 0.255)</td>
</tr>
<tr>
<td></td>
<td>$b_7$</td>
<td>0.967</td>
<td>0.052</td>
<td>(0.869, 1.073)</td>
</tr>
<tr>
<td></td>
<td>$b_8$</td>
<td>0.384</td>
<td>0.043</td>
<td>(0.297, 0.471)</td>
</tr>
<tr>
<td></td>
<td>$b_9$</td>
<td>1.716</td>
<td>0.074</td>
<td>(1.569, 1.861)</td>
</tr>
<tr>
<td></td>
<td>$b_{10}$</td>
<td>0.411</td>
<td>0.044</td>
<td>(0.326, 0.498)</td>
</tr>
<tr>
<td></td>
<td>$b_{11}$</td>
<td>1.308</td>
<td>0.060</td>
<td>(1.193, 1.432)</td>
</tr>
<tr>
<td></td>
<td>$b_{12}$</td>
<td>0.517</td>
<td>0.042</td>
<td>(0.435, 0.601)</td>
</tr>
<tr>
<td></td>
<td>$b_{13}$</td>
<td>0.566</td>
<td>0.045</td>
<td>(0.476, 0.654)</td>
</tr>
<tr>
<td></td>
<td>$b_{14}$</td>
<td>0.659</td>
<td>0.045</td>
<td>(0.572, 0.748)</td>
</tr>
<tr>
<td></td>
<td>$b_{15}$</td>
<td>1.061</td>
<td>0.054</td>
<td>(0.959, 1.169)</td>
</tr>
</tbody>
</table>

*continued on next page*
### Table (6.1) continued from previous page

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Mean</th>
<th>Standard Error</th>
<th>95% Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( b_{16} )</td>
<td>1.418</td>
<td>0.063</td>
<td>(1.294, 1.548)</td>
</tr>
<tr>
<td></td>
<td>( b_{17} )</td>
<td>0.839</td>
<td>0.049</td>
<td>(0.746, 0.938)</td>
</tr>
<tr>
<td></td>
<td>( b_{18} )</td>
<td>0.752</td>
<td>0.046</td>
<td>(0.664, 0.846)</td>
</tr>
<tr>
<td></td>
<td>( b_{19} )</td>
<td>0.453</td>
<td>0.042</td>
<td>(0.370, 0.536)</td>
</tr>
<tr>
<td></td>
<td>( b_{20} )</td>
<td>0.768</td>
<td>0.047</td>
<td>(0.676, 0.863)</td>
</tr>
<tr>
<td></td>
<td>( b_{21} )</td>
<td>0.813</td>
<td>0.048</td>
<td>(0.720, 0.911)</td>
</tr>
<tr>
<td></td>
<td>( b_{22} )</td>
<td>1.272</td>
<td>0.060</td>
<td>(1.155, 1.390)</td>
</tr>
<tr>
<td></td>
<td>( b_{23} )</td>
<td>0.770</td>
<td>0.047</td>
<td>(0.678, 0.864)</td>
</tr>
<tr>
<td></td>
<td>( b_{24} )</td>
<td>0.526</td>
<td>0.044</td>
<td>(0.438, 0.614)</td>
</tr>
<tr>
<td></td>
<td>( b_{25} )</td>
<td>1.260</td>
<td>0.060</td>
<td>(1.144, 1.379)</td>
</tr>
<tr>
<td></td>
<td>( b_{26} )</td>
<td>0.550</td>
<td>0.045</td>
<td>(0.464, 0.643)</td>
</tr>
<tr>
<td></td>
<td>( b_{27} )</td>
<td>0.519</td>
<td>0.043</td>
<td>(0.433, 0.604)</td>
</tr>
<tr>
<td></td>
<td>( b_{28} )</td>
<td>1.445</td>
<td>0.064</td>
<td>(1.324, 1.574)</td>
</tr>
<tr>
<td></td>
<td>( b_{29} )</td>
<td>1.967</td>
<td>0.083</td>
<td>(1.808, 2.135)</td>
</tr>
<tr>
<td></td>
<td>( b_{30} )</td>
<td>2.840</td>
<td>0.118</td>
<td>(2.617, 3.079)</td>
</tr>
<tr>
<td></td>
<td>( b_{31} )</td>
<td>1.303</td>
<td>0.062</td>
<td>(1.182, 1.429)</td>
</tr>
<tr>
<td></td>
<td>( b_{32} )</td>
<td>0.885</td>
<td>0.048</td>
<td>(0.788, 0.978)</td>
</tr>
<tr>
<td></td>
<td>( \phi_2 )</td>
<td>0.026</td>
<td>0.017</td>
<td>(0.001, 0.069)</td>
</tr>
<tr>
<td></td>
<td>( \phi_3 )</td>
<td>0.282</td>
<td>0.015</td>
<td>(0.252, 0.313)</td>
</tr>
<tr>
<td></td>
<td>( \phi_4 )</td>
<td>0.554</td>
<td>0.011</td>
<td>(0.531, 0.578)</td>
</tr>
<tr>
<td>Regression Model</td>
<td>( \rho )</td>
<td>0.230</td>
<td>0.021</td>
<td>(0.186, 0.271)</td>
</tr>
<tr>
<td></td>
<td>( \beta_0(t_0) )</td>
<td>-0.020</td>
<td>0.067</td>
<td>(-0.153, 0.114)</td>
</tr>
<tr>
<td></td>
<td>( \beta_0(t_1) )</td>
<td>1.539</td>
<td>0.099</td>
<td>(1.343, 1.732)</td>
</tr>
<tr>
<td></td>
<td>( \beta_0(t_2) )</td>
<td>1.692</td>
<td>0.098</td>
<td>(1.498, 1.884)</td>
</tr>
<tr>
<td></td>
<td>( \beta_0(t_3) )</td>
<td>1.634</td>
<td>0.108</td>
<td>(1.425, 1.851)</td>
</tr>
</tbody>
</table>

*continued on next page*
Table 6.1 shows the estimated parameters and their standard errors with 95% credible interval of the joint model that consists of the stereotype model, the regression model and the competing risks model. The following results are not based on the real data since we generated some of the data to get enough information to demonstrate the method. The coefficients $b$ in the stereotype model suggest the influence of the quality of life score on how patients selected the response category. The coefficients $\phi$ imply how close adjacent response categories are. In the regression model, the coefficients $\beta$ could indicate the treatment effects. It can be seen that there are two sets of coefficients in the competing risks model since we are

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Mean</th>
<th>Standard Error</th>
<th>95% Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereotype Model</td>
<td>$\beta_0(t_4)$</td>
<td>1.765</td>
<td>0.125</td>
<td>(1.516,2.001)</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_5)$</td>
<td>1.623</td>
<td>0.165</td>
<td>(1.276,1.935)</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_6)$</td>
<td>2.048</td>
<td>0.371</td>
<td>(1.409,2.875)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_0)$</td>
<td>0.055</td>
<td>0.115</td>
<td>(-0.163,0.278)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_1)$</td>
<td>0.164</td>
<td>0.150</td>
<td>(-0.127,0.469)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_2)$</td>
<td>-0.176</td>
<td>0.148</td>
<td>(-0.475,0.113)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_3)$</td>
<td>0.167</td>
<td>0.163</td>
<td>(-0.164,0.477)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_4)$</td>
<td>-0.231</td>
<td>0.191</td>
<td>(-0.606,0.147)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_5)$</td>
<td>-0.117</td>
<td>0.239</td>
<td>(-0.584,0.343)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_6)$</td>
<td>-0.778</td>
<td>0.527</td>
<td>(-1.893,0.155)</td>
</tr>
<tr>
<td>Competing Risks Model</td>
<td>$\delta_0(1)$</td>
<td>-0.297</td>
<td>0.198</td>
<td>(-0.692,0.085)</td>
</tr>
<tr>
<td></td>
<td>$\delta_1(1)$</td>
<td>-0.199</td>
<td>0.474</td>
<td>(-1.179,0.690)</td>
</tr>
<tr>
<td></td>
<td>$\delta_0(2)$</td>
<td>-1.712</td>
<td>0.395</td>
<td>(-2.561,-0.990)</td>
</tr>
<tr>
<td></td>
<td>$\delta_1(2)$</td>
<td>-0.413</td>
<td>0.663</td>
<td>(-1.810,0.790)</td>
</tr>
</tbody>
</table>
interested in two failure types: CDC-B(1) and CDC-B(2). We use the competing risks model because sometimes the covariate effects will differ significantly across the different failure types. However, the example in this chapter gives similar results in both failure types. In this particular case, the negative coefficients $\delta_1(1)$ and $\delta_1(2)$ show the patients who received a continuous treatment have a lower hazard of progression to CDC-B(1) and CDC-B(2) respectively. In the same way, the negative coefficients $\delta_0(1)$ and $\delta_0(2)$ suggest that a higher quality of life is related to a reduced hazard of having CDC-B(1) and CDC-B(2).
Chapter 7

Conclusion

We developed joint models for the association between longitudinal ordinal responses and the survival process using data from the Staccato study for the application. The aim of the analysis was to investigate the relationship between the quality of life and the time to CDC-B progression. Also, another aim was to simultaneously determine the factors that affect the quality of life and time to CDC-B progression. We used the stereotype model to model the quality of life of patients since this model is capable of capturing the ordinal nature of the data. This model also gives a natural interpretation. That is, a patient at a higher score of quality of life should have a higher probability of responding in a higher response category. This is because the direction of ordering can be recorded from poor to excellent where a low response implies a poor quality of life. Moreover, the stereotype model can improve the fit by adding extra score parameters, but it still has the advantage of requiring only a single parameter to
describe the effect of a predictor. We would prefer the stereotype model on the grounds of parsimony and flexibility over other ordinal logistic regression models.

We subsequently considered the Cox proportional hazards model to model time to CDC-B progression. The covariate variables such as sex, age and treatment were included in the joint model to predict the hazard of an individual. However, the model comparison test showed that only treatment and the shared latent variable were enough to model the change of hazard function although other covariates could be important. The results we obtained here are the same for both the Bayesian joint model and the semiparametric joint model. Note that the shared latent variable indicates the unobservable quality of life of an individual. The advantage of using the latent variable is that one can reduce the number of variables since it contains all information on the quality of life. That is, in the stereotype model and the Cox proportional hazards model, the latent variable was assumed to be a covariate representing quality of life for each individual. Another advantage is that the latent variable generates a more realistic quality of life score than a summed scale score of the observed responses. This is because the latent variable derives from a model which allow a nonlinear relationship between satisfaction and question response categories.

For the Bayesian joint model, all questions in the MOS-HIV questionnaire were modeled by the stereotype model and the binary model. Al-
though the number of response categories was different, we still can use the stereotype model to model all items/questions simultaneously. We used the binary model which is a special case of the stereotype model with $L = 2$ to model the questions that have binary responses. The partial survival likelihood function was used in the survival component. This means the baseline hazard was canceled out in the joint likelihood function. The shared latent variable was assumed to be a continuous variable. The models were fitted using the Markov Chain Monte Carlo (MCMC) techniques through the Metropolis-Hasting Algorithm. The results from the Bayesian joint model suggested that the patients who had a better quality of life were associated with the lower hazard of the progression to CDC-B. The patients on continuous treatment had a better quality of life compared with the patients on CD4-guided interruption treatment.

For the semiparametric joint model, the baseline hazard was included in the joint likelihood function. This caused our joint model to become more complicated since the baseline hazard had no parametric form and was treated as a step function. The shared latent variable was defined as a discrete variable. The benefit of having a discrete latent variable in the model is that we could classify patients to their most likely group. Each group was composed of a set of individuals who were homogeneous with respect to their characteristics or their attitudes from quality of life survey responses. The latent variables then represented the average of quality of life scores of patients in each group. The stereotype model was used
to model the ordinal responses which had five categories from the MOS-HIV questionnaire. We determined the joint likelihood function by using the idea of finite mixture models which are often used for the purpose of clustering. The EM algorithm was employed to estimate all unknown parameters including the unknown baseline hazard through iterative estimation. The results from the semiparametric joint model showed that we could classify the patients from the Staccato study by their quality of life scores into five groups. The patients who were in the group of high value of quality of life score had a better quality of life.

The results from both models the Bayesian joint model and the semiparametric joint model are consistent in view of assessing quality of life and survival analysis. The better quality of life was related to a reduced hazard of having CDC-B. Furthermore, the patients on continuous treatment had a lower hazard of CDC-B progression and had a better quality of life compared with the patients on CD4-guided interruption treatment. In this thesis, we demonstrate how these two approaches can be done in terms of joint modeling. We cannot compare the value of parameter estimates between these two approaches directly. The Bayesian joint model is more flexible in that the quality of life score, $\theta_i$, is allowed to vary by individual and by time point. Furthermore, grouping of individuals to quality of life score categories as is done in the semiparametric approach can be quite arbitrary. We also can add more covariates that we assume to affect quality of life in the regression part of the Bayesian joint model. In the re-
gression model, we can relate different covariates with the latent variable and the stereotype model can in turn relate the questionnaire responses with the latent variables.

Generally, standard methods for the joint model of longitudinal and survival data allow for one failure type with an assumption of independent censoring. However, we extended the joint model approach to allow for competing risks data. The method was applied to modified data from the Staccato study, since the original data did not suit the competing risks model and our aim was to illustrate the joint model approach in another aspect. At the survival endpoint, we presented two competing risks: CDC-B(1) and CDC-B(2). The competing risks model is useful when a follow-up setting may terminate due to more than one failure type and the covariate effects might differ significantly across the different failure types. The results from our analysis were similar in both cases. That is, the patients on the continuous treatment had a lower hazard on CDC-B(1) and CDC-B(2). The results also showed the lower hazard on CDC-B(1) and CDC-B(2) when the patients had a better quality of life. However, these results are not based on the real data since we used pseudo data to demonstrate the method.

Joint model approaches are appropriate when the quality of life and the survival endpoints are likely to be related. The significance of the contribution of this thesis is the development of two joint model approaches incorporated with the stereotype models. This is a new application since
no literature has examined the stereotype model within a joint model. As mentioned, the stereotype model is very useful and flexible, especially with the ordinal data. In spite of this, our joint model can be applied to any ordinal data and survival data. Furthermore, the joint model can cope with the problem of missing data due to dropout prior to death or withdrawal before the end of the study. This is because the joint model approach has the ability to gain information from both components effectively. In practice, it is relatively complicated by the nature of the data; for example, the quality of life data might not be completely collected at the follow-up time and the uncensored survival information might not be available for all subjects. These two components could be linked through the latent variable when the joint model method is used. The latent variable is able to gather all information from several quality of life variables together. This could reduce the number of parameters in the model.

The models were implemented in the R programming language. However, we found that the model evaluation is computationally time consuming since our models have a complex structure and high-dimensional data. Although MCMC techniques are particularly useful for a complex model, the drawback of MCMC is computational time required. The sampling procedures that underlie the MCMC methods generally require a very large number of iterations before model parameters can be reliably estimated. It also needs to reach convergence of all parameters for estimates to be reliable. The difficulty is when there are many parameters in
the model to estimate. It takes several days for estimation run and more for more complex models or when analyzing large amount of data. Parameter estimation by using the EM algorithm for semiparametric joint model was much quicker in this study. According to the slow execution, MCMC should be implemented on a high level programming environment that can execute much faster such as C programming language.

Some research areas will be considered in future work. First, we will study the model misspecification issue for the joint model. If the item response model is misspecified, are the parameter estimates for the survival model (or the regression model) still valid? If not, then a good choice of the item response model is important and it is better to choose the item response model which has as few assumptions as possible. Also, we will evaluate the situations where the choice of the item response model is less important.

As mentioned in Chapter 4, the discrete latent variable can be extended to the cluster analysis. This methodology can classify the individuals to their most likely group. The other benefit is to reduce the dimensionality of the problem so that there are fewer parameters in the model to be estimated. Another possible further research is improvement on the R-code of programming to make it faster and easy to use, such as writing an R package for the joint model.
Appendix A

Convergence Diagnostics and Posterior Summary
Model 1

This model consists of two stereotype models, the binary model, the regression model and the Cox proportional hazard model. There are two predictors (quality of life and treatment) in the survival component. The posterior summary for model 1 has already been given in chapter 5. Thus, this section will show only the convergence diagnostics of model 1.

Trace and Density Plots

Figure A.1: $\rho$, a correlation of time
Figure A.2: $a_2$, $a_3$, $a_4$ and $a_5$ from the stereotype model ($L = 5$)

Figure A.3: $b_1$, $b_2$, $b_3$ and $b_4$ from the stereotype model ($L = 5$)
Figure A.4: $b_5$, $b_6$, $b_7$, $b_8$, $b_9$, $b_{10}$, $b_{11}$ and $b_{12}$ from the stereotype model ($L = 5$)
Figure A.5: \( b_{13}, b_{14}, b_{15}, b_{16}, b_{17}, b_{18}, b_{19}, \) and \( b_{20} \) from the stereotype model (\( L = 5 \))
Figure A.6: $b_{21}, b_{22}, b_{23}, b_{24}, b_{25}, b_{26}, b_{27}$ and $b_{28}$ from the stereotype model ($L = 5$)
Figure A.7: $b_{29}, b_{30}, b_{31}$ and $b_{32}$ from the stereotype model ($L = 5$)

Figure A.8: $\phi_2, \phi_3$ and $\phi_4$ from the stereotype model ($L = 5$)
Figure A.9: $a_2$ and $a_3$ from the stereotype model ($L = 3$)

Figure A.10: $b_1, b_2, b_3$ and $b_4$ from the stereotype model ($L = 3$)
Figure A.11: $b_5$ and $b_6$ from the stereotype model ($L = 3$)

Figure A.12: $\phi_2$ from the stereotype model ($L = 3$)
Figure A.13: $a^*$, $b_1^*$ and $b_2^*$ from the binary model

Figure A.14: $\beta_0$ at time $t_0 - t_3$ from the regression model
Figure A.15: $\beta_0$ at time $t_4 - t_6$ and $\beta_1$ at time $t_0 - t_4$ from the regression model
Figure A.16: $\beta_1$ at time $t_5 - t_6$ from the regression model

Figure A.17: $\delta_0$ and $\delta_1$ from the Cox proportional hazards model
Model 2

This model consists of two stereotype models, the binary model, the regression model and the Cox proportional hazard model. There are three predictors (quality of life, treatment and initial CD4 count) in the survival component. Graphical diagnostics by trace and density plots and two statistical diagnostics: Gelman and Rubin’s multiple sequence diagnostic and Heidelberger and Welch Diagnostic are given. The posterior results are summarized as follows.

Trace and Density Plots

Figure A.18: $\rho$, a correlation of time
APPENDIX A. CONVERGENCE DIAGNOSTICS AND POSTERIOR SUMMARY

(a) $a_2$ and $a_3$

(b) $a_4$ and $a_5$

Figure A.19: $a_2$, $a_3$, $a_4$ and $a_5$ from the stereotype model ($L = 5$)

(a) $b_1$ and $b_2$

(b) $b_3$ and $b_4$

Figure A.20: $b_1$, $b_2$, $b_3$ and $b_4$ from the stereotype model ($L = 5$)
Figure A.21: \( b_5, b_6, b_7, b_8, b_9, b_{10}, b_{11} \) and \( b_{12} \) from the stereotype model \((L = 5)\)
Figure A.22: $b_{13}, b_{14}, b_{15}, b_{16}, b_{17}, b_{18}, b_{19}$ and $b_{20}$ from the stereotype model ($L = 5$)
Figure A.23: $b_{21}, b_{22}, b_{23}, b_{24}, b_{25}, b_{26}, b_{27}$ and $b_{28}$ from the stereotype model ($L = 5$)
Figure A.24: $b_{29}, b_{30}, b_{31}$ and $b_{32}$ from the stereotype model ($L = 5$)

Figure A.25: $\phi_2, \phi_3$ and $\phi_4$ from the stereotype model ($L = 5$)
Figure A.26: $a_2$ and $a_3$ from the stereotype model ($L = 3$)

Figure A.27: $b_1$, $b_2$, $b_3$ and $b_4$ from the stereotype model ($L = 3$)
Figure A.28: $b_5$ and $b_6$ from the stereotype model ($L = 3$)

Figure A.29: $\phi_2$ from the stereotype model ($L = 3$)
Figure A.30: $a^*$, $b^*_1$ and $b^*_2$ from the binary model

(a) $a^*$

(b) $b^*_1$ and $b^*_2$

Figure A.31: $\beta_0(t_0)$ and $\beta_0(t_1)$

(a) $\beta_0(t_0)$ and $\beta_0(t_1)$

(b) $\beta_0(t_2)$ and $\beta_0(t_3)$

Figure A.31: $\beta_0$ at time $t_0 - t_3$ from the regression model
Figure A.32: $\beta_0$ at time $t_4 - t_6$ and $\beta_1$ at time $t_0 - t_4$ from the regression model
Figure A.33: $\beta_1$ at time $t_5 - t_6$ from the regression model

(a) $\delta_0$ and $\delta_1$
(b) $\delta_2$

Figure A.34: $\delta_0$, $\delta_1$ and $\delta_2$ from the Cox proportional hazards model
### Statistical Diagnostics

#### Gelman and Rubin’s Multiple Sequence Diagnostic

Table A.1: **Gelman and Rubin Diagnostic, Model 2**

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Point est. $(\hat{R})$</th>
<th>97.5% quantile</th>
<th>Multivariate psrf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereotype Model</td>
<td>$a_2$</td>
<td>1.06</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>(5 levels)</td>
<td>$a_3$</td>
<td>1.29</td>
<td>1.98</td>
<td>1.31</td>
</tr>
<tr>
<td>$L = 5$ levels</td>
<td>$a_4$</td>
<td>1.11</td>
<td>1.42</td>
<td></td>
</tr>
<tr>
<td>$J = 32$ questions</td>
<td>$a_5$</td>
<td>1.13</td>
<td>1.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_1$</td>
<td>1.01</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_2$</td>
<td>1.01</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_3$</td>
<td>1.00</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_4$</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_5$</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_6$</td>
<td>1.01</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_7$</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_8$</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_9$</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{10}$</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{11}$</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{12}$</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{13}$</td>
<td>1.00</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{14}$</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

*continued on next page*
<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Point est. ( (\hat{R}) )</th>
<th>97.5% quantile</th>
<th>Multivariate psrf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multivariate</td>
<td>( b_{15} )</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{16} )</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{17} )</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{18} )</td>
<td>1.01</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{19} )</td>
<td>1.00</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{20} )</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{21} )</td>
<td>1.00</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{22} )</td>
<td>1.02</td>
<td>1.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{23} )</td>
<td>1.04</td>
<td>1.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{24} )</td>
<td>1.03</td>
<td>1.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{25} )</td>
<td>1.03</td>
<td>1.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{26} )</td>
<td>1.00</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{27} )</td>
<td>1.01</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{28} )</td>
<td>1.00</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{29} )</td>
<td>1.00</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{30} )</td>
<td>1.00</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{31} )</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_{32} )</td>
<td>1.00</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \phi_2 )</td>
<td>1.02</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \phi_3 )</td>
<td>1.35</td>
<td>2.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \phi_4 )</td>
<td>1.05</td>
<td>1.22</td>
<td></td>
</tr>
<tr>
<td>Stereotype Model</td>
<td>( a_2 )</td>
<td>1.03</td>
<td>1.07</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>( a_3 )</td>
<td>1.04</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_1 )</td>
<td>1.00</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b_2 )</td>
<td>1.01</td>
<td>1.01</td>
<td></td>
</tr>
</tbody>
</table>

**continued on next page**
<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Point est. ( \hat{R} )</th>
<th>97.5% quantile</th>
<th>Multivariate psrf</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi_2 )</td>
<td></td>
<td>1.02</td>
<td>1.06</td>
<td>-</td>
</tr>
<tr>
<td>Binary Model</td>
<td>( \phi_2 )</td>
<td>1.02</td>
<td>1.06</td>
<td>-</td>
</tr>
<tr>
<td>( Q = 2 ) questions</td>
<td>( \phi_2 )</td>
<td>1.02</td>
<td>1.06</td>
<td>-</td>
</tr>
<tr>
<td>Regression Model</td>
<td>( \beta )</td>
<td>1.00</td>
<td>1.00</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( \beta_0(t_0) )</td>
<td>1.00</td>
<td>1.01</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( \beta_0(t_1) )</td>
<td>1.00</td>
<td>1.00</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( \beta_0(t_2) )</td>
<td>1.00</td>
<td>1.01</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( \beta_0(t_3) )</td>
<td>1.00</td>
<td>1.02</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>( \beta_0(t_4) )</td>
<td>1.00</td>
<td>1.00</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( \beta_0(t_5) )</td>
<td>1.01</td>
<td>1.05</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( \beta_0(t_6) )</td>
<td>1.08</td>
<td>1.32</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( \beta_1(t_0) )</td>
<td>1.00</td>
<td>1.01</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( \beta_1(t_1) )</td>
<td>1.01</td>
<td>1.03</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( \beta_1(t_2) )</td>
<td>1.00</td>
<td>1.00</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( \beta_1(t_3) )</td>
<td>1.00</td>
<td>1.01</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>( \beta_1(t_4) )</td>
<td>1.00</td>
<td>1.01</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( \beta_1(t_5) )</td>
<td>1.02</td>
<td>1.08</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( \beta_1(t_6) )</td>
<td>1.02</td>
<td>1.03</td>
<td>-</td>
</tr>
<tr>
<td>Cox’s Model</td>
<td>( \delta )</td>
<td>1.02</td>
<td>1.11</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( \delta_1 )</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>
Heidelberger and Welch Diagnostic

Table A.2: **Heidelberger and Welch Diagnostic, Model 2**  
(running the Heidelberger and Welch Diagnostic at $\epsilon = 0.1$ and $\alpha = 0.05$)

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Stationary Test</th>
<th>Halfwidth Test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereotype Model</td>
<td>$a_2$</td>
<td>passed</td>
<td>passed</td>
<td>0.234</td>
</tr>
<tr>
<td>(5 levels)</td>
<td>$a_3$</td>
<td>passed</td>
<td>passed</td>
<td>0.297</td>
</tr>
<tr>
<td>$L = 5$ levels</td>
<td>$a_4$</td>
<td>passed</td>
<td>passed</td>
<td>0.120</td>
</tr>
<tr>
<td>$J = 32$ questions</td>
<td>$a_5$</td>
<td>passed</td>
<td>passed</td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td>$b_1$</td>
<td>passed</td>
<td>passed</td>
<td>0.576</td>
</tr>
<tr>
<td></td>
<td>$b_2$</td>
<td>passed</td>
<td>passed</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>$b_3$</td>
<td>passed</td>
<td>passed</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td>$b_4$</td>
<td>passed</td>
<td>passed</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td>$b_5$</td>
<td>passed</td>
<td>passed</td>
<td>0.214</td>
</tr>
<tr>
<td></td>
<td>$b_6$</td>
<td>passed</td>
<td>passed</td>
<td>0.229</td>
</tr>
<tr>
<td></td>
<td>$b_7$</td>
<td>passed</td>
<td>passed</td>
<td>0.258</td>
</tr>
<tr>
<td></td>
<td>$b_8$</td>
<td>passed</td>
<td>passed</td>
<td>0.373</td>
</tr>
<tr>
<td></td>
<td>$b_9$</td>
<td>passed</td>
<td>passed</td>
<td>0.334</td>
</tr>
<tr>
<td></td>
<td>$b_{10}$</td>
<td>passed</td>
<td>passed</td>
<td>0.155</td>
</tr>
<tr>
<td></td>
<td>$b_{11}$</td>
<td>passed</td>
<td>passed</td>
<td>0.191</td>
</tr>
</tbody>
</table>

*continued on next page*
### APPENDIX A. CONVERGENCE DIAGNOSTICS AND POSTERIOR SUMMARY

#### Table (A.2) continued from previous page

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Stationary Test</th>
<th>Halfwidth Test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b_{12}$</td>
<td>passed</td>
<td>passed</td>
<td>0.595</td>
</tr>
<tr>
<td></td>
<td>$b_{13}$</td>
<td>passed</td>
<td>passed</td>
<td>0.305</td>
</tr>
<tr>
<td></td>
<td>$b_{14}$</td>
<td>passed</td>
<td>passed</td>
<td>0.334</td>
</tr>
<tr>
<td></td>
<td>$b_{15}$</td>
<td>passed</td>
<td>passed</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>$b_{16}$</td>
<td>passed</td>
<td>passed</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td>$b_{17}$</td>
<td>passed</td>
<td>passed</td>
<td>0.189</td>
</tr>
<tr>
<td></td>
<td>$b_{18}$</td>
<td>passed</td>
<td>passed</td>
<td>0.169</td>
</tr>
<tr>
<td></td>
<td>$b_{19}$</td>
<td>passed</td>
<td>passed</td>
<td>0.505</td>
</tr>
<tr>
<td></td>
<td>$b_{20}$</td>
<td>passed</td>
<td>passed</td>
<td>0.356</td>
</tr>
<tr>
<td></td>
<td>$b_{21}$</td>
<td>passed</td>
<td>passed</td>
<td>0.315</td>
</tr>
<tr>
<td></td>
<td>$b_{22}$</td>
<td>passed</td>
<td>passed</td>
<td>0.348</td>
</tr>
<tr>
<td></td>
<td>$b_{23}$</td>
<td>passed</td>
<td>passed</td>
<td>0.172</td>
</tr>
<tr>
<td></td>
<td>$b_{24}$</td>
<td>passed</td>
<td>passed</td>
<td>0.500</td>
</tr>
<tr>
<td></td>
<td>$b_{25}$</td>
<td>passed</td>
<td>passed</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>$b_{26}$</td>
<td>passed</td>
<td>passed</td>
<td>0.188</td>
</tr>
<tr>
<td></td>
<td>$b_{27}$</td>
<td>passed</td>
<td>passed</td>
<td>0.214</td>
</tr>
<tr>
<td></td>
<td>$b_{28}$</td>
<td>passed</td>
<td>passed</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>$b_{29}$</td>
<td>passed</td>
<td>passed</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>$b_{30}$</td>
<td>passed</td>
<td>passed</td>
<td>0.268</td>
</tr>
<tr>
<td></td>
<td>$b_{31}$</td>
<td>passed</td>
<td>passed</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>$b_{32}$</td>
<td>passed</td>
<td>passed</td>
<td>0.599</td>
</tr>
<tr>
<td></td>
<td>$\phi_2$</td>
<td>passed</td>
<td>passed</td>
<td>0.210</td>
</tr>
<tr>
<td></td>
<td>$\phi_3$</td>
<td>passed</td>
<td>passed</td>
<td>0.354</td>
</tr>
<tr>
<td></td>
<td>$\phi_4$</td>
<td>passed</td>
<td>passed</td>
<td>0.199</td>
</tr>
<tr>
<td>Stereotype Model</td>
<td>$\alpha_2$</td>
<td>passed</td>
<td>passed</td>
<td>0.724</td>
</tr>
</tbody>
</table>

*continued on next page*
<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Stationary Test</th>
<th>Halfwidth Test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3 levels)</td>
<td>$a_3$</td>
<td>passed</td>
<td>passed</td>
<td>0.363</td>
</tr>
<tr>
<td>$L = 3$ levels</td>
<td>$b_1$</td>
<td>passed</td>
<td>passed</td>
<td>0.254</td>
</tr>
<tr>
<td>$J = 6$ questions</td>
<td>$b_2$</td>
<td>passed</td>
<td>passed</td>
<td>0.171</td>
</tr>
<tr>
<td></td>
<td>$b_3$</td>
<td>passed</td>
<td>passed</td>
<td>0.405</td>
</tr>
<tr>
<td></td>
<td>$b_4$</td>
<td>passed</td>
<td>passed</td>
<td>0.465</td>
</tr>
<tr>
<td></td>
<td>$b_5$</td>
<td>passed</td>
<td>passed</td>
<td>0.778</td>
</tr>
<tr>
<td></td>
<td>$b_6$</td>
<td>passed</td>
<td>passed</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>$\phi_2$</td>
<td>passed</td>
<td>passed</td>
<td>0.752</td>
</tr>
<tr>
<td>Binary Model</td>
<td>$a$</td>
<td>passed</td>
<td>passed</td>
<td>0.887</td>
</tr>
<tr>
<td>$Q = 2$ questions</td>
<td>$b_1$</td>
<td>passed</td>
<td>passed</td>
<td>0.740</td>
</tr>
<tr>
<td></td>
<td>$b_2$</td>
<td>passed</td>
<td>passed</td>
<td>0.213</td>
</tr>
<tr>
<td>Regression Model</td>
<td>$\rho$</td>
<td>passed</td>
<td>passed</td>
<td>0.933</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_0)$</td>
<td>passed</td>
<td>passed</td>
<td>0.602</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_1)$</td>
<td>passed</td>
<td>passed</td>
<td>0.407</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_2)$</td>
<td>passed</td>
<td>passed</td>
<td>0.101</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_3)$</td>
<td>passed</td>
<td>passed</td>
<td>0.278</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_4)$</td>
<td>passed</td>
<td>passed</td>
<td>0.495</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_5)$</td>
<td>passed</td>
<td>passed</td>
<td>0.323</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_6)$</td>
<td>passed</td>
<td>passed</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_0)$</td>
<td>passed</td>
<td>passed</td>
<td>0.680</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_1)$</td>
<td>passed</td>
<td>passed</td>
<td>0.875</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_2)$</td>
<td>passed</td>
<td>passed</td>
<td>0.962</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_3)$</td>
<td>passed</td>
<td>passed</td>
<td>0.515</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_4)$</td>
<td>passed</td>
<td>passed</td>
<td>0.340</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_5)$</td>
<td>passed</td>
<td>passed</td>
<td>0.416</td>
</tr>
</tbody>
</table>

*continued on next page*
APPENDIX A. CONVERGENCE DIAGNOSTICS AND POSTERIOR SUMMARY

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Stationary Test</th>
<th>Halfwidth Test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>β1 (t0)</td>
<td>passed</td>
<td>passed</td>
<td></td>
<td>0.263</td>
</tr>
<tr>
<td>Cox’s Model</td>
<td>δ0</td>
<td>passed</td>
<td>passed</td>
<td>0.153</td>
</tr>
<tr>
<td></td>
<td>δ1</td>
<td>passed</td>
<td>passed</td>
<td>0.417</td>
</tr>
<tr>
<td></td>
<td>δ2</td>
<td>passed</td>
<td>passed</td>
<td>0.824</td>
</tr>
</tbody>
</table>

The Posterior Summary

Table A.3: The Posterior Summary, Model 2

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Mean</th>
<th>Standard Error</th>
<th>95% Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereotype Model</td>
<td>a2</td>
<td>0.8983</td>
<td>0.026</td>
<td>(0.843,0.945)</td>
</tr>
<tr>
<td>(5 levels)</td>
<td>a3</td>
<td>1.843</td>
<td>0.024</td>
<td>(1.793,1.890)</td>
</tr>
<tr>
<td>L = 5 levels</td>
<td>a4</td>
<td>1.686</td>
<td>0.029</td>
<td>(1.632,1.739)</td>
</tr>
<tr>
<td>J = 32 questions</td>
<td>a5</td>
<td>0.961</td>
<td>0.037</td>
<td>(0.882,1.035)</td>
</tr>
<tr>
<td></td>
<td>b1</td>
<td>0.228</td>
<td>0.067</td>
<td>(0.094,0.362)</td>
</tr>
<tr>
<td></td>
<td>b2</td>
<td>1.764</td>
<td>0.083</td>
<td>(1.603,1.927)</td>
</tr>
<tr>
<td></td>
<td>b3</td>
<td>2.714</td>
<td>0.106</td>
<td>(2.509,2.920)</td>
</tr>
<tr>
<td></td>
<td>b4</td>
<td>3.705</td>
<td>0.137</td>
<td>(3.438,3.982)</td>
</tr>
<tr>
<td></td>
<td>b5</td>
<td>2.419</td>
<td>0.105</td>
<td>(2.212,2.625)</td>
</tr>
<tr>
<td></td>
<td>b6</td>
<td>0.763</td>
<td>0.071</td>
<td>(0.626,0.907)</td>
</tr>
<tr>
<td></td>
<td>b7</td>
<td>2.372</td>
<td>0.103</td>
<td>(2.178,2.579)</td>
</tr>
<tr>
<td></td>
<td>b8</td>
<td>1.185</td>
<td>0.075</td>
<td>(1.038,1.335)</td>
</tr>
<tr>
<td></td>
<td>b9</td>
<td>3.789</td>
<td>0.140</td>
<td>(3.516,4.067)</td>
</tr>
</tbody>
</table>

continued on next page
<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Mean</th>
<th>Standard Error</th>
<th>95% Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>$b_{10}$</td>
<td>1.282</td>
<td>0.077</td>
<td>(1.135,1.434)</td>
</tr>
<tr>
<td>11</td>
<td>$b_{11}$</td>
<td>3.046</td>
<td>0.119</td>
<td>(2.818,3.282)</td>
</tr>
<tr>
<td>12</td>
<td>$b_{12}$</td>
<td>1.241</td>
<td>0.077</td>
<td>(1.090,1.396)</td>
</tr>
<tr>
<td>13</td>
<td>$b_{13}$</td>
<td>1.510</td>
<td>0.081</td>
<td>(1.351,1.434)</td>
</tr>
<tr>
<td>14</td>
<td>$b_{14}$</td>
<td>1.664</td>
<td>0.084</td>
<td>(1.501,1.835)</td>
</tr>
<tr>
<td>15</td>
<td>$b_{15}$</td>
<td>2.667</td>
<td>0.112</td>
<td>(2.451,2.895)</td>
</tr>
<tr>
<td>16</td>
<td>$b_{16}$</td>
<td>3.502</td>
<td>0.136</td>
<td>(3.235,3.774)</td>
</tr>
<tr>
<td>17</td>
<td>$b_{17}$</td>
<td>2.117</td>
<td>0.096</td>
<td>(1.930,2.304)</td>
</tr>
<tr>
<td>18</td>
<td>$b_{18}$</td>
<td>1.773</td>
<td>0.085</td>
<td>(1.606,1.944)</td>
</tr>
<tr>
<td>19</td>
<td>$b_{19}$</td>
<td>1.086</td>
<td>0.072</td>
<td>(0.940,1.231)</td>
</tr>
<tr>
<td>20</td>
<td>$b_{20}$</td>
<td>1.673</td>
<td>0.083</td>
<td>(1.508,1.841)</td>
</tr>
<tr>
<td>21</td>
<td>$b_{21}$</td>
<td>1.847</td>
<td>0.086</td>
<td>(1.684,2.021)</td>
</tr>
<tr>
<td>22</td>
<td>$b_{22}$</td>
<td>2.718</td>
<td>0.112</td>
<td>(2.499,2.943)</td>
</tr>
<tr>
<td>23</td>
<td>$b_{23}$</td>
<td>1.709</td>
<td>0.084</td>
<td>(1.547,1.877)</td>
</tr>
<tr>
<td>24</td>
<td>$b_{24}$</td>
<td>1.311</td>
<td>0.077</td>
<td>(1.157,1.465)</td>
</tr>
<tr>
<td>25</td>
<td>$b_{25}$</td>
<td>2.749</td>
<td>0.110</td>
<td>(2.532,2.965)</td>
</tr>
<tr>
<td>26</td>
<td>$b_{26}$</td>
<td>1.203</td>
<td>0.075</td>
<td>(1.060,1.354)</td>
</tr>
<tr>
<td>27</td>
<td>$b_{27}$</td>
<td>1.121</td>
<td>0.072</td>
<td>(0.986,1.266)</td>
</tr>
<tr>
<td>28</td>
<td>$b_{28}$</td>
<td>2.772</td>
<td>0.108</td>
<td>(2.563,2.991)</td>
</tr>
<tr>
<td>29</td>
<td>$b_{29}$</td>
<td>3.459</td>
<td>0.140</td>
<td>(3.196,3.793)</td>
</tr>
<tr>
<td>30</td>
<td>$b_{30}$</td>
<td>4.232</td>
<td>0.173</td>
<td>(3.897,4.584)</td>
</tr>
<tr>
<td>31</td>
<td>$b_{31}$</td>
<td>2.498</td>
<td>0.105</td>
<td>(2.302,2.709)</td>
</tr>
<tr>
<td>32</td>
<td>$b_{32}$</td>
<td>1.913</td>
<td>0.092</td>
<td>(1.737,2.095)</td>
</tr>
<tr>
<td>33</td>
<td>$\phi_2$</td>
<td>0.099</td>
<td>0.010</td>
<td>(0.075,0.115)</td>
</tr>
<tr>
<td>34</td>
<td>$\phi_3$</td>
<td>0.272</td>
<td>0.009</td>
<td>(0.252,0.289)</td>
</tr>
</tbody>
</table>

*continued on next page*
## APPENDIX A. CONVERGENCE DIAGNOSTICS AND POSTERIOR SUMMARY

### Table (A.3) continued from previous page

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Mean</th>
<th>Standard Error</th>
<th>95% Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\phi_4$</td>
<td>0.588</td>
<td>0.008</td>
<td>(0.571, 0.602)</td>
</tr>
<tr>
<td>Stereotype Model</td>
<td>$a_2$</td>
<td>1.501</td>
<td>0.052</td>
<td>(1.397, 1.605)</td>
</tr>
<tr>
<td>(3 levels)</td>
<td>$a_3$</td>
<td>2.101</td>
<td>0.055</td>
<td>(1.989, 2.210)</td>
</tr>
<tr>
<td>$L = 3$ levels</td>
<td>$b_1$</td>
<td>0.608</td>
<td>0.067</td>
<td>(0.475, 0.747)</td>
</tr>
<tr>
<td>$J = 6$ questions</td>
<td>$b_2$</td>
<td>1.306</td>
<td>0.097</td>
<td>(1.127, 1.502)</td>
</tr>
<tr>
<td></td>
<td>$b_3$</td>
<td>0.981</td>
<td>0.080</td>
<td>(0.826, 1.147)</td>
</tr>
<tr>
<td></td>
<td>$b_4$</td>
<td>1.043</td>
<td>0.085</td>
<td>(0.884, 1.217)</td>
</tr>
<tr>
<td></td>
<td>$b_5$</td>
<td>1.444</td>
<td>0.103</td>
<td>(1.251, 1.657)</td>
</tr>
<tr>
<td></td>
<td>$b_6$</td>
<td>1.547</td>
<td>0.116</td>
<td>(1.332, 1.789)</td>
</tr>
<tr>
<td></td>
<td>$\phi_2$</td>
<td>0.129</td>
<td>0.031</td>
<td>(0.088, 0.208)</td>
</tr>
<tr>
<td>Binary Model</td>
<td>$a$</td>
<td>2.172</td>
<td>0.101</td>
<td>(1.977, 2.369)</td>
</tr>
<tr>
<td></td>
<td>$b_1$</td>
<td>1.025</td>
<td>0.108</td>
<td>(0.816, 1.239)</td>
</tr>
<tr>
<td></td>
<td>$b_2$</td>
<td>0.799</td>
<td>0.080</td>
<td>(0.643, 0.961)</td>
</tr>
<tr>
<td>Regression Model</td>
<td>$\rho$</td>
<td>0.342</td>
<td>0.028</td>
<td>(0.284, 0.397)</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_0)$</td>
<td>-0.009</td>
<td>0.066</td>
<td>(-0.142, 0.120)</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_1)$</td>
<td>0.789</td>
<td>0.077</td>
<td>(0.637, 0.945)</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_2)$</td>
<td>0.783</td>
<td>0.079</td>
<td>(0.629, 0.945)</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_3)$</td>
<td>0.800</td>
<td>0.085</td>
<td>(0.632, 0.961)</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_4)$</td>
<td>0.818</td>
<td>0.098</td>
<td>(0.614, 1.008)</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_5)$</td>
<td>0.722</td>
<td>0.134</td>
<td>(0.460, 0.978)</td>
</tr>
<tr>
<td></td>
<td>$\beta_0(t_6)$</td>
<td>0.643</td>
<td>0.237</td>
<td>(0.145, 1.098)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_0)$</td>
<td>0.026</td>
<td>0.112</td>
<td>(-0.196, 0.244)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_1)$</td>
<td>0.033</td>
<td>0.127</td>
<td>(-0.222, 0.279)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_2)$</td>
<td>0.201</td>
<td>0.132</td>
<td>(-0.057, 0.456)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1(t_3)$</td>
<td>0.225</td>
<td>0.138</td>
<td>(-0.045, 0.496)</td>
</tr>
</tbody>
</table>

continued on next page
<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Mean</th>
<th>Standard Error</th>
<th>95% Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1(t_4)$</td>
<td>0.089</td>
<td>0.158</td>
<td>(-0.228,0.401)</td>
<td></td>
</tr>
<tr>
<td>$\beta_1(t_5)$</td>
<td>0.159</td>
<td>0.217</td>
<td>(-0.278,0.582)</td>
<td></td>
</tr>
<tr>
<td>$\beta_1(t_6)$</td>
<td>0.358</td>
<td>0.414</td>
<td>(-0.403,1.194)</td>
<td></td>
</tr>
<tr>
<td>Cox’s Model</td>
<td>$\delta_0$</td>
<td>-0.162</td>
<td>0.189</td>
<td>(-0.531,0.212)</td>
</tr>
<tr>
<td></td>
<td>$\delta_1$</td>
<td>-1.037</td>
<td>0.463</td>
<td>(-2.023,-0.195)</td>
</tr>
<tr>
<td></td>
<td>$\delta_2$</td>
<td>0.341</td>
<td>0.654</td>
<td>(-0.794,1.787)</td>
</tr>
</tbody>
</table>
APPENDIX A. CONVERGENCE DIAGNOSTICS AND POSTERIOR SUMMARY
Appendix B

EM procedure and Results

The EM algorithm for the semiparametric joint model is presented in Figure B.1. In the initialization step (iteration 0), $\Theta^{(0)}$ is defined as the initial values for set of parameters in the stereotype model, $\lambda^{(0)}$ is defined as the baseline hazard initial values for all individuals and $\delta^{(0)}$ is defined as the initial values for set of parameters in the Cox proportional hazards model. Next, we iterate between the expectation and maximization step until convergence. At iteration $t$, we substitute $\Theta^{(t-1)}$, $\lambda^{(t-1)}$ and $\delta^{(t-1)}$ into equation (4.32) to obtain the log likelihood value, $\ell(\Theta^{(t)}, \lambda^{(t)}, \delta^{(t)})$, at iteration $t$. Convergence is reached if the absolute relative difference between two consecutive iterations is close to zero.
APPENDIX B. EM PROCEDURE AND RESULTS

\[ \Theta^{(0)}, \lambda^{(0)}, \delta^{(0)} \]

\[ t = 1 \]

Given \( \Theta^{(t-1)}, \lambda^{(t-1)} \) and \( \delta^{(t-1)} \), obtain \( E^{(t)} \)

Maximize \( \ell(\Theta^{(t-1)}, \lambda^{(t-1)}, \delta^{(t-1)}) \) to obtain \( \Theta^{(t)} \) and \( \delta^{(t)} \) by using an alternating algorithm (fixing between \( \varphi_{i} \) and \( (b_{j} + \theta_{r}) \))

Calculate \( \lambda^{(t)} \) from \( \Theta^{(t)}, \delta^{(t)} \) and \( E^{(t)} \)

\[ t \geq 2 \]

No \( \rightarrow t = t + 1 \)

Yes \( \rightarrow \]

\[ \frac{L(\Theta^{(t+1)}|Y,Z) - L(\Theta^{(t)}|Y,Z)}{L(\Theta^{(t)}|Y,Z)} \approx 0 \]

No \( \rightarrow \]

Yes \( \rightarrow \)

STOP

Figure B.1: The modified profile likelihood EM algorithm for the semi-parametric joint model
225

Estimated parameters
Table B.1: EM estimate of the competing models
(Model(a) - Model(g))
Parameter

Estimated value (SE)
Model(a)

Model(b)

Model(c)

Model(d)

Model(e)

Model(f)

Model(g)

a2

0.844(0.027)

0.844(0.026)

0.844(0.027)

0.844(0.027)

0.844(0.027)

0.844(0.027)

0.845(0.027)

a3

1.086(0.044)

0.832(0.049)

0.300(0.043)

0.300(0.043)

0.295(0.043)

0.297(0.043)

0.300(0.043)

a4

1.184(0.065)

0.154(0.066)

-0.649(0.061)

-0.649(0.061)

-0.653(0.061)

-0.649(0.061)

-0.650(0.061)

a5

0.145(0.113)

-1.836(0.109)

-3.342(0.106)

-3.343(0.106)

-3.350(0.106)

-3.341(0.106)

-3.353(0.106)

b2

1.688(0.132)

2.406(0.133)

2.504(0.131)

2.505(0.131)

2.510(0.131)

2.504(0.131)

2.515(0.131)

b3

3.110(0.138)

4.328(0.137)

4.578(0.136)

4.578(0.136)

4.588(0.136)

4.582(0.136)

4.584(0.136)

b4

4.193(0.144)

5.135(0.142)

5.465(0.142)

5.466(0.142)

5.475(0.142)

5.470(0.142)

5.474(0.142)

b5

2.275(0.134)

2.863(0.133)

3.035(0.133)

3.035(0.133)

3.043(0.133)

3.038(0.133)

3.044(0.133)

b6

-0.259(0.136)

0.233(0.138)

0.166(0.135)

0.167(0.135)

0.170(0.135)

0.1649(0.135)

0.177(0.135)

b7

2.168(0.134)

2.687(0.133)

2.836(0.133)

2.837(0.133)

2.844(0.133)

2.839(0.133)

2.846(0.133)

b8

0.084(0.135)

0.667(0.137)

0.609(0.133)

0.610(0.133)

0.613(0.133)

0.607(0.133)

0.620(0.133)

b9

3.704(0.141)

4.619(0.139)

4.903(0.140)

4.903(0.140)

4.913(0.140)

4.907(0.140)

4.911(0.140)

b10

-0.050(0.135)

0.495(0.138)

0.427(0.133)

0.428(0.133)

0.430(0.133)

0.424(0.133)

0.438(0.134)

b11

2.801(0.136)

3.584(0.134)

3.785(0.134)

3.786(0.134)

3.794(0.134)

3.789(0.134

3.794(0.134)

b12

1.286(0.133)

1.625(0.132)

1.710(0.132)

1.711(0.132)

1.720(0.132)

1.715(0.132)

1.719(0.132)

b13

0.837(0.133)

1.414(0.135)

1.409(0.132)

1.410(0.132)

1.412(0.132)

1.407(0.132)

1.421(0.132)

b14

1.274(0.133)

1.657(0.132)

1.719(0.132)

1.720(0.132)

1.726(0.132)

1.721(0.132)

1.729(0.132)

b15

2.179(0.134)

2.747(0.133)

2.881(0.132)

2.881(0.132)

2.888(0.132)

2.883(0.132)

2.891(0.132)

b16

2.849(0.136)

3.533(0.135)

3.716(0.134)

3.716(0.134)

3.723(0.134)

3.718(0.134)

3.726(0.134)

b17

1.556(0.133)

2.051(0.133)

2.127(0.132)

2.128(0.132)

2.134(0.132)

2.128(0.132)

2.138(0.132)

b18

1.730(0.133)

2.224(0.132)

2.326(0.131)

2.327(0.131)

2.333(0.131)

2.328(0.131)

2.336(0.131)

b19

1.392(0.133)

1.782(0.132)

1.861(0.131)

1.862(0.131)

1.869(0.131)

1.863(0.131)

1.871(0.131)

b20

1.949(0.133)

2.500(0.133)

2.623(0.132)

2.623(0.132)

2.631(0.132)

2.625(0.132)

2.633(0.132)

b21

1.843(0.133)

2.364(0.133)

2.472(0.132)

2.473(0.132)

2.479(0.132)

2.474(0.132)

2.482(0.132)

b22

2.105(0.134)

3.133(0.136)

3.228(0.133)

3.229(0.133)

3.231(0.133)

3.225(0.133)

3.239(0.133)

b23

1.096(0.133)

2.170(0.135)

2.223(0.131)

2.224(0.131)

2.227(0.131)

2.221(0.131)

2.232(0.131)

b24

0.447(0.133)

1.413(0.136)

1.420(0.132)

1.420(0.132)

1.424(0.132)

1.418(0.132)

1.429(0.132)

b25

2.254(0.134)

3.144(0.136)

3.238(0.133)

3.239(0.133)

3.241(0.133)

3.236(0.133)

3.251(0.133)

b26

1.295(0.132)

2.406(0.133)

2.528(0.131)

2.528(0.131)

2.537(0.131)

2.531(0.131)

2.533(0.131)

b27

1.894(0.132)

2.307(0.131)

2.433(0.131)

2.433(0.131)

2.442(0.131)

2.436(0.131)

2.442(0.131)

b28

3.596(0.140)

4.445(0.138)

4.716(0.138)

4.716(0.138)

4.725(0.138)

4.720(0.138)

4.725(0.138)

b29

4.705(0.151)

5.697(0.149)

6.097(0.149)

6.097(0.149)

6.109(0.149)

6.104(0.149)

6.105(0.149)

b30

6.410(0.182)

7.671(0.181)

8.242(0.180)

8.242(0.180)

8.257(0.180)

8.252(0.180)

8.245(0.180)

b31

3.473(0.140)

4.276(0.139)

4.539(0.139)

4.540(0.139)

4.549(0.139)

4.544(0.139)

4.548(0.139)

b32

2.156(0.135)

2.757(0.134)

2.900(0.133)

2.901(0.133)

2.908(0.133)

2.903(0.133)

2.909(0.133)

continued on next page


Table B.2: EM estimate of the competing models (Model(h) - Model(o))

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model(h)</th>
<th>Model(i)</th>
<th>Model(j)</th>
<th>Model(k)</th>
<th>Model(l)</th>
<th>Model(m)</th>
<th>Model(o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>θ</td>
<td>3.085(0.052)</td>
<td>3.085(0.052)</td>
<td>3.085(0.052)</td>
<td>3.085(0.052)</td>
<td>3.085(0.052)</td>
<td>3.085(0.052)</td>
<td>3.085(0.052)</td>
</tr>
<tr>
<td>a</td>
<td>0.848(0.027)</td>
<td>0.848(0.027)</td>
<td>0.848(0.027)</td>
<td>0.848(0.027)</td>
<td>0.848(0.027)</td>
<td>0.848(0.027)</td>
<td>0.848(0.027)</td>
</tr>
<tr>
<td>b</td>
<td>2.738(0.131)</td>
<td>2.738(0.131)</td>
<td>2.738(0.131)</td>
<td>2.738(0.131)</td>
<td>2.738(0.131)</td>
<td>2.738(0.131)</td>
<td>2.738(0.131)</td>
</tr>
<tr>
<td>c</td>
<td>1.725(0.131)</td>
<td>1.725(0.131)</td>
<td>1.725(0.131)</td>
<td>1.725(0.131)</td>
<td>1.725(0.131)</td>
<td>1.725(0.131)</td>
<td>1.725(0.131)</td>
</tr>
<tr>
<td>d</td>
<td>1.401(0.131)</td>
<td>1.401(0.131)</td>
<td>1.401(0.131)</td>
<td>1.401(0.131)</td>
<td>1.401(0.131)</td>
<td>1.401(0.131)</td>
<td>1.401(0.131)</td>
</tr>
<tr>
<td>e</td>
<td>1.717 (0.131)</td>
<td>1.717 (0.131)</td>
<td>1.717 (0.131)</td>
<td>1.717 (0.131)</td>
<td>1.717 (0.131)</td>
<td>1.717 (0.131)</td>
<td>1.717 (0.131)</td>
</tr>
<tr>
<td>f</td>
<td>3.124 (0.131)</td>
<td>3.124 (0.131)</td>
<td>3.124 (0.131)</td>
<td>3.124 (0.131)</td>
<td>3.124 (0.131)</td>
<td>3.124 (0.131)</td>
<td>3.124 (0.131)</td>
</tr>
</tbody>
</table>

continued on next page
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model(h)</th>
<th>Model(i)</th>
<th>Model(j)</th>
<th>Model(k)</th>
<th>Model(l)</th>
<th>Model(m)</th>
<th>Model(o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_{20}$</td>
<td>2.622(0.132)</td>
<td>2.632(0.132)</td>
<td>2.624(0.132)</td>
<td>2.625(0.132)</td>
<td>2.695(0.135)</td>
<td>2.696(0.135)</td>
<td>2.734(0.134)</td>
</tr>
<tr>
<td>$b_{21}$</td>
<td>2.470(0.132)</td>
<td>2.481(0.132)</td>
<td>2.473(0.132)</td>
<td>2.474(0.132)</td>
<td>2.539(0.135)</td>
<td>2.540(0.135)</td>
<td>2.577(0.134)</td>
</tr>
<tr>
<td>$b_{22}$</td>
<td>3.221(0.133)</td>
<td>3.237(0.133)</td>
<td>3.230(0.133)</td>
<td>3.230(0.133)</td>
<td>3.342(0.136)</td>
<td>3.342(0.136)</td>
<td>3.379(0.135)</td>
</tr>
<tr>
<td>$b_{23}$</td>
<td>2.217(0.131)</td>
<td>2.233(0.131)</td>
<td>2.222(0.131)</td>
<td>2.225(0.131)</td>
<td>2.321(0.134)</td>
<td>2.321(0.134)</td>
<td>2.339(0.134)</td>
</tr>
<tr>
<td>$b_{24}$</td>
<td>3.231(0.133)</td>
<td>3.248(0.133)</td>
<td>3.237(0.133)</td>
<td>3.240(0.133)</td>
<td>3.346(0.136)</td>
<td>3.347(0.136)</td>
<td>3.385(0.135)</td>
</tr>
<tr>
<td>$b_{25}$</td>
<td>2.530(0.131)</td>
<td>2.537(0.131)</td>
<td>2.529(0.131)</td>
<td>2.530(0.131)</td>
<td>2.634(0.134)</td>
<td>2.644(0.134)</td>
<td>2.663(0.133)</td>
</tr>
<tr>
<td>$b_{26}$</td>
<td>2.434(0.131)</td>
<td>2.442(0.131)</td>
<td>2.435(0.131)</td>
<td>2.435(0.131)</td>
<td>2.560(0.133)</td>
<td>2.570(0.133)</td>
<td>2.543(0.133)</td>
</tr>
<tr>
<td>$b_{27}$</td>
<td>4.717(0.130)</td>
<td>4.728(0.130)</td>
<td>4.718(0.130)</td>
<td>4.719(0.130)</td>
<td>4.829(0.141)</td>
<td>4.830(0.141)</td>
<td>4.889(0.141)</td>
</tr>
<tr>
<td>$b_{28}$</td>
<td>6.102(0.149)</td>
<td>6.107(0.149)</td>
<td>6.100(0.149)</td>
<td>6.100(0.149)</td>
<td>6.217(0.152)</td>
<td>6.219(0.152)</td>
<td>6.302(0.152)</td>
</tr>
<tr>
<td>$b_{29}$</td>
<td>8.253(0.180)</td>
<td>8.258(0.180)</td>
<td>8.246(0.180)</td>
<td>8.246(0.180)</td>
<td>8.413(0.183)</td>
<td>8.416(0.183)</td>
<td>8.520(0.183)</td>
</tr>
<tr>
<td>$b_{30}$</td>
<td>4.541(0.139)</td>
<td>4.549(0.139)</td>
<td>4.542(0.139)</td>
<td>4.542(0.139)</td>
<td>4.634(0.141)</td>
<td>4.635(0.141)</td>
<td>4.796(0.141)</td>
</tr>
<tr>
<td>$b_{31}$</td>
<td>2.899(0.133)</td>
<td>2.909(0.133)</td>
<td>2.901(0.133)</td>
<td>2.902(0.133)</td>
<td>2.972(0.136)</td>
<td>2.972(0.136)</td>
<td>3.013(0.136)</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.368</td>
<td>0.366</td>
<td>0.367</td>
<td>0.367</td>
<td>0.363</td>
<td>0.363</td>
<td>0.365</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>0.568</td>
<td>0.568</td>
<td>0.568</td>
<td>0.568</td>
<td>0.573</td>
<td>0.573</td>
<td>0.570</td>
</tr>
<tr>
<td>$\delta_0$</td>
<td>-0.549(0.078)</td>
<td>-0.546(0.078)</td>
<td>-0.193(0.087)</td>
<td>-0.226(0.078)</td>
<td>-0.175(0.041)</td>
<td>-0.196(0.041)</td>
<td>-0.175(0.050)</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>-0.429(0.291)</td>
<td>-0.334(0.290)</td>
<td>-0.99(0.047)</td>
<td>-0.897(0.435)</td>
<td>-0.95(0.442)</td>
<td>-0.966(0.442)</td>
<td>-0.937(0.435)</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.047(0.004)</td>
<td>1.861(1.162)</td>
<td>0.036(0.007)</td>
<td>0.714(0.286)</td>
<td>-</td>
<td>0.020(0.099)</td>
<td>-</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>-0.537(0.328)</td>
<td>-0.2(0.307)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>5.286(0.050)</td>
<td>5.281(0.058)</td>
<td>5.281(0.058)</td>
<td>5.279(0.058)</td>
<td>2.220(0.097)</td>
<td>3.973(0.097)</td>
<td>3.789(0.064)</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>2.586(0.047)</td>
<td>2.586(0.047)</td>
<td>2.584(0.047)</td>
<td>2.586(0.047)</td>
<td>3.971(0.694)</td>
<td>6.275(0.694)</td>
<td>11.019(0.716)</td>
</tr>
<tr>
<td>$\theta_4$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6.271(1.115)</td>
<td>2.271(1.115)</td>
<td>5.407(0.073)</td>
</tr>
<tr>
<td>$\theta_5$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2.234(0.082)</td>
</tr>
<tr>
<td>$\theta_6$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.171(0.087)</td>
</tr>
</tbody>
</table>
Appendix C

MOS-HIV questionnaire
MOS-HIV HEALTH SURVEY

INSTRUCTIONS TO THE STUDY COORDINATOR:

The following questionnaire asks the patient about many aspects of his/her health and health care. It should be given to the patient prior to the clinical exam and preferably in a quiet secluded area (e.g., exam room or other office).

It is important to be familiar with the content and format of the questionnaire before giving it to study participants. At the first visit, please begin by telling the participant:

"We would like you to answer some questions about how you are feeling and the kinds of things you are able to do. Your answers will help us understand the effects of the medication you are taking. We appreciate your filling out this questionnaire."

You should then briefly go over the format of the questions and how to complete them. Have the participant complete the questionnaire before vital signs, history and physical are completed.

The questionnaire is very brief and should take no more than 10 minutes to complete. Before giving the patient the questionnaire, please fill out the header(s) and DETACH THIS PAGE.

Each question is in the same general format. Note that the patient is always asked to check one box for each question. All questions refer to the PAST 4 WEEKS.

Collect the completed questionnaire before the clinical exam. Before going on, review the questionnaire for omissions. If the participant missed any of the questions, point this out and have him/her complete the omissions.

PLEASE COMPLETE THE FOLLOWING ITEMS AFTER PATIENT COMPLETES THE QUESTIONNAIRE OR AFTER YOU ASCERTAIN THAT THIS IS NOT POSSIBLE:

1. How was the questionnaire completed? ............ 1 □ Self administered by the study participant
   2 □ Face-to-face interview that you conducted
   3 □ Phone interview
   4 □ Not completed
   5 □ Other

   If Other, specify: ________________________________

2. If you answered 2 or 4, please indicate the reason(s) why:

   Patient refused initially: .......................... 1 □ Yes 2 □ No
   Patient's reading level not adequate: .................. 1 □ Yes 2 □ No
   There was not enough time: .......................... 1 □ Yes 2 □ No
   Patient forgot reading glasses: .......................... 1 □ Yes 2 □ No
   Other reason: .......................... 1 □ Yes 2 □ No

   If Other, specify: ________________________________
INSTRUCTIONS TO PATIENT: Please answer the following questions by placing a " " in the appropriate box.

1. In general, would you say your health is: (Check One)
   - Excellent ................................................................. 1 □
   - Very Good ............................................................. 2 □
   - Good ................................................................. 3 □
   - Fair ................................................................. 4 □
   - Poor ................................................................. 5 □

2. How much bodily pain have you generally had during the past 4 weeks? (Check One)
   - None ................................................................. 1 □
   - Very Mild ............................................................ 2 □
   - Mild ................................................................. 3 □
   - Moderate ............................................................ 4 □
   - Severe ............................................................... 5 □
   - Very Severe .......................................................... 6 □

3. During the past 4 weeks, how much did pain interfere with your normal work (or your normal activities, including work outside the home and housework)? (Check One)
   - Not at all ................................................................. 1 □
   - A little bit ............................................................. 2 □
   - Moderately ............................................................ 3 □
   - Quite a bit ............................................................ 4 □
   - Extremely ............................................................. 5 □
4. The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

<table>
<thead>
<tr>
<th>Activity Description</th>
<th>YES, limited a lot</th>
<th>YES, limited a little</th>
<th>NO, not limited</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. The kinds or amounts of vigorous activities you can do, like lifting heavy objects, running or participating in strenuous sports.</td>
<td>1 □</td>
<td>2 □</td>
<td>3 □</td>
</tr>
<tr>
<td>b. The kinds or amounts of moderate activities you can do, like moving a table, carrying groceries or bowling.</td>
<td>1 □</td>
<td>2 □</td>
<td>3 □</td>
</tr>
<tr>
<td>c. Walking uphill or climbing (a few flights of stairs).</td>
<td>1 □</td>
<td>2 □</td>
<td>3 □</td>
</tr>
<tr>
<td>d. Bending, lifting or stooping.</td>
<td>1 □</td>
<td>2 □</td>
<td>3 □</td>
</tr>
<tr>
<td>e. Walking one block.</td>
<td>1 □</td>
<td>2 □</td>
<td>3 □</td>
</tr>
<tr>
<td>f. Eating, dressing, bathing or using the toilet.</td>
<td>1 □</td>
<td>2 □</td>
<td>3 □</td>
</tr>
</tbody>
</table>

5. Does your health keep you from working at a job, doing work around the house or going to school? (Check One)

- [ ] Yes
- [ ] No

6. Have you been unable to do certain kinds or amounts of work, housework, or schoolwork because of your health? (Check One)

- [ ] Yes
- [ ] No
For each of the following questions, please check the box for the one answer that comes closest to the way you have been feeling during the past 4 weeks.

<table>
<thead>
<tr>
<th>Question</th>
<th>All of the Time</th>
<th>Most of the Time</th>
<th>A Good Bit of the Time</th>
<th>Some of the Time</th>
<th>A Little of the Time</th>
<th>None of the Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. How much of the time, during the past 4 weeks, has your health limited your social activities (like visiting with friends or close relatives)?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>8. How much of the time, during the past 4 weeks:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Have you been a very nervous person?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>b. Have you felt calm and peaceful?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>c. Have you felt downhearted and blue?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>d. Have you been a happy person?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>e. Have you felt so down in the dumps that nothing could cheer you up?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
For each of the following questions, please check the box for the one answer that comes closest to the way you have been feeling during the past 4 weeks.

<table>
<thead>
<tr>
<th></th>
<th>All of the Time</th>
<th>Most of the Time</th>
<th>A Good Bit of the Time</th>
<th>Some of the Time</th>
<th>A Little of the Time</th>
<th>None of the Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.</td>
<td>How often during the past four weeks:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a.</td>
<td>Did you feel full of pep?</td>
<td>1 ☐</td>
<td>2 ☐</td>
<td>3 ☐</td>
<td>4 ☐</td>
<td>5 ☐</td>
</tr>
<tr>
<td>b.</td>
<td>Did you feel worn out?</td>
<td>1 ☐</td>
<td>2 ☐</td>
<td>3 ☐</td>
<td>4 ☐</td>
<td>5 ☐</td>
</tr>
<tr>
<td>c.</td>
<td>Did you feel tired?</td>
<td>1 ☐</td>
<td>2 ☐</td>
<td>3 ☐</td>
<td>4 ☐</td>
<td>5 ☐</td>
</tr>
<tr>
<td>d.</td>
<td>Did you have enough energy to do the things you wanted to do?</td>
<td>1 ☐</td>
<td>2 ☐</td>
<td>3 ☐</td>
<td>4 ☐</td>
<td>5 ☐</td>
</tr>
<tr>
<td>e.</td>
<td>Did you feel weighed down by your health problems?</td>
<td>1 ☐</td>
<td>2 ☐</td>
<td>3 ☐</td>
<td>4 ☐</td>
<td>5 ☐</td>
</tr>
<tr>
<td>f.</td>
<td>Were you discouraged by your health problems?</td>
<td>1 ☐</td>
<td>2 ☐</td>
<td>3 ☐</td>
<td>4 ☐</td>
<td>5 ☐</td>
</tr>
<tr>
<td>g.</td>
<td>Did you feel despair over your health problems?</td>
<td>1 ☐</td>
<td>2 ☐</td>
<td>3 ☐</td>
<td>4 ☐</td>
<td>5 ☐</td>
</tr>
<tr>
<td>h.</td>
<td>Were you afraid because of your health?</td>
<td>1 ☐</td>
<td>2 ☐</td>
<td>3 ☐</td>
<td>4 ☐</td>
<td>5 ☐</td>
</tr>
</tbody>
</table>
10. How much of the time, during the past 4 weeks:

<table>
<thead>
<tr>
<th></th>
<th>All of the Time</th>
<th>Most of the Time</th>
<th>A Good Bit of the Time</th>
<th>Some of the Time</th>
<th>A Little of the Time</th>
<th>None of the Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Did you have difficulty reasoning and solving problems, for example, making plans, making decisions, learning new things?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>b. Did you forget things that happened recently, for example, where you put things and when you had appointments?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>c. Did you have trouble keeping your attention on any activity for long?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>d. Did you have difficulty doing activities involving concentration and thinking?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

11. Please check the box that best describes whether each of the following statements is true or false for you. (Check one box on each line)

<table>
<thead>
<tr>
<th></th>
<th>Definitely True</th>
<th>Mostly True</th>
<th>Not Sure</th>
<th>Mostly False</th>
<th>Definitely False</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. I am somewhat ill.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>b. I am as healthy as anybody I know.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>c. My health is excellent.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>d. I have been feeling bad lately.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
12. How has the quality of your life been during the past 4 weeks? That is, how have things been going for you? (Check One)

- Very well; could hardly be better ........................................ 1 □
- Pretty good ........................................................................ 2 □
- Good and bad parts about equal ........................................... 3 □
- Pretty bad ........................................................................ 4 □
- Very bad; could hardly be worse ......................................... 5 □

13. How would you rate your physical health and emotional condition now compared to 4 weeks ago? (Check One)

- Much better ........................................................................ 1 □
- A little better ...................................................................... 2 □
- About the same .................................................................. 3 □
- A little worse ...................................................................... 4 □
- Much worse ....................................................................... 5 □

THANK YOU VERY MUCH
Bibliography


Ananworanich, J., Siangphoe, U., Hill, A. and et al. (2005), ‘Highly active antiretroviral therapy (harrt) retreatment in patients on cd4-guided
therapy achieved similar virologic suppression compared with patients on continuous harrt’, *Journal of Acquired Immune Deficiency Syndromes* 39, 523–529.


Centers for Disease Control and Prevention (1992), ‘1993 revised classification system for hiv infection and expanded surveillance case definition


on staging hiv-1 infection’, *Journal of Acquired Immune Deficiency Syndrome* 5, 794–802.


